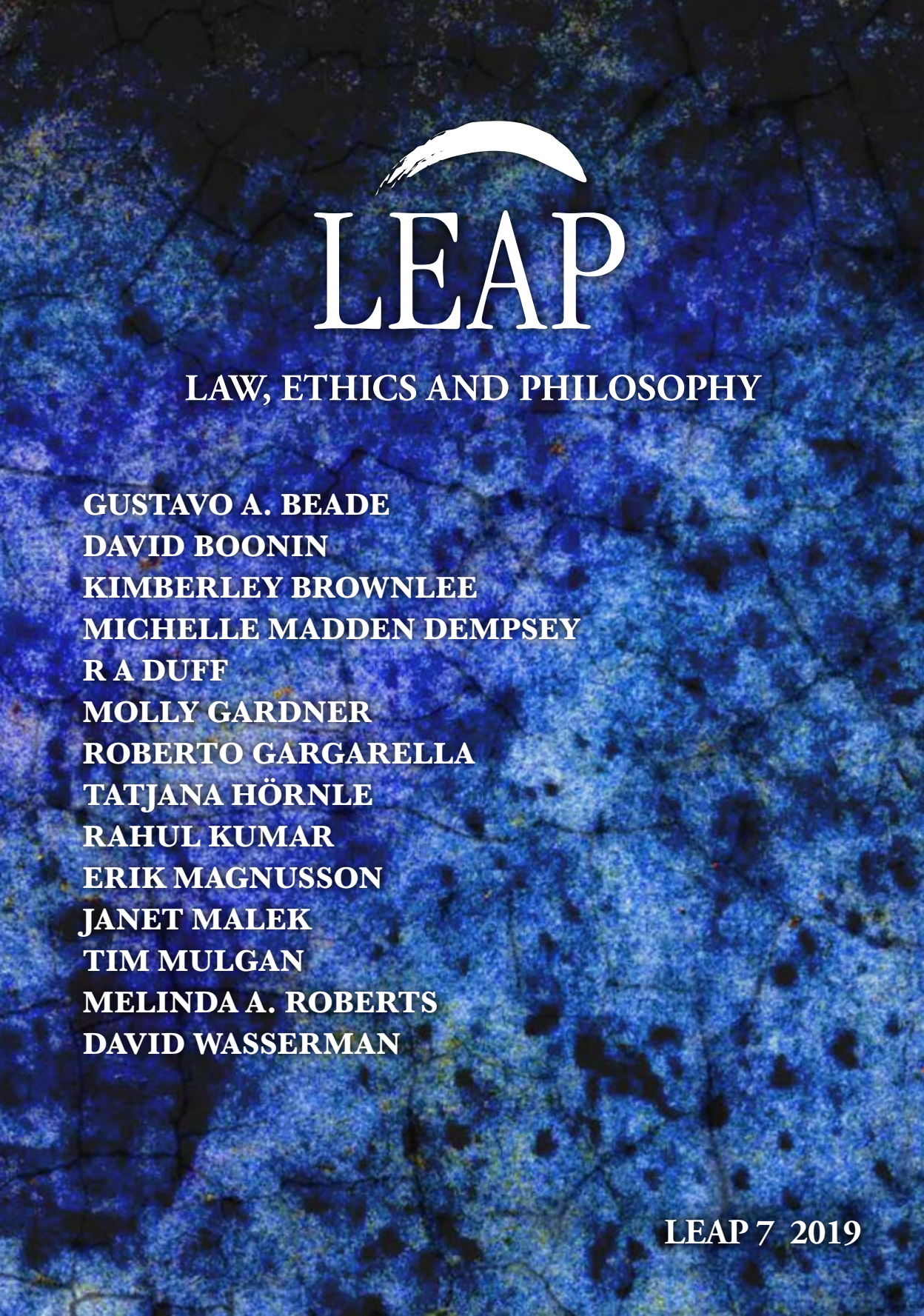
LEAP
LAW, ETHICS AND PHILOSOPHY
GUSTAVO A. BEADE
DAVID BOONIN
KIMBERLEY BROWNLEE
MICHELLE MADDEN DEMPSEY
R A DUFF
MOLLY GARDNER
ROBERTO GARGARELLA
TATJANA HÖRNLE
RAHUL KUMAR
ERIK MAGNUSSON
JANET MALEK
TIM MULGAN
MELINDA A. ROBERTS
DAVID WASSERMAN
LEAP 7 2019

LAW, ETHICS AND PHILOSOPHY

LEAP
Law, Ethics and Philosophy

Law, Ethics and Philosophy (LEAP)

Online ISSN: 2341-1465

LEAP is a peer-reviewed, open access international journal dedicated to work in ethics, legal theory, and social and political philosophy. It welcomes clear, rigorous and original submissions that address concrete issues of public concern as well as more abstract theoretical questions. It also has the distinctive aims of (a) fostering work drawing on a variety of disciplines within the social and natural sciences, including biology, economics, history, and psychology; and (b) promoting dialogue between the Anglophone and non-Anglophone worlds.

We invite submissions of articles up to 10,000 words, discussion notes up to 6,000 words, and replies and exchanges not exceeding 3,000 words. Please send an anonymized version of the submission, along with a cover letter, to:leap.journal@upf.edu. All published submissions will have undergone blind review, and the journal will notify authors of submitted pieces about the progress of their submission within six weeks.

Although LEAP accepts exclusively submissions in English, the journal strongly encourages submissions from authors who also write in languages other than English, and will always strive to ensure that their work is assessed on the basis of its content and not primarily its mode of expression. Where necessary the editorial process provides authors with guidance regarding matters of English style. The journal is published by Pompeu Fabra University and hosted by RACO, the Catalan repository of open access journals (http://raco.cat/index.php/leap)Enquiries regarding the journal may be directed to: leap.journal@upf.edu.

LEAP
Law, Ethics and Philosophy

Vol. 7
2019

Editorial Board

Pompeu Fabra University
http://www.raco.cat/index.php/LEAP

CONTENTS

Law, Ethics and Philosophy (LEAP) Vol. 7, 2019

Page

SYMPOSIUM
ON DAVID BOONIN'S *THE NON-IDENTITY PROBLEM AND THE ETHICS OF FUTURE PEOPLE*

SYMPOSIUM
ON ANTONY DUFF'S
THE REALM OF CRIMINAL LAW

Acknowledgments

This issue of *Law, Ethics and Philosophy* has been possible thanks to the Catalan Agency for Management of University and Research Grants, whose grant to the Law and Philosophy Research Group (SGR 823) has funded the correction, proofreading, copy editing, and typesetting of this volume. It has also been funded by the European Research Council Consolidator Grant (648610) on Justice and the Family: An Analysis of the Normative Significance of Procreation and Parenthood in a Just Society (PI: Serena Olsaretti). We also wish to thank Peter Josse and Martín Comesaña, respectively, for their proofreading and typesetting work..

Symposium on David Boonin's *The Non-Identity Problem and the Ethics of Future People*

DOI: 10.31009/LEAP.2019.V7.01

Introduction

ERIK MAGNUSSON
University of Manitoba

Consider the following case:

> *Wilma:* Wilma has decided to have a baby. She goes to the doctor for a checkup and the doctor tells her that there is some good news and some bad news. The bad news is that as things now stand, if Wilma conceives, her child will have a disability. The doctor cannot say precisely what the disability will be, but she can tell Wilma three things about it. First, it will be the kind of disability that clearly has a substantially negative impact on a person's quality of life. This could be because of features intrinsic to the disability itself, because Wilma's society discriminates against or fails to sufficiently accommodate people with the disability, or because of some combination of these reasons. Second, even if it is possible for a life to be worse than no life at all, this particular disability will clearly not be so serious as to render the child's life worse than no life at all. So while the disability will be considerably far from trivial, the child's life will nonetheless clearly be worth living. Finally, the disability will be irreversible. There will be no way to eliminate it or mitigate its effects.
>
> The good news is that Wilma can prevent this from happening. If she takes a tiny pill once a day for two months before conceiving, her child will be perfectly healthy. The pill is easy to take, has no side effects, and will be paid for by her health insurance. Fully understanding all of the facts about the situation, Wilma decides that having to take a pill once a day for two months before conceiving is a bit too inconvenient and so chooses to throw the pills away and conceive at once. As a result of this choice, her child [Pebbles] is born with a significant and irreversible disability (Boonin, 2014: 2).

Initially, it is tempting to claim that Wilma has acted wrongly by refusing to delay conception, and that the wrongness of her action can be explained in terms of its negative impact on the welfare of Pebbles. But as Derek Parfit (1984: chapter 16) and others have argued, this intuitive reaction is complicated by the fact that our genetic identities are dependent on the timing of conception. Had Wilma followed her doctor's advice and delayed conception for two months, this would not have improved Pebbles'

condition, but would rather have resulted in a different child altogether, one composed of a unique set of reproductive gametes. Assuming that life with her particular disability is better than no life at all, Pebbles has not in fact been made any worse off as a result of Wilma's decision, as she would not have existed had Wilma decided differently. This, in turn, raises a puzzling question: what, if anything, is morally objectionable about Wilma's decision?

Providing an answer to this question and others like it has become the preoccupation of a vast literature on what is now known as the *non-identity problem*. The non-identity problem arises in cases where an action or decision that appears to harm a person is also a necessary condition of their existence. It is generated from the mundane fact that our genetic identities are dependent on the timing of conception, such that if your parents had not conceived when they in fact did, the resultant child would not have been you, but rather a genetically distinct (or 'non-identical') child composed of a unique set of reproductive gametes. Parfit (1984: 352) expresses this fact as the Time-Dependence Claim, which holds that "If any particular person had not been conceived when he was in fact conceived, it is *in fact* true that he would never have existed."[1] According to Parfit and others, the Time-Dependence Claim forces us to concede that so long as a person has a life worth living, she cannot be harmed by a set of actions or decisions that led to her conception, for her very existence depended on that conception taking place when it did.[2]

In the case of *Wilma*, this presents a problem: it appears as though Wilma has acted wrongly by her refusal to delay conception, yet our most readily available explanation of wrongdoing—that Wilma has *harmed* Pebbles—does not apply in her particular case. Must we conclude, then, that Wilma has acted permissibly? Most philosophers have attempted to avoid this conclusion by explaining how an action or decision can be morally wrong even if it does not make any particular person worse off than they otherwise would have been. For example, some have argued that it not necessary to be made worse off in order to be harmed; that it is not necessary to be harmed in order to be wronged; or that it is not necessary

[1] Parfit (1984: 352) actually presents two different versions of the Time-Dependence Claim, the latter of which holds that "If any particular person had not been conceived within a month of the time when he was in fact conceived, he would in fact never have existed." Parfit includes the latter version to account for the fact that a differently timed conception that occurred within the same month would still involve the same ovum, and thus the resultant person would not be entirely genetically distinct

[2] I draw here on my presentation of the non-identity problem in Magnusson (2019).

for an action or decision to wrong a particular person in order to be morally wrong. However, in The *Non-Identity Problem and the Ethics of Future People*, David Boonin argues that all of these varied attempts fail, and that we must therefore accept the 'implausible conclusion' that decisions like Wilma's are in fact morally permissible.

Boonin's book is noteworthy not only for its striking conclusion, but also for the comprehensiveness of the argument that leads to it. In the four-plus decades since the independent discovery of the non-identity problem by Parfit (1976), Thomas Schwartz (1978), and Robert Adams (1979), an extensive literature has emerged in response to that problem across a wide range of disciplines, including philosophy, law, political science, and bioethics, to name only a few. Through seven chapters and six appendices, Boonin carefully collects, categorizes, and analyzes an impressive amount of this literature, making his book both a useful entry point and required reading for students and professional academics working on the non-identity problem.

Central to Boonin's argument is his distinctive understanding of the non-identity problem, which he sees as emerging from a tension between "five plausible premises and one implausible conclusion." (Boonin, 2014: 27) In relation to the case of *Wilma* described above, he lists these premises as follows:

> P1: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been
>
> P2: If A's act harms B, then A's act makes B worse off than B would otherwise have been
>
> P3: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not harm anyone other than Pebbles
>
> P4: If an act does not harm anyone, then the act does not wrong anyone
>
> P5: If an act does not wrong anyone, then the act is not morally wrong
>
> C: Wilma's act of conceiving Pebbles is not morally wrong (Boonin, 2014: 27)

Boonin's objective in his book is to show that we should actually accept this conclusion and thereby reject the non-identity *problem* in favour of a coherent non-identity *argument*. To do so, he dedicates each of his chapters to showing how attempts at resisting it by rejecting one of the premises fail, ultimately concluding that (a) we must accept the Implausible Conclusion, but (b) it is not actually as implausible as it initially seems.

Reflecting Boonin's understanding, this Symposium brings together six philosophers whose strategies for addressing the non-identity problem involve rejecting one of the premises that Boonin claims cannot plausibly be rejected. Janet Malek opens the Symposium with a paper exploring two possible strategies for rejecting P1. The first (the *de re*/*de* dicto strategy) is based on the idea that we can conceive of Pebbles in two distinct ways—as the particular individual 'Pebbles' or as an individual who has the property of being 'Wilma's child'—and that the latter conception supports the claim that Wilma's refusal to delay conception has in fact made Pebbles worse off than she otherwise would have been. The second (the *metaphysical* strategy) challenges the extent to which the possible children in non-identity cases are actually distinct by emphasizing the non-genetic determinants of identity that will go on to shape the individual who turns out to be 'Wilma's child'. According to Malek, both of these strategies provide plausible ways of constructing continuity of identity among the possible persons implicated in non-identity cases, and therefore provide grounds for rejecting P1.

Molly Gardner follows Malek's discussion with a paper challenging P2. As stated above, P2 assumes that the notion of harm must be understood in a counterfactual sense, according to which a person is harmed by an act if they are made worse off than they otherwise would have been had the act not been performed. On this conception, Pebbles cannot be harmed as a result of Wilma's decision, as she has a life worth living and would not have existed had Wilma acted differently. However, Gardner argues that there is a plausible alternative conception of harm that allows us to claim that Pebbles *has* in fact been harmed in a morally relevant sense. According to the Existence Account of harming, a state of affairs, T, is a harm for an individual, S, if and only if:

> (i) There is an essential component of T that is a condition with respect to which S can be intrinsically better or worse off; and
>
> (ii) If S existed and T had not obtained, then S would be better off with respect to that condition.

According to Gardner, the Existence Account is capable of providing a coherent harm-based solution to the non-identity problem, one that involves less bullet-biting than alternative harm-based accounts that have been proposed in the literature, and one that meets Boonin's own criteria for a satisfactory solution.

Rahul Kumar is next with a paper challenging P4. As presented above, P4 assumes that being harmed is a necessary condition of being wronged,

though this assumption has notably been challenged by proponents of rights-based solutions to the non-identity problem, who argue that it is possible to violate the rights of a person without making them worse off. Boonin considers and responds to a variety of rights-based solutions in Chapter 5 of his book, though Kumar argues that many of these accounts misconstrue non-consequentialist thinking about rights and their moral significance, and thus Boonin's objections to them do not foreclose the possibility of a tenable rights-based solution. According to Kumar, such a solution is possible, though it must focus on the general right to have one's vital interests adequately considered in the regulation of others' conduct, and this general right must be understood as attaching to prospective children in the *de dicto* sense, rather than to any particular individual.

David Wasserman follows Kumar's discussion with a paper exploring three challenges that might be raised against either P4 or P5, all of which emerge from possible deficiencies in Wilma's attitude toward her procreative decision. The first involves conceiving of Wilma as the steward of a precious gift—the 'gift of life'—that may be bestowed in ways that express disrespect to the recipient or to the gift itself. The second involves seeing Wilma as deviously creating an obligation to her child that is impossible to satisfy, namely, the obligation to protect her child from the harm of blindness. Finally, the third involves seeing Wilma as displaying insensitivity toward the harm of blindness in the way that she responds to the prospect of that harm, by choosing to expose her child to it for selfish, trivial reasons. After canvassing each challenge, Wasserman ultimately endorses a version of the third, concluding that it is possible to claim that Wilma has displayed an objectionable insensitivity in her decision to create a blind child, and that this insensitivity may leave Pebbles with legitimate grounds for complaint.

Tim Mulgan follows Wasserman's discussion with a paper that exclusively challenges P5, or the claim that if an act does not wrong anyone then it is not morally wrong. Mulgan structures his argument around a case in which this claim appears to be false, namely, a case in which humanity elects to become extinct in the face of a planetary crisis rather than to continue existing in space through a 'generation starship' program. Mulgan argues that the decision to allow the extinction of humanity is morally wrong despite wronging no one, and that any plausible moral theory must be able to accommodate this intuition. He then explains how rule consequentialism has the resources to support this intuition, and in turn to provide a way of avoiding Boonin's Implausible Conclusion in the case of *Wilma*.

Melinda Roberts follows Mulgan's discussion with a final paper

challenging Boonin's general approach toward analyzing the non-identity problem. As indicated above, Boonin approaches the non-identity problem via a core 'same-number' choice in which a woman, Wilma, must choose between creating one of two children who will exist at different levels of welfare. He then extends his analysis to other cases in which the non-identity problem arises, including 'different-number' choices in which we must choose between courses of action that will result in different numbers of people existing at different levels of welfare. However, Roberts argues that there are two interrelated problems with this type of approach. First, she argues that the focus on *Wilma* leads Boonin to take an excessively narrow view of the facts in non-identity cases and thereby treat what are plausibly three-option cases (in which an agent can choose between (a) not causing a person to exist, (b) causing that person to exist with a disadvantage, or (c) causing a person to exist *without* a disadvantage) as two-option cases (in which an agent can choose only between (a) and (b)). Second, she argues that this narrow view of the facts in non-identity cases leads Boonin to take an excessively narrow view of the conceptual elements of the non-identity problem, particularly the relevant baseline against which to determine whether a person has been made worse off. Following Roberts' paper, the Symposium concludes with a reply by David Boonin, who clarifies his arguments and responds to the various objections that have been raised by the commentators.

Like Boonin's book, I hope this Symposium contributes further to our understanding of the non-identity problem and the many challenges it poses in the context of procreative and intergenerational ethics. In closing, I would like to express my gratitude to David Boonin and the authors for their participation, to the anonymous referees for their insightful feedback, and to Serena Olsaretti and Iñigo González-Ricoy for their constant assistance throughout the editorial process.

BIBLIOGRAPHY

Adams, R., 1979: "Existence, Self-Interest, and the Problem of Evil", Noûs, 13: 53-65.

Boonin, D., 2014: *The Non-Identity Problem and the Ethics of Future People*, Oxford: Oxford University Press.

Magnusson, E., 2019: "Children's Rights and the Non-Identity Problem", *Canadian Journal of Philosophy*, 49(5): 580-605.

Parfit, D., 1976: "On Doing the Best for our Children", in *Ethics and Population*, ed. M. Bayles, 100-115, Cambridge, MA: Schenkman.

Parfit, D., 1984: *Reasons and Persons*, Oxford: Clarendon Press.

Schwartz, T., 1978: "Obligations to Posterity", in *Obligations to Future Generations*, eds. R. Sikora and B. Barry, 3-13, Philadelphia: Temple University Press.

DOI: 10.31009/LEAP.2019.V7.02

The Non-Identity Non-Problem: Constructing Continuity of Identity among Possible Persons

JANET MALEK
Baylor College of Medicine

ABSTRACT

In *The Non-Identity Problem and the Ethics of Future People*, David Boonin carefully analyzes five plausible premises that together entail the "Implausible Conclusion" that a woman (Wilma) does nothing morally wrong by declining to take a pill every day for two months that will prevent her future child (Pebbles) from being blind. Boonin finds that there is no defensible reason to reject any of these premises, ultimately claiming that the non-identity problem should not be understood as a "problem" but as a sound argument with an unintuitive conclusion that should be accepted. This paper examines Boonin's defense of just one of these five premises, P1, that Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been. I first offer some preliminary comments about Boonin's approach, then review and critique his rebuttals of various objections that have been made to P1. I finally defend the claim that there is at least one morally relevant sense in which P1 is false and therefore in which Wilma's act of conceiving now makes Pebbles worse off than she otherwise would have been.

Keywords: Non-Identity Problem, Possible Persons, Future Children, Harm, Conception, Derek Parfit.

1. INTRODUCTION

David Boonin (2014) describes the non-identity problem from the perspective of Wilma, who is told by her physician that if she conceives a child immediately, her child (referred to as Pebbles) will be blind. However, if Wilma takes a pill every day for two months and then conceives a child, her child (referred to as Rocks) will have normal sight. Despite having this

information, Wilma decides to conceive immediately and her child is born blind. Although many people share the intuition that Wilma has done something morally wrong by declining to wait to conceive, Derek Parfit (1984) has pointed out that explaining why Wilma's act is wrong is more difficult than it first appears. Because the genetic identity of Wilma's child is dependent upon the timing of her conception, Parfit argues, it is not coherent to say that Pebbles was harmed by Wilma's conception decision because otherwise she would not have existed—Rocks would have existed instead. And if Pebbles is not harmed, there is no justification for claiming that Wilma's choice is wrong.

In The Non-Identity Problem and the Ethics of Future People, Boonin (2014) carefully analyzes five plausible premises that together entail the "Implausible Conclusion" that Wilma does nothing morally wrong by choosing to conceive Pebbles (who is blind) rather than waiting to conceive Rocks (who has normal sight). These are:

> P1: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been;
>
> P2: If A's act harms B, then A's act makes B worse off than B would otherwise have been;
>
> P3: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not harm anyone other than Pebbles;
>
> P4: If an act does not harm anyone, then the act does not wrong anyone;
>
> P5: If an act does not wrong anyone, then the act is not morally wrong;
>
> C: Wilma's act of conceiving Pebbles is not morally wrong (27).

Boonin finds that there is no defensible reason to reject any of these premises, ultimately claiming that the non-identity problem should not be understood as a "problem" but rather as a sound argument with an unintuitive conclusion that should be accepted. This paper examines Boonin's defense of just one of these five premises, P1. I first offer some preliminary comments about Boonin's approach, then review and critique his rebuttals of various objections that have been made to P1. I finally defend the claim that there is at least one morally relevant sense in which P1 is false and therefore in which Wilma's act of conceiving now makes Pebbles worse off than she otherwise would have been.

2. BOONIN'S APPROACH

P1 is a complex premise that incorporates the fundamental observations that motivate Parfit's non-identity problem. It is therefore worth making a few observations about Boonin's formulation of the premise and how his formulation shapes his analysis.

Negating a Negative vs. Proving a Positive

Because of its negative construction (i.e. "Wilma's act…does not make Pebbles worse off…"), there are two different cases in which P1 would be true: Wilma's act is unrelated to Pebbles' well-being (the weaker case) or Wilma's act has an effect on Pebbles' well-being and that effect is not bad (the stronger case). Although not required to prove that P1 is false, a true solution to the non-identity problem (one that meets what Boonin calls the "Robustness Requirement" by offering a reason to reject the Implausible Conclusion) needs to demonstrate that P1 is false in both of these cases. As a result, the threshold necessary for "solving" the non-identity problem is somewhat higher than proving that Boonin's five-premise argument is unsound. It may therefore be helpful to reformulate a contradictory version of P1 to use in analyzing the various objections as a test for whether they indicate P1 is false in the stronger case. I will call this contradictory premise ~P1.

~P1: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving makes Pebbles worse off than she otherwise would have been.

If a justification for rejecting P1 also provides grounds for accepting ~P1, it is sufficiently strong to explain the intuition that Wilma acts wrongly in conceiving Pebbles. Further, if there is a sound argument to demonstrate the truth of ~P1—that Wilma's conception choice makes Pebbles worse off than she otherwise would have been—we can clearly and robustly reject P1. I will therefore use ~P1 as a test premise throughout my discussion.

Non-Identity vs. a Life Worth Living

Looking more closely at P1, it becomes apparent that its truth turns on two separate ideas: 1) that Pebbles and Rocks are not the same person and 2) that Pebbles is not worse off than she would have been because her life is worth living despite her blindness. Breaking P1 into two sub-premises can therefore offer conceptual clarity about the relevant objections and rebuttals (Malek 2017). These two premises can be described as follows for

the case of Wilma:

> *The Non-Identity Premise:* If Wilma had conceived two months later than she actually did, a different child would have been brought into being and Pebbles never would have existed.
>
> *The Life Worth Living Premise:* Pebbles is not made worse off than she otherwise would have been by Wilma's decision to conceive if her life is worth living.

Of the five objections Boonin addresses in his analysis of P1, three (the Incoherence, Equivocation, and Asymmetry Objections) relate to the Life Worth Living Premise and two (the *De Re*/*De Dicto* and Metaphysical Objections) concern the Non-Identity Premise.

Although proving either sub-premise false would be sufficient to undermine the soundness of the Implausible Conclusion, objections to the Life Worth Living Premise have different explanatory powers than those that address the Non-Identity Premise because of a difference in the possible worlds they compare. The Life Worth Living Premise compares two possible worlds: one in which an individual exists and another in which that individual does not exist. Rebutting this premise alone cannot therefore offer an adequate solution to the non-identity problem, which compares two possible worlds, each with one individual who exists. The Non-Identity Premise, in contrast, compares two possible worlds each with one existing individual and so has the potential to offer a sufficiently robust solution. As the following analysis will show, the *De Re*/*De Dicto* and Metaphysical Objections are therefore more promising starting points for an argument proving ~P1.

Same Number Problems vs. Different Number Problems

Relatedly, it is worth noting that most of Boonin's analysis focuses on Same Number Problems—those in which the same number of people exist in various possible worlds—as opposed to Different Number Problems—those in which an individual's decision changes the number of people brought into being. Wilma's case is a paradigmatic Same Number Problem. Boonin explains this focus by claiming that "since a solution to the non-identity problem cannot succeed in different number problems if it cannot succeed in the simpler same number cases, I will not explicitly discuss different number cases." (8) The decision to focus on Same Number Problems may be reasonable for the sake of conceptual clarity, but Boonin's justification for taking this approach is not convincing. It is far from obvious that Same Number Problems are any simpler than Different

Number Problems. They may in fact be more complex because of the challenges associated with comparing possible worlds in which genetically distinct individuals are brought into existence rather than comparing the existence or non-existence of one individual. Furthermore, it is entirely possible that a solution to Different Number Problems is independent of a solution to Same Number Problems.

Regardless of whether Boonin's focus is justifiable, awareness of that focus will be useful for analyzing his arguments. Same Number and Different Number Problems raise fundamentally different ethical questions. The former asks whether someone acts wrongly by choosing to conceive in the way they did while the latter addresses whether someone acts wrongly by choosing to conceive at all. Comparing two possible worlds in which the same number of people exist requires different tools than comparing two worlds in which different numbers of people exist. A full account of these differences is beyond the scope of this paper, but some key considerations will become evident in what follows.

3. OBJECTIONS CONSIDERED

Boonin considers and rebuts five different objections that have been levelled against P1, arguing that if these objections fail we have good evidence for the truth of the claim that Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she otherwise would have been. The following sections critique Boonin's responses to these five arguments. His analyses of three of the objections (those relating to the Life Worth Living Premise) are relatively convincing in demonstrating that these objections are not sufficient to solve the non-identity problem. Boonin's responses to the other two objections (those relating to the Non-Identity Premise), however, do not adequately refute those objections, leaving open plausible accounts of how Wilma's act of conceiving now makes Pebbles worse off than she otherwise would have been (~P1).

Incoherence, Equivocation, and Asymmetry Objections

As described by Boonin, the Incoherence Objection holds that in order for P1 to be true, we must be able to coherently compare Pebbles' level of well-being if she is conceived now to her state of well-being if she is not conceived in order to sensibly affirm that she is not worse off than she otherwise would have been. Proponents of this objection argue that existence and non-existence are so fundamentally different that we cannot coherently

compare the two, making it impossible to claim that Pebbles is not worse off than she otherwise would have been. It therefore clearly targets the Life Worth Living Premise by rejecting this possibility. However, as Boonin argues, this objection only refutes P1 in a weak sense because it equally undermines any attempt to show that Wilma *does*, in fact, make Pebbles worse off by conceiving now rather than waiting two months. As noted above, any true resolution to the non-identity problem must offer support for ~P1 and the Incoherence Objection fails this Robustness test.

The Equivocation Objection also targets the Life Worth Living Premise, that is, the idea that a person whose life is worth living cannot be worse off than he would have been if he had not existed. It clarifies that the phrase "a life worth living" has at least two possible meanings: a life worth continuing or a life worth conceiving. When used in reference to someone who exists, the phrase "a life worth living" means "a life worth continuing" – that is, a life that is good enough that the person living it is better off alive than dead. In contrast, when the same phrase is used in reference to a possible person who does not yet exist, it means "a life worth conceiving" – that is, a life that is good enough to justify creating it. It has been argued that these two standards describe different levels of well-being and that a life worth creating is a higher standard than a life worth continuing (Benatar 2000, Malek 2017). As a result of this difference, there are some people – those living lives that are just marginally worth continuing but not worth conceiving – who are, in fact, made worse off than they otherwise would have been by being conceived. Even if there are different thresholds for lives worth conceiving and lives worth continuing, Boonin argues, the Equivocation Objection would only be successful in resolving the non-identity problem in a small subset of cases – those in which Pebbles' life is only marginally worth living. It, too, therefore fails the Robustness Requirement.

The Asymmetry Objection is derived from David Benatar's observation that there is an intuitive asymmetry between how we value the absence of pain and the absence of pleasure (Benatar 2006). As summarized by Boonin, "the absence of pain is better than the presence of pain, even if there is nobody for whom the absence of pain is a benefit" and "the absence of pleasure is worse than the presence of pleasure only if there is somebody for whom this absence is a deprivation." (47) Boonin illustrates this asymmetry by contrasting judgments about a "Cursed Couple" (who are guaranteed to have a child who only experiences pain) and a "Blessed Couple" (who are guaranteed to have a child who only experiences pleasure), each of whom decides not to conceive. The intuitive response to this thought experiment is that the Cursed Couple's decision results in a

relatively better state of affairs by not conceiving while the Blessed Couple's decision has not resulted in a worse state of affairs. In other words, there is a common intuition that when a possible person is not brought into being, there is value in the fact that pain is not experienced but there is no disvalue in the fact that pleasure is not experienced.

Benatar's Asymmetry Objection therefore categorically refutes the Life Worth Living sub-premise. That is, it not only claims that Pebbles *may* have been made worse off than she otherwise would have been by Wilma's decision to conceive her despite the fact that her life is worth living, but argues that she—and every other person ever conceived—has been made worse off by her own conception. Boonin plausibly argues that Benatar's Asymmetry Objection fails to offer a resolution to the non-identity problem. If we take Benatar's argument seriously, it would show why Wilma does something morally wrong by conceiving *either* now or later, which Boonin points out is problematic in that it does not offer an analysis of Wilma's choice to conceive that anyone finds intuitive. To be fair, this is the provocative thesis embraced by Benatar. The implausibility of the conclusion is therefore not a reason Benatar would find sufficiently convincing to abandon his argument.

Boonin's analyses of these three objections reveal a fundamental conceptual flaw in his account. While his arguments about their limitations are correct, there is a sense in which they tear down a straw man. These objections are intended to resolve Different Number problems, not Same Number Problems. They address the question of whether a life is worth living, therefore comparing existence and non-existence, which is *not* the question at the heart of the non-identity problem as described in Wilma's case. Boonin therefore tackles three arguments that are conceptually irrelevant to the problem he wants to solve. Given that these objections are not intended to offer a solution to Same Number Problems, successful rebuttal of them does not constitute evidence for the truth of P1 or give reason to embrace the Implausible Conclusion. It will therefore be necessary to turn to the other two objections, those targeting the Non-Identity Premise, to determine whether Boonin's rebuttals of those objections are sufficiently convincing to erase doubts about the truth of P1.

The De Re/De Dicto Objection

The idea behind the *De Re/De Dicto* Objection is that there are at least two different ways of thinking about things in the world. Conceptualizing something *de dicto*, meaning "from what is said", focuses on the properties of a thing in the world (e.g. Wilma's child) whereas *de re*, meaning "about the thing" (OED), refers to the particular thing that has those properties

(e.g. Pebbles). Boonin follows Hare (2007) in arguing that moral claims can be made using both *de re* and *de dicto* formulations. The objection holds that it is possible to coherently claim that Wilma's child is worse off if Wilma conceives now rather than waiting two months because the child she conceives now is, in a *de dicto* sense, identical to the child she would conceive two months from now. It therefore responds to the Non-Identity Premise by offering an account that establishes continuity of identity between Pebbles and Rocks.

I have defended this objection (Malek 2017) using the similar construct of rigid and non-rigid designators described by Saul Kripke. Kripke (1972) suggests that it is appropriate to "call something a rigid designator if in every possible world it designates the same object, a non-rigid or accidental designator if that is not the case." I prefer using Kripke's distinction to the *de re*/*de dicto* one in formulating the objection because rigid and non-rigid designation are both semantic tools for referencing an entity. In contrast, *de re* is used to make metaphysical claims while *de dicto* makes semantic ones. The rigid/non-rigid distinction therefore seems to be a better tool for building analogies and describing differences. Further, it explicitly is set in the context of possible worlds, which is more straightforwardly applicable to the non-identity problem. Nonetheless, the two formulations of the objection are derived from the same core idea that an entity may be identified in two different ways.

Parfit dismisses the move as a "verbal trick" (1984, p397), but others (Hare 2007, Malek 2017) have argued that there are compelling reasons to use the *de dicto*/non-rigid approach when constructing moral claims about future children. Prior to conception, it is impossible to use *de re*/rigid concepts to refer to a future child because that individual does not yet exist. As a result, the only way to refer to future entities is in terms of the properties they will have, such as their relationships with other people (e.g. Wilma's child). The *de dicto*/non-rigid approach also can be used to make coherent claims that span a period of time during which a child is conceived (e.g. "When I graduated from college I promised my future children that I would cover their college tuition, and I have fulfilled that promise.") Further, *de dicto*/non-rigid designation of a child better captures the way parents think about their future children: the infinite variety of genetic identities does not prevent them from seeing "their child" as an object of moral consideration.

Boonin offers an in-depth critique of this approach using Hare's example of Officer Tess, who is charged with the task of creating policies related to automobile safety features in her community. Hare argues that Tess has *de dicto*/non-rigid obligations to her community's accident

victims. Even if different members of the community are involved in accidents as a result of her policies, Tess's duty is to promote the well-being of whoever experiences them. In the same way, Hare argues, Wilma has an obligation to promote the well-being of her future child *de dicto*/non-rigidly designated. Boonin rebuts this argument by noting that there are two different interpretations of this obligation, both of which are problematic: (1) Tess's obligation is to create a policy under which her accident victims (*de dicto*/non-rigidly designated) are as well off as possible and that obligation could be discharged by creating a policy under which accidents are redirected to happen to the most well-off members of the community (e.g. Jack), so that even though their well-being is compromised they are still better off than those who would have been accident victims under an alternative policy (e.g. Jill), which is absurd; or (2) Tess's obligation is to create a policy that reduces the harm suffered by accident victims in her community, an obligation that cannot apply to Wilma's case "because neither of Wilma's choices will cause any harm at all to the particular child that she conceives" (36). Boonin claims that this two-pronged analysis demonstrates that Hare's scenario fails to show that a *de dicto*/non-rigid approach resolves the non-identity problem.

Neither of Boonin's rebuttals are persuasive in my view. First, while the idea that Tess could fulfill her obligation by redirecting accidents from Jill to Jack is clearly absurd, this interpretation is a straw man due to important differences between Tess and Wilma's situations. Tess is tasked with making decisions among possible worlds in which both Jack and Jill exist in a *de re*/rigid sense and have established levels of well-being. In contrast, Wilma must compare one possible world in which Pebbles exists but Rocks does not to another possible world in which Rocks exists and Pebbles does not. At the point a decision is made, Pebbles and Rocks cannot be referred to in a *de re*/rigid sense and have no established level of well-being. Intuitions about the absurdity of Tess implementing a redirection policy are generated by the fact that Tess is not fulfilling the spirit of her obligation to protect her community but is instead using an alternative approach that is available because there are two *de re*/rigidly designated people affected by her decision. In contrast, by waiting to conceive, Wilma is fulfilling the spirit of her obligation in the only way that is available to her. Boonin's rebuttal therefore exploits a disanalogy in the example rather than demonstrating the inadequacy of Hare's primary point about the possible existence of *de dicto*/non-rigidly formulated obligations.

Perhaps more interestingly, Boonin's claim that Hare's example does not apply to Wilma's case under the harm reduction interpretation is based on circular reasoning. The statement that Wilma's choice to conceive Pebbles will not result in harm to the child she brings into the world relies

on a *de re*/rigid designation of that particular child. If, instead, a *de dicto*/non-rigid designation is used, Wilma's child may be made worse off as a result of her decision and may therefore be harmed. The argument therefore only goes through if its conclusion is assumed: that Pebbles and Rocks should be referred to as two different possible children.

Boonin's next step is to hypothetically grant that this objection does show that Wilma's child is made worse off but then demonstrate that it generates even more problematic results than the Implausible Conclusion. In doing so, he develops a scenario in which he is a newly-minted physician who has moved to a new town with an overwhelming demand for medical care and is faced with the question of who to accept as patients. According to Boonin, "Anyone I accept as a patient will be *de re* better off as a result of my taking them on, and so de re considerations can't tell me whom to accept as a patient…In deciding which people to take on as patients, that is, I should be guided by the *de dicto* concern for what would be best for the health of 'my patients.' And this means that I must choose the healthier people to accept as patients" (37-8). This unintuitive result, he argues, demonstrates that *de dicto*/non-rigid formulations of our moral claims cause problems that are worse than those generated by non-identity.

As with Tess's case, the in-demand doctor scenario has a fundamentally different setup than Wilma's. Creating a pool of patients—selecting among people who already exist—and creating a new person—selecting among possible people who do not yet exist—are not the same type of decision. The use of *de dicto*/non-rigid designation in the former case may have results that are not applicable to situations like the latter, undermining the effectiveness of Boonin's argument. Unfortunately, developing analogies that do not suffer from this problem are impossible due to the unique nature of procreation.

If, as I've argued, Boonin's rebuttal of the *De Re*/*De Dicto* Objection is unsuccessful, it is possible that the use of *de dicto*/non-rigid designation offers a genuine solution to the non-identity problem. This approach offers a way to formulate coherent moral claims that creates continuity of identity among possible future children. It therefore demonstrates that there is at least one morally relevant sense in which Pebbles is made worse off by Wilma's decision to conceive her immediately, and thereby offers support for ~P1.

The Metaphysical Objection

While the *De Re*/*De Dicto* Objection concerns the way we designate or refer to entities in the world, the Metaphysical Objection concerns the identity

of those entities. The Metaphysical Objection could therefore be considered a stronger version of the *De Re*/*De Dicto* Objection. It goes beyond the stance that we can make coherent moral claims involving continuity of identity among possible people to the position that those various possible people actually are the same person (who therefore is made worse off by Wilma's choice).

The intuition being tested by the Metaphysical Objection is whether an individual's genetic makeup is the only morally relevant determinant of his identity – whether he would have existed if his parents had not conceived that individual exactly when and how they did. While there is a sense in which the answer to this question is "no" – because a different egg and sperm would have united and therefore a child with a different genetic identity would have been created – there is also a sense in which the answer to the question is "yes" – because that individual fills the same role and has the same set of relationships regardless of his genetic identity. Continuity of identity among different possible worlds can be established if identity is understood in this latter sense.

Boonin rebuts this objection by offering examples of cases in which this "solution" to the non-identity problem causes more problems than it resolves. First, if Wilma decides not to conceive Pebbles but ends up having twins because she waited to conceive, a non-genetic paradigm for identity would entail that Pebbles is identical to both of the resulting twins, a result Boonin claims is less plausible than the Implausible Conclusion because that would imply that the twins are identical to each other. A second scenario Boonin uses (following Parsons 2003) to demonstrate the problem with the metaphysical objection involves Wilma waiting a year (rather than two months) to conceive. As in the original case, the metaphysical objection would say that Rocks is identical to Pebbles and so Wilma makes Pebbles worse off than she otherwise would have been by conceiving her. But, Boonin argues, in this version Wilma could theoretically also conceive Rocks and, if she does, the objection would entail that that Rocks and Pebbles are identical even if they both exist.

The effectiveness of Boonin's refutation of this objection hinges on commitments about the degree to which a person's essence is "fragile" (Hare 2009). Essence is perfectly fragile on a view in which any possible difference between two people in different possible worlds breaks the continuity of identity between those two people. Essence is robust if substantial differences between those two people can be tolerated while identity is maintained. Several philosophers have explored aspects and implications of different accounts of personal essence that would, if defensible, offer robust solutions to the non-identity problem (Hare 2009,

Mackie 2006, Wrigley 2012). Defending a detailed account of exactly what continuity of identity across possible worlds requires is beyond the scope of this paper, though for present purposes it is sufficient to consider whether the two scenarios above would be devastating to a plausible account.

Both of the scenarios Boonin uses to refute the Metaphysical Objection are Different Number Cases. As discussed above, the tools required to analyze Different Number Cases are significantly different from those used for Same Number cases. As a result, it seems plausible to admit that continuity of identity is disrupted in Different Number Cases even though it may be preserved in Same Number Cases. Further, depending on the criteria established by a theory of personal essence, conception at a substantially different time or under significantly different circumstances could have a similar effect. But the fact that continuity of identity is disrupted under some circumstances does not entail that it is not preserved under other, more ordinary circumstances, such as those facing Wilma. Boonin's argument therefore demonstrates that there are limits to how robust personal essence may be, but does not foreclose the possibility that it may sometimes survive across different possible worlds.

Like the *De Re*/*De Dicto* Objection, then, the Metaphysical Objection resolves Same Number Non-identity Problems by offering an alternative account to our traditional ways of thinking about future people that creates continuity of identity among Wilma's possible children. In the context of comparing different possible worlds, where identity is a concept rather than something that can be physically demonstrated, the distinction between these two objections shrinks. In both cases, the claim can be made that there is a morally relevant sense in which there is continuity of identity between Pebbles and Rocks. If this continuity exists, Wilma makes her child worse off by not waiting to conceive and demonstrates that P1 can be true under some circumstances.

4. CONCLUSION

David Boonin's The Non-Identity Problem and the Ethics of Future People takes a sensible approach to a difficult problem and defends an unpopular conclusion. Boonin's arguments are clear, thoughtful, and thorough, offering helpful examples and creative (but not overly outlandish) thought experiments. Although I appreciate Boonin's analysis of the objections to the claim that Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she otherwise would have been, I am not convinced that all of these

objections have been shown to be misguided and that we therefore must accept the Implausible Conclusion.

Although distinct types of claims, both the *De Re*/*De Dicto* and the Metaphysical Objections develop frameworks for creating the continuity of identity among possible future people necessary for resolving the non-identity problem. Whether using a semantic argument about how moral claims should be constructed or a metaphysical argument about how future human beings are defined, these objections offer plausible alternative views on how future children should be conceptualized. As a result, these alternative perspectives change the moral claims that can be made about them. Specifically, they show that it is not always incoherent to state that someone has been made worse off as a result of the method of their conception.

These two objections allow us to affirm ~P1, that Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving makes Pebbles worse off than she otherwise would have been if we talk about Pebbles in a *de dicto*/non-rigidly designated way or if we use traits beyond her genetic makeup to define her personal identity. It is worth noting that this alternative approach is not the only sense in which people can be understood and referred to, but it is still a morally relevant one.

I suspect that Boonin might charge my view with violating his "Independence Requirement" in that its primary support comes from the fact that it offers a way to avoid the Implausible Conclusion. To meet this requirement, the resolution must be able to provide insight to problems arising in other contexts as well. However, I would respond that bringing a person into existence is a fundamentally different type of thing than what happens in any of the other examples he offers. True analogies to creating a human being are difficult if not impossible to construct. As others have argued (Heyd 1992), the ethics of conception may therefore require the development of tools that are fundamentally different from those used to analyze other sorts of questions. Just as an electron but not a stone can exist in two places at one time, so can a future child but not an existing one be harmed by certain types of decisions. This exceptionalist approach may seem indefensible but finds support in the widespread recognition that making a person is a truly exceptional thing to do.

BIBLIOGRAPHY

Benatar, D., 2000: The wrong of wrongful life. *American Philosophical Quarterly* 37: 175-83.

— 2006: *Better never to have been: the harm of coming into existence, New York*: Oxford University Press.

Boonin, D., 2014: The non-identity problem and the ethics of future people, New York: Oxford University Press.

Hare, C., 2007: Voices from another world: must we respect the interests of people who do not, and never will, exist? *Ethics* 117:498-523.

— 2009: The ethics of morphing. *Philosophical Studies* 145:111-130.

Heyd, D. 1992. *Genethics*, Berkeley: University of California Press.

Kripke, S., 1972: *Naming and necessity*, Cambridge, MA: Harvard University Press.

Mackie, P. 2006: *How things might have been: individuals, kinds and essential properties*. New York: Oxford University Press.

Malek, J., 2017: The possibility of being harmed by one's own conception, in *The Oxford handbook of reproductive ethics*, ed. L. Francis, 571-91 New York, New York:Oxford University Press.

— 2006: Identity, harm and the ethics of reproductive technology, *Journal of medicine and philosophy* 31:83-95.

Oxford English Dictionary: https://en.oxforddictionaries.com/definition/de_re Accessed 1 July, 2018.

Oxford English Dictionary: https://en.oxforddictionaries.com/definition/de_dicto Accessed 1 July, 2018.

Parfit, D., 1984: *Reasons and persons*, New York: Oxford University Press.

Parsons, J. 2003: Why the handicapped child case is hard. *Philosophical Studies* 112: 147-62.

Wrigley, A.J., 2012: Harm to future persons: non-identity problems and counterpart solutions, *Ethical Theory and Moral Practice* 15:175-190.

DOI: 10.31009/LEAP.2019.V7.03

David Boonin on the Non-Identity Argument: Rejecting the Second Premise[1]

MOLLY GARDNER
Bowling Green State University

ABSTRACT

According to various "harm-based" approaches to the non-identity problem, an action that brings a particular child into existence can also harm that child, even if his or her life is worth living. In the third chapter of *The Non-Identity Problem and the Ethics of Future People*, David Boonin surveys a variety of harm-based approaches and argues that none of them are successful. In this paper I argue that his objections to these various approaches do not impugn a harm-based approach that Boonin does not consider, an approach I call the "existence solution to the non-identity problem". I also argue that the existence solution is more plausible than Boonin's own proposed solution.

Keywords: non-identity problem, David Boonin, harm-based approach, existence solution, causal selection, mere condition, backtracking counterfactuals.

1. INTRODUCTION

One strategy for solving the non-identity problem is to hold that in some cases, an action that brings a particular child into existence can also harm him or her, even if the child's life is worth living. This approach, which I will call the "harm-based" approach, is the focus of the third chapter of David Boonin's (2014) book, *The Non-Identity Problem and the Ethics of Future People*. In the chapter, Boonin considers versions of the harm-based approach developed by Matthew Hanser (2008, 2009), Elizabeth Harman (2004, 2009), Seana Shiffrin (1999), and others, and he argues that either these views fail to deliver a sufficiently robust solution to the non-identity problem or else they lack plausibility (which Boonin refers to as "modesty").

1 For helpful feedback on this paper, I thank Eden Lin and two anonymous referees for this journal.

He concludes that *no* version of the harm-based approach will succeed, for the weaknesses of the views he criticizes will afflict all other versions of the approach.

Has Boonin really identified fatal flaws in the views he considers, and do these flaws generalize to all other versions of the harm-based approach? I agree that some of Boonin's objections are persuasive: Boonin has, indeed, shown that some versions of the harm-based approach are either immodest or insufficiently robust. However, I disagree with Boonin's claim that these objections generalize to *all* versions of the harm-based approach. In what follows, I will argue that a particular version of the harm-based approach that Boonin does not consider—one that features the *existence account of harming* [2]—does, in fact, deliver a robust solution to the non-identity problem. Moreover, the costs of what I will call "the existence solution" are minimal: it requires less bullet-biting than other versions of the harm-based approach, and indeed, less bullet-biting than Boonin's own view.

To make my case, I will first provide a more precise characterization of both the non-identity problem and the harm-based approach to solving it. I will then explain what Boonin means when he objects to other versions of the harm-based approach on the grounds of either immodesty or a lack of robustness. In the fourth section, I will explain both the existence account of harming and the existence solution that it supports, and I will argue that the existence solution is neither insufficiently robust nor immodest. Finally, I will argue that the implications of Boonin's own solution to the non-identity problem are less plausible, on balance, than are those of the existence solution.

2. THE NON-IDENTITY PROBLEM AND THE HARM-BASED APPROACH

The non-identity problem arises in cases where the existence of an individual whose life is worth living depends counterfactually upon an action that appears to be objectionable in virtue of some negative consequence the action has for the individual's well-being. To illustrate the problem, Boonin constructs a case in which a doctor tells a woman named Wilma that if she conceives a child now, the child will be blind, but if she takes a pill once a day for two months and conceives after that, her child's vision will be normal. Importantly, the blind child who would be born now

2 I develop the existence account of harming more fully in Gardner (2015) and Gardner (2017).

("Pebbles") is not identical to the child who would be born two months from now ("Rocks")—hence the term "non-identity".[3] The doctor tells Wilma that the pills are easy to take and have no side effects. Nevertheless, Wilma decides against taking the pills, and she conceives Pebbles instead of Rocks. Pebbles' blindness is a significant detriment to her quality of life, but on the whole, her life is worth living.

To see how this case generates a philosophical problem, consider the tension between various intuitions people might have about the case. On the one hand, Pebbles is worse off than Rocks would have been had Wilma acted differently, and this consideration might suggest that there was something wrong with Wilma's action. But on the other hand, Pebbles is no worse off than Pebbles would have been had Wilma acted differently; after all, if Wilma had not conceived now, *Pebbles* would not have been at all. This second consideration lends support to the notion that Wilma's action was unobjectionable. Boonin attempts to capture the tension between these considerations with what he calls the "non-identity argument", an argument whose five premises are highly plausible in themselves, but whose conclusion is so counterintuitive that Boonin refers to it as "the Implausible Conclusion":

(P1) Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been.

(P2) If A's act harms B, then A's act makes B worse off than B would otherwise have been.

(P3) Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not harm anyone other than Pebbles.

(P4) If an act does not harm anyone, then the act does not wrong anyone.

(P5) If an act does not wrong anyone, then the act is not morally wrong.

Therefore,

(C) Wilma's act of conceiving Pebbles is not morally wrong.

As Boonin notes, to solve the non-identity problem, we must either reject at least one of the plausible premises or accept the Implausible Conclusion. In the various chapters of his book, Boonin considers attempts to reject each of the five premises. He concludes that none of these attempts succeed, and he ends the book by endorsing the Implausible Conclusion.

3 The term "non-identity problem" was coined by Derek Parfit (1984) to describe a problem that appears to have been independently discovered by Parfit (1976), Thomas Schwartz (1978), and Robert Adams (1979).

For my purposes, the relevant chapter is the third one, in which Boonin criticizes various attempts to reject P2. What I'm calling the "harm-based" approach to the non-identity problem can be understood as any approach that tries to solve the non-identity problem by rejecting P2. To reject P2 is to maintain that an action can harm someone even if it doesn't make the victim worse off than he or she would otherwise have been. Those who reject P2 usually appeal to some set of harm-based principles or some account of harming in order to justify their view. In Section 3, I will defend the account of harming that I believe can *best* justify the rejection of P2, namely, the existence account. First, however, I will review the two main objections that Boonin raises against other versions of the harm-based approach.

3. ROBUSTNESS AND MODESTY

In organizing his objections to other versions of the harm-based approach, Boonin appeals to what he calls "the robustness requirement" and "the modesty requirement": two requirements that a successful solution to the non-identity problem must meet. He claims that all versions of the harm-based approach violate either the robustness requirement, the modesty requirement, or both; this is why he believes that no version of the harm-based approach will successfully solve the non-identity problem. In this section, I will explain Boonin's account of each requirement in turn.

According to Boonin, a principle that purports to solve the non-identity problem by rejecting a premise of the non-identity argument meets the *robustness* requirement only if it is "strong enough to warrant rejecting any weakened version of the premise that would still be strong enough to generate the unqualified version of the Implausible Conclusion" (22). There are two things to note about this account of robustness. First, the phrase, "unqualified version of the Implausible Conclusion" is supposed to refer to the conclusion that Wilma's act of conceiving Pebbles—in the case as originally described—is not morally wrong. A *qualified* version of that conclusion would involve adding new facts to the case, such as the fact that by conceiving Pebbles, Wilma violates a just law or breaks a promise. If we added the latter fact to the case, the qualified conclusion would be that Wilma's act of conceiving Pebbles *and thereby breaking* a *promise* is not morally wrong. It is easy to undermine such a qualified conclusion, but as Boonin points out, doing so does not really help us solve the non-identity problem.

The second thing to note about Boonin's account of robustness is his reference to a "weakened version of the premise". Recall that premise two

of the non-identity argument says that if A's act harms B, then A's act makes B worse off than B would otherwise have been. Here is a weakened version of that premise:

(W2) If Wilma's act (in the case as originally described) harms Pebbles, then Wilma's act makes Pebbles worse off than Pebbles would otherwise have been.

Boonin's point is that it might be possible to find counterexamples to P2 that would not be counterexamples to W2. But since W2 is still strong enough, on its own, to generate the unqualified Implausible Conclusion, a counterexample to P2 that did not undermine W2 would not really solve the non-identity problem. Thus, a harm-based solution to the non-identity problem satisfies the robustness requirement if and only if it undermines W2.

In addition to the robustness requirement, Boonin's objections to the harm-based views also appeal to what he calls "the modesty requirement". According to Boonin, a proposed solution to the non-identity problem will fail the modesty requirement if its implications are less plausible than the Implausible Conclusion. He writes, "The modesty requirement constitutes a reasonable constraint on a satisfactory solution to the non-identity problem because considerations about intuitive plausibility are precisely what give rise to the problem in the first place" (23). Thus, to meet the modesty requirement, a solution based on the existence account of harming (or any other view) must have implications that are more plausible than the claim that Wilma's act of conceiving Pebbles is not morally wrong.

As I noted above, Boonin's main objections to various versions of the harm-based approach are that each version violates either the robustness requirement, the modesty requirement, or both. Thus, in order to show that the existence solution to the non-identity problem is impervious to Boonin's main objections, I will need to show that it satisfies both requirements. In the next section, I will explain the existence solution to the non-identity problem and I will argue that it does, indeed, satisfy both.

4. SOLVING THE NON-IDENTITY PROBLEM WITH THE EXISTENCE ACCOUNT OF HARMING

I will introduce the existence account of harming by contrasting it with the second premise of the non-identity argument. Recall that according to P2, if A's act harms B, then A's act makes B worse off than B would otherwise have been. When we interpret this principle, it is natural to construe the word 'makes' as a causal verb: another way of stating P2 is to say that a

harmful action always *causes* the victim to be worse off than he or she otherwise would have been. And indeed, Boonin often uses causal language and counterfactual language interchangeably. Thus, he seems committed to the view that the following sentence implies that A causes B:

(D) If A hadn't happened, then B wouldn't have happened.

However, the existence solution expressly rejects the notion that the D states a sufficient condition for causation. According to the existence solution, causation may be reducible to counterfactual dependence, but not in a way that makes the relation expressed by D sufficient for causation: the relationship between causation and counterfactual dependence is more complicated, and the two concepts play distinct roles in the solution to the non-identity problem.[4] Here is the full existence account of harming:

Harming (def.): An event, E, harms an individual, S, if and only if E causes a state of affairs that is a harm for S.

Harm (def.): A state of affairs, T, is a harm for an individual, S, if and only if

(i) There is an essential component of T that is a condition with respect to which S can be intrinsically better or worse off; and

(ii) If S existed and T had not obtained, then S would be better off with respect to that condition. [5]

This account has at least three important features. First, it is compatible with the rejection of P2 in the non-identity argument. The existence account of harming does not imply that an action harms a victim only if the action makes the victim worse off than he or she would have been, had the *action* not been performed. Instead, it implies that an action harms a victim only if it makes the victim worse off than he or she would have been, had the victim existed without the upshot of the action. The existence account of harming thereby locates harm in the *effects* of various actions, and not in the *causes* of those effects.[6]

Second, to assess whether an effect qualifies as a harm, the account does not ask us to compare a world in which the victim suffers the alleged harm to a possible world in which the victim does not exist at all. Instead, the account asks us to compare the world containing both the victim and

4 I say more about the relationship between counterfactual dependence and causation in Gardner (forthcoming).

5 This formulation of the account comes from Gardner (2015).

6 If the reader is still uncertain what this means, don't despair. Very shortly, I will illustrate how the view works by applying it to the case of Wilma and Pebbles. For more on the advantages of an "effect-relative" account of harming over an "action-relative" account, see Gardner (2017).

the alleged harm to a world where the victim exists without the alleged harm; we then check to see whether, in the latter world, the victim is better off. Such a comparison is legitimate because it is metaphysically possible for someone to exist without, say, being blind, even if, as a matter of fact, she *would* not have existed if the cause of her blindness hadn't obtained. Even if your formative years are essential to your identity, such that you could not have existed if you had not been born blind, it is still metaphysically possible to *become* sighted at the age of, say, 15. So there is a possible world where you exist and are sighted, and we can compare your well-being in that possible world to your well-being in the world in which you are blind. Such a comparison is appropriate for ascertaining whether a state of affairs qualifies as a harm because a harm is the sort of thing you would be better off without. If we don't consider the possible worlds where you exist in the absence of the alleged harm, it is impossible to determine whether you would be better off without it.[7]

A third important feature of the existence account of harming is that it yields a solution to the non-identity problem that is impervious to the two main objections that Boonin raises against other versions of the harm-based approach. To show that this is so, I will explain how the existence solution satisfies both the robustness requirement and the modesty requirement, in turn.

Recall that a solution to the non-identity problem satisfies the robustness requirement if and only if it undermines W2, the claim that if Wilma's act (in the case as originally described) harms Pebbles, then Wilma's act makes Pebbles worse off than Pebbles would otherwise have been. Thus, I need to show that the existence account of harming implies, not only that an act can harm someone without making the victim worse off, but that *Wilma's* act, in the case as originally described, harms Pebbles without making her worse off. I will show this by arguing that (a) there is a state of affairs that instantiates a comparatively poor showing along some intrinsic dimension of Pebbles' well-being and (b) Wilma's act causes that state of affairs.

To see why there is a state of affairs that instantiates a comparatively poor showing along some intrinsic dimension of Pebbles' well-being, recall the original description of the case involving Wilma and Pebbles. I noted there that Pebbles' blindness is a significant detriment to her quality of life. This stipulation about the case (which is also featured in Boonin's description of the case) means that for Pebbles, things are going comparatively worse along some dimension of her well-being. (If her well-being has only one dimension, then things are going comparatively worse

7 For more on why we should build the assumption that S exists into the antecedent of the relevant counterfactual, see Gardner (2015: 436 – 437).

for her, full stop.) What, exactly, is measured by that dimension of well-being depends on which theory of well-being is correct; it may be that Pebbles is doing poorly with respect to her capabilities, her happiness, the satisfaction of her desires, or some other objective list of goods. Indeed, to more precisely describe the state of affairs that constitutes the harm for Pebbles, we would need to know which theory of well-being is correct. For example, if an objective list theory were true and if vision were on the objective list of goods, then the harm for Pebbles would be *the state of affairs in which she is blind*. But we do not need to describe the state of affairs so precisely in order to know that there is such a state of affairs. All we need to know is that there is some well-being-related state of affairs such that, if Pebbles existed and that state of affairs had not obtained, Pebbles would be faring better. And since we do know this, conditions (i) and (ii) in the above definition of harm are both satisfied.

We have so far established that Pebbles suffers a harm. The next question is whether Wilma's act causes that harm. Does it? Someone might worry that we need to describe the state of affairs that is the harm more precisely in order to know whether Wilma's act causes that harm. However, this is not so: in the case as described, the harmful state of affairs is closely connected to (if not *identical* to) the state of affairs in which Pebbles is blind. Therefore, if Wilma's act causes the state of affairs in which Pebbles is blind, then *by causing* that state of affairs, her act also causes the state of affairs that is a harm for Pebbles. Thus, in order to show that Wilma's act harms Pebbles, we need only to show that Wilma's act causes the state of affairs in which Pebbles is blind.

Notice that a paradigmatic way of causing a state of affairs in which someone is blind is to blind someone after they have already been born. For example, suppose that Ted throws acid in Fred's eyes, and from that point on Fred is blind. This is a clear case in which (a) Fred suffers a harm and (b) Ted caused that harm.

Wilma's act is different from Ted's. Instead of inducing a certain biological property in an individual who already exists, Wilma's act *brings into existence* an individual with that property. However, this difference is not as significant as it might initially appear to be. That is because the existence account of harming does not say that harming is a matter of causing someone to have a certain property. Instead, it says that harming is a matter of causing a particular state of affairs. And although the case of Ted and Fred shows that causing someone *to have* a certain property is *one* way of causing a particular state of affairs, it is not the only way. There are, in other words, at least two ways to cause the harm of blindness: you can cause someone who already exists to be blind, or you can cause someone

to exist who *will* be blind. Wilma causes the existence of Pebbles, who will be blind, and so she causes a state of affairs that is a harm for Pebbles. This means that the definition of harming provided by the existence account of harming is also satisfied: Wilma's act harms Pebbles. Because the existence account of harming yields this result, Boonin's robustness requirement is satisfied.

What about the modesty requirement? Recall that to meet the modesty requirement, a solution to the non-identity problem must not have implications that are less plausible than the Implausible Conclusion. Boonin does not consider the existence solution in his chapter, so I cannot be entirely sure which implications of the existence solution he would select for entry into a showdown with the Implausible Conclusion. Nevertheless, the two implications of the existence solution that Boonin and other opponents would probably find the most objectionable are the following:

(1) Counterfactual dependence is not a sufficient condition for causation; and

(2) Some so-called "backtracking counterfactuals" are true.

In the remainder of this section, I will explain why the existence solution has each of these implications and then argue that such implications are not as bad as they might initially seem.

First, however, I want to make an observation. In his critique of other harm-based views, the implications that Boonin objects to are not pure metaphysical principles in the way that (1) and (2) are. Instead, they are implications about concrete cases. Some of the implications that Boonin objects to, for example, are the following:

(3) When I prevent you from receiving the medicine you need, I don't harm you (60).

(4) When I "shoot you in self-defense or as part of the prosecution of a just war", I don't harm you (63).

(5) When a doctor performs a procedure that causes a patient named Ray to regain his vision for forty years and then lose it again, the doctor wrongfully harms Ray (68).

(6) Every act of conception harms the person who is conceived (70).

The relative concreteness of these claims makes it easier for Boonin to argue that they are less plausible than the Implausible Conclusion. This is because Boonin often appeals to what his colleagues and students actually think in order to establish the comparative plausibility of various

propositions (25). And in my experience, students, in particular, are relatively confident when they form judgments about concrete cases of stealing medicine, shooting people, operating on people, conceiving children, and so on. However, I find that students tend to be less confident when they are asked how plausible claims like (1) and (2) are; there is less of a basis in what we might call "commonsense metaphysics" for either affirming or rejecting (1) and (2) directly. So to test (1) and (2) against common sense, we need to see how these abstract claims function within broader metaphysical theories that, through complicated chains of entailment, might eventually support or undercut our more easily testable, commonsense intuitions. I believe that (1) and (2) fit extraordinarily well within a broader metaphysical picture that supports a great deal of our commonsense intuitions about causation, harm, and wrongdoing. I cannot fully support such a claim in this article, but I will present some of the main reasons I believe this is so. At this point, however, I wish to simply to acknowledge the fact that the ostensibly implausible implications of the existence solution are slightly more abstract than the ostensibly implausible implications of other harm-based views.

Here is why the existence solution is committed to (1), the claim that counterfactual dependence is not a sufficient condition for causation. According to the existence solution, Wilma harms Pebbles by conceiving her, and this helps explain why Wilma's action was morally wrong. However, if counterfactual dependence *were* a sufficient condition for causation, then *every* parent would harm his or her child by bringing that child into existence. After all, we all suffer numerous harms throughout our lives: we get sick, we get injured, we suffer distress, and of course, we die. All of the corresponding harms depend counterfactually on the proposition that our parents conceived us. If our parents had *not* conceived us, we would not have suffered such harms. Thus, if counterfactual dependence were sufficient for causation, there would be little point in focusing on the harm associated with Pebbles's blindness—just one of the *myriad* harms that Wilma causes Pebbles, and probably not nearly the worst—and supposing that our knowledge of this one harm could, by itself, settle the question of whether Wilma acted wrongly.

Instead of endorsing counterfactual dependence as a sufficient condition for causation, the existence solution affirms a distinction between a cause and a mere condition. According to the existence solution, there is a good reason for Wilma to focus on Pebbles' blindness, rather than all the other harms that Pebbles will suffer: namely, Wilma causes the state of affairs in which Pebbles is blind, but her action is merely the condition of those other harms. From Wilma's perspective, Pebbles' blindness generates

a relatively strong reason against conceiving Pebbles, and those other harms do not. And if, as seems to be true in the case Boonin constructs, this reason against conceiving Pebbles is not outweighed by countervailing reasons, then conceiving Pebbles is wrong.

That is why the existence solution is committed to the claim that counterfactual dependence is not sufficient for causation. But is this commitment more implausible than the Implausible Conclusion? It *would* be more implausible if there were no defensible principles of causation that could accommodate such a commitment. But fortunately for the existence solution, there are coherent principles of causation that reject counterfactual dependence as sufficient for causation and instead support a distinction between a cause and a mere condition.[8] Moreover, there is good reason to accept such causal principles; they cohere remarkably well with the judgements about fault and blame that we make in law and everyday morality.[9] We don't, for example, blame our parents for our deaths (unless they do, in fact, kill us); we don't blame the fall of Rome on the meteor that wiped out the dinosaurs; and we don't attribute the election of Donald Trump to the existence of Earth's magnetic field. Thus, insofar as common sense morality has a role to play in vindicating Boonin's view, it can play that role equally well or better in vindicating the distinction between a cause and a mere condition.

As I noted above, the second implication that I believe opponents of the existence solution would object to is the claim that some so-called "backtracking counterfactuals" are true. To understand what a backtracking counterfactual is, consider two events, A and B, and suppose that B happens later in time than A. A foretracking counterfactual—the kind I have rejected as a sufficient condition for causation—takes this form:

(O) If A hadn't happened, then B wouldn't have happened.

By contrast, a backtracking counterfactual takes this form:

(B) If B hadn't happened, then A wouldn't have happened.

Thus, backtracking counterfactuals express a way of reasoning backwards: we reason that if a later event had not happened, an earlier event would not

8 For some causal principles that support a distinction between a cause and a mere condition, see Hart and Honoré (1985), Broadbent (2008), Gardner (forthcoming), and McGrath (2005). The existence solution is committed to the claim that at least *one* such account is correct.

9 Indeed, Hitchcock (2007) contrasts what he calls a "folk attributive" concept of causation—a concept that includes a distinction between a cause and a mere condition—with both a "metaphysical" concept of causation and a "scientific" concept of causation. He argues that the folk attributive concept and the scientific concept have promise, but he expresses skepticism that we need a theory of the "metaphysical" concept.

have happened. When uttered, many backtracking counterfactuals sound somewhat more awkward than foretracking counterfactuals, and this may be one of the reasons that philosophers tend not to appeal to them. As Alex Broadbent argues, it sounds odd to say, "If I had been richer, I would have chosen a different career". Nevertheless, he notes that the inference, *itself*, is not odd, for we can make sense of the claim, "For me to have been richer, I would have to have chosen a different career". Moreover, he notes that not all backtracking counterfactuals sound so odd, even when we leave out constructions like 'would have to have'; there is nothing objectionable about saying "If the librarian had not gone into the library, she would not have left her fingerprints on the door"[10].

To see why the existence solution is committed to the claim that some backtracking counterfactuals are true, consider the following case:

> *Dim Vision*. Because of retinal damage he suffered long ago, Jones has been blind for many years. Dr. Smith performs an operation that repairs some but not all of the damage. After the operation, Jones can see, but not very well: he has a condition we can call *dim vision*.[11]

A common intuition about this case is that Dr. Smith does not harm Jones; instead, he benefits Jones by improving his vision. The existence account of harming can justify this intuition if we appeal to the following backtracking counterfactual:

> (V) If Jones existed and the state of affairs in which he has dim vision did not obtain, then Dr. Smith wouldn't have operated and Jones would have been worse off in some respect.

If V is true, then dim vision isn't a harm for Jones, and Dr. Smith didn't harm Jones by causing it. But if no backtracking counterfactuals are true, then V *isn't* true, and the existence account of harming cannot justify the claim that Dr. Smith did not harm Jones. This would be a problem for the existence account of harming, given that there are many cases of helping that bear a structural similarity to Dim Vision, and the existence account would have to hold that all of these were cases of harming, rather than

10 See his unpublished book manuscript, *No Smoke Without Fire*: 18 – 19.

11 This case comes from Hanser (2009) and is also discussed in Harman (2009), Thomson (2010), and Gardner (2015).

helping.[12]

Fortunately, I do not think it is implausible to hold that some backtracking counterfactuals are true. As I noted above, such counterfactuals appear to support many of our inferences, especially those that are inferences to the best explanation. Moreover, there is support for the truth of some backtracking counterfactuals in the metaphysics literature.[13] To my knowledge, we currently lack a fully developed semantics for evaluating the truth or falsity of backtracking counterfactuals. Nevertheless, we lack an uncontroversial semantics for evaluating the truth or falsity of foretracking counterfactuals as well. This hasn't stopped us from incorporating foretracking counterfactuals into our theories and making intuitive judgments about whether they are true or false, so it should not stop us from appealing to and using our intuitions to evaluate backtracking counterfactuals.

5. A BULLET-BITING SHOWDOWN

Let us take stock. I have so far explained the non-identity problem, the harm-based approach to solving it, and how the existence solution fits into the harm-based approach. I have also reviewed Boonin's two requirements for successfully solving the non-identity problem—the robustness requirement and the modesty requirement—and I have argued that the existence solution to the non-identity problem satisfies both of them. In this final section, I will present some reasons for thinking that the existence solution is a *better* solution than Boonin's own proposal.

As I noted in the introduction, Boonin's own proposal for solving the non-identity problem is to accept the Implausible Conclusion. His reason for doing so is that he believes that every other proposal for solving the

12 Notice that the backtracking counterfactual we appeal to in Dim Vision has us "imagine away" Dr. Smith's action and then consider how much better or worse off Jones would be. In effect, then, the possible world we consider when we imagine the absence of Jones's new condition is the same possible world we imagine when we employ the principle that serves as P2 in the non-identity argument. This is why, in cases that don't involve either preemption or bringing a person into existence, the existence account of harming delivers all the same verdicts that are delivered by the account of harming that Boonin favors (2014: 53). Importantly, however, the existence account of harming delivers different verdicts in preemption and non-identity cases. To see why the verdict is different in non-identity cases, compare the following two counterfactuals: (1) If Jones existed and didn't have dim vision, Dr. Smith wouldn't have operated; and (2) If Pebbles existed and weren't blind, then Wilma wouldn't have conceived her. Whereas the first counterfactual is true, the second is obviously false: the possibility that Pebbles existed and Wilma didn't conceive her is remote, to say the least.

13 See, for example, Elga (2001) and Vihvelin (1994).

problem either doesn't really solve the problem, as advertised, or has implications that are even *less* plausible than the Implausible Conclusion. It is tempting to think, then, that Boonin's endorsement of the Implausible Conclusion is the *least* plausible feature of his view, and that his view as a whole is the *most plausible* of all the approaches to the non-identity problem.

However, as he proceeds towards accepting the Implausible Conclusion, Boonin bites some other bullets along the way. If we count the Implausible Conclusion as one implausible tenet of his view, then his view as a whole includes at least *three* implausible tenets. In all three cases, Boonin provides independent arguments for thinking that the three tenets are not as implausible as they initially seem. However, it's not clear that these other arguments are fully successful. In this final section, I want to explain the two implausible commitments Boonin's view has in addition to the Implausible Conclusion; challenge some of his reasons for holding that these commitments are plausible, after all; and show why the existence solution avoids both of them.

The first bullet Boonin bites has to do with pre-emption cases like the following:

> *Preemption*. Shooter shoots Victim in the abdomen, and although Victim does not die, he must spend the next two weeks in the hospital. If Shooter hadn't shot Victim, then Booter would have, and Victim would have ended up in the hospital with exactly the same injuries.[14]

Recall that according to P2, if A's act harms B, then A's act makes B worse off than B would otherwise have been. Because Boonin accepts P2, he is committed to the implausible claim that in *Preemption*, Shooter does not harm Victim.

Boonin accepts this implication of his view and argues that it is more plausible than it might initially appear to be. To show that such a claim is plausible, he constructs a number of thought experiments that he takes to be structurally similar to Preemption. He argues that in each case, the relevant action does not harm the victim, and that *if* it does not harm the victim, then Shooter's action in Preemption doesn't harm Victim, either. Here is a representative case:

> *Trolley*. Philip is trapped on a runaway trolley. The trolley is about to smash into a brick wall, which will kill him instantly. You can prevent this from happening by pulling a switch before the trolley approaches

14 This case is modeled on the case Boonin calls "The Two Hit Men" (2014: 57).

> a branch in the line. If you pull the switch, the trolley will be diverted and will instead smash into a second brick wall, which will also kill Philip instantly and at the exact same moment that the first wall would have killed him. You pull the switch. As a result, Philip is killed by smashing into the second brick wall rather than the first (58).

The problem with Boonin's appeal to Trolley and similar cases is that our intuitions about these cases may not be finely tuned enough to support the judgment that

(A) The action did not harm the victim

rather than

(B) The reason against the action was vanishingly weak.

In other words, both A and B might be reasonable ways to articulate our intuitions about Boonin's cases. However, only A can support Boonin's claim that it is plausible to think that Shooter's action does not harm Victim in Preemption; B cannot. And it is not clear why we should think that A captures our intuition better than B.

Moreover, there are reasons to think that the judgment we should really form about Boonin's cases is B, rather than A. If we conclude that in Preemption, Shooter does not harm Victim, various features of the case are difficult to explain. Why does Victim end up in the hospital for two weeks, if he wasn't harmed? Why is there blood leaking out of his abdomen? Importantly, the claim that Shooter harms Victim is compatible with the claim that the reason against shooting Victim was vanishingly weak: there is always a reason against harming, but the strength of that reason clearly varies from one case to another.[15] The reason for me to refrain from pinching you on the arm in order to prevent a bomb from exploding is obviously weaker than the reason for me to refrain from pinching you on the arm in order to annoy you. Preemption and Trolley may well be cases where an action causes harm, but because of the unusual circumstances, it is simply less important to avoid causing that harm. Thus, taking these additional considerations into account, it remains implausible to hold that in Preemption, Shooter does not harm Victim.

Fortunately, a proponent of the existence solution to the non-identity problem does not have to bite the bullet that Boonin bites in the case of Preemption. According to the existence account of harming, Shooter harms Victim: Victim has suffered a harm, and by shooting Victim, Shooter caused that harm. This is a major advantage of the view, and it counts towards thinking that the existence solution is more plausible, on the

15 See Gardner (2017).

whole, than Boonin's solution to the non-identity problem.

The other implausible implication of Boonin's view is illustrated by the following case:

> *Angry Alastair:* Angry Alastair slaps you in the face. This is because he was trying to restrain himself; if he hadn't slapped you in the face, he would have torn your arms off. [16]

Recall, once more, that Boonin accepts P2, which tells us that A's act harms B only if A's act makes B worse off than B would otherwise have been. This principle implies that Alastair does not harm you by slapping you in the face, since if he hadn't slapped you in the face, you would have been even worse off. Boonin accepts this implication of his view and argues that it is more plausible that it might initially seem. His argument appeals to the following case:

> *Parking Meter.* Your car is parked in a space with an expired parking meter. A police officer is about to write you a $25 ticket, but then he notices that your door is unlocked and that there are four quarters in your cup holder. The police officer opens your door, takes the change, and feeds the meter. If he hadn't taken the dollar's worth of change from your cup holder, he would have written you the $25 ticket. [17]

Boonin's intuition about the case is that the police officer doesn't harm you. But since Parking Meter is structurally similar to Angry Alastair, Alastair must not harm you in Angry Alastair, either.

Boonin's argument is not convincing. In the case of Parking Meter, even if we have the intuition that you aren't harmed, it's not clear that the best explanation for this intuition is that P2 is true. An alternative explanation is that losing a dollar simply isn't a big deal (at least for affluent people who own cars).[18] Because it is easy to confuse a small harm with no harm, we need to test whether our intuitions are responding to the size of the loss or the absence of a necessary condition on harming. I propose that we

16 A slightly longer description of the case appears in Boonin (2014: 62).

17 This is my paraphrase of the parking meter case in Boonin (2014: 62).

18 An anonymous reviewer suggests a third explanation: that money is not the currency of harm. While I agree with the reviewer that the concept of "financial harm" is less compelling than the concept of harm to one's body or property, I disagree that Boonin's case relies on the assumption that money really is the currency of harm. Boonin elsewhere makes it clear that he thinks well-being is the currency of harm. Thus, in Parking Meter, I take him to be suggesting that when the police officer takes your dollar, your well-being drops by an amount that corresponds to the loss of one dollar. (And when I say "drops," I am making a temporal comparison, rather than a counterfactual comparison: your well-being is lower than it was before, but your well-being is not as low as it would have been had the police officer given you the ticket.) For more discussion about what we should take to be the currency of harm, see Tadros (2014).

consider a slightly different case:

> *Theft.* Your car is parked in a space with an expired parking meter. Unfortunately, you have left a duffel bag containing $5,000 in the backseat. A police officer is angry that about the expired meter, and he wants to ram his police car into your parked car, causing you $8,000 in damage. However, he changes his mind and simply takes the duffel bag.

Here, you are out $5,000, and I think it's clearer that you are harmed. Moreover, Angry Alastair more closely resembles Theft than it resembles Parking Meter, given the size of the harms at issue. Thus, it remains implausible to claim that in Angry Alastair, you are not harmed.

Here is how the existence account of harming supports the judgment that Alastair harms you: there is some general state of affairs either identical to or closely associated with the state of affairs in which you have been physically injured *at all*. If that state of affairs had not obtained, you would have been better off along some injury-related dimension of your well-being. Moreover, Alastair caused that general state of affairs, and so he harmed you. Importantly, the account does not necessarily imply that some *specific* upshot of the slap (such as a stinging cheek) is a harm to you. It may well be that if you hadn't suffered the harm associated with the stinging cheek, you would have been worse off along the same dimension on which the cheek pain registers.[19] But in this case, it's reasonable to think that the slap is a benefit to you anyway, given that it prevents a greater harm.

In other words, in Angry Alastair, the harm you suffer is that your well-being is set back at all along a certain dimension, and not that your well-being is set back some specific number of units along that dimension. This implication of the existence solution is more plausible, I think, than the claim that you are not harmed in any sense.

6. CONCLUSION

Although Boonin critiques a number of different versions of the harm-based approach to the non-identity problem, he does not critique the existence solution. I have argued that the existence solution holds up well

19 On the other hand, if a stinging cheek corresponds to a completely different dimension of well-being than torn-off-arm pain, then there is a respect in which you would have been better off without the stinging cheek. In that case, Alastair's action causes you multiple harms: the general harm of being injured at all and the specific harm of the stinging cheek.

against the objections Boonin raises to the other harm-based views. I have also argued that the existence solution is more plausible that Boonin's own proposal for solving the non-identity problem. Nevertheless, the existence solution is committed to some controversial claims about counterfactuals and causation. To fully vindicate the existence solution, we will need a more complete account of the semantics of both foretracking and backtracking counterfactuals, and we will need to situate them within a plausible theory of causation that distinguishes between causes and mere conditions.

BIBLIOGRAPHY

Adams, R., 1979: "Existence, Self-Interest, and the Problem of Evil", *Nous* 13: 53 – 65.
Boonin, D., 2014: *The Non-Identity Problem and the Ethics of Future People*, Oxford: Oxford University Press.
Broadbent, A., Unpublished book manuscript: *No Fire Without Smoke: Causation, Counterfactuals, and How to Make a Difference.*
— 2008: "The Difference Between Cause and Condition", *Proceedings of the Aristotelian Society* 108: 355 – 364.
Elga, A., 2001: "Statistical Mechanics and the Asymmetry of Counterfactual Dependence", *Philosophy of Science* 68: S313 – S324.
Gardner, M., 2015: "A Harm-Based Solution to the Non-Identity Problem", *Ergo* 2: 427 – 44.
—2017: "On the Strength of the Reason Against Harming", *Journal of Moral Philosophy* 14: 73 – 87.
— Forthcoming: "When Good things Happen to Harmed People", *Ethical Theory and Moral Practice.*
Hanser, M., 2008: "The Metaphysics of Harm", *Philosophy and Phenomenological Research* 77: 421 – 50.
— 2009: "Harming and Procreating", in *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, ed. M. Roberts and D. Wasserman, 179 – 99, New York: Springer.
Harman, E., 2004: "Can We Harm and Benefit in Creating?" *Philosophical Perspectives* 18: 89 – 113.
— 2009: "Harming as Causing Harm", in *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, ed. M. Roberts and D. Wasserman, 137 – 54, New York: Springer.
Hart, H.L.A. and T. Honoré, 1985: *Causation in the Law, 2nd* ed. Oxford: Oxford University Press.
Hitchcock, C., 2007: "Three Concepts of Causation", *Philosophy Compass* 2: 508 – 516.
Lewis, D., 1973: "Causation", *Journal of Philosophy* 70: 556 – 67.
McGrath, S., 2005: "Causation by Omission: A Dilemma", *Philosophical Studies* 123: 125 – 148.
Parfit, D., 1976: "On Doing the Best for Our Children", in *Ethics and*

Population, ed. M. Bayles, 100 – 15, Cambridge, MA: Schenkman.
— 1984: *Reasons and Persons*, Oxford: Oxford University Press.
Shiffrin, S., 1999: "Wrongful Life, Procreative Responsibility, and the Significance of Harm", *Legal Theory* 5: 117 – 48.
Schwartz, T., 1978: "Obligations to Posterity", in *Obligations to Future Generations*, ed. R. Sikora and B. Barry, 3 – 13, Philadelphia: Temple University Press.
Tadros, V., 2014: "What Might Have Been", in P*hilosophical Foundations of the Law of Torts*, ed. J. Oberdiek, 171 – 92, Oxford: Oxford University Press.
Thomson, J. J., 2011: "More on the Metaphysics of Harm", *Philosophy and Phenomenological Research* 82: 436 – 58.
Vihvelin, K., 1994: "Causes, Effects, and Counterfactual Dependence", *Australasian Journal of Philosophy* 73: 560 – 573.

DOI: 10.31009/LEAP.2019.V7.04

Rights, Wronging, and the Snares of Non-Identity

RAHUL KUMAR
Queen's University

ABSTRACT

Chapter five of Boonin's The *Non-Identity Problem and the Ethics of Future People* examines what he identifies as the fourth premise of the non-identity argument: if an act harms no one it wrongs no one. He argues that the reasons for rejecting it offer no grounds for resisting the conclusion of the non-identity problem. I argue that key arguments for this claim subtly misconstrue non-consequentialist reasoning about wronging. This shows, I suggest, that the prospects for resisting the non-identity problem by harnessing the resources of a general account of what it is for one person to wrong another are in fact quite bright.

Keywords: Rights, non-consequentialism, deontology, non-identity, wronging.

1. INTRODUCTION

The non-identity problem starts with the observation that what a person chooses now can determine the metaphysical identity of those who will exist in the future. Seemingly innocuous, it yields, together with a few independently plausible premises, a startling conclusion, one nicely illustrated by the case of Wilma, who has to decide when to conceive.[1] If she conceives now, she will give birth to a blind child, Pebbles. But if she postpones conception and accepts the minor inconvenience of taking a pill each day for the next two months before proceeding, she will give birth to a sighted child, Rocks.

Intuitively, Wilma deciding to conceive now rather than waiting two months is wrong. It is wrong because, in choosing not to spare her child a significant disadvantage in life, when she can easily do so, she wrongs it.

1 This is the central example Boonin works throughout his book.

What the non-identity problem draws our attention to is the significance of blind Pebbles being the product of the union of a different sperm and egg than that which yields sighted Rocks. What turns on Wilma's decision, therefore, is not the fate of a particular person who will either be born better or worse off, depending on what she decides, but which of two distinct individuals will be brought into existence—blind Pebbles or sighted Rocks.

Neither Pebbles nor Rocks is rendered better or worse off by being conceived than she or he would have been had Wilma decided differently. But if Wilma deciding to conceive now rather than wait will not make the resulting child, Pebbles, worse off than she otherwise would have been, it is hard to see how she is harmed by Wilma's choice. And if Pebbles is not harmed, what basis do we have for the intuitive conviction that it would be wrong for Wilma not to wait because not waiting would wrong her future child? The unintuitive conclusion, that Wilma deciding to conceive now is not wrong because doing so would not harm her child, ends up looking difficult to resist.

One of the many delights and challenges of The *Non-Identity Problem and the Ethics of Future People* is the care with which David Boonin interrogates an array of arguments that aim to show one of the premises of the argument for the unintuitive conclusion to be mistaken. Finding them all unsatisfactory, he concludes that the only philosophically defensible option is to simply embrace it as true: the non-identity problem should be viewed not as a problem to be overcome, but as an argument establishing an important moral truth.

I remain unconvinced. The challenges to the non-identity problem Boonin surveys may not, as stated, succeed. But they can be improved upon, rendering them more effective against the premises of the argument they target. In particular, I think it is worth looking more closely at what Boonin identifies as P4 of the non-identity argument, which holds that if an act does not harm anyone, then the act does not wrong anyone. What he asks is whether there is a way of understanding an act as wronging another, even if it could not, in principle, harm her, that can be employed to dissolve the non-identity problem. His conclusion is that there is not (Boonin, 2014:148). Certain arguments for why this is so, however, seem to me to subtly misconstrue non-consequentialist reasoning about wronging, and in ways that close off the possibility that a better general understanding of what it is for one person to wrong another might give us good grounds for resisting P4 (and hence the Implausible Conclusion).

To develop this claim, I want to first consider Boonin's arguments as to why appealing to the rights of those who will be brought into existence

cannot help dissolve the non-identity problem. I will then turn to his critical discussion of what he calls the 'actual and *de dicto*' approach. Both strands of argument capture insights, I will argue, that can be put to work as part of a defense of the rejection of the fourth premise as crucial to dissolving the non-identity problem.

2. BOONIN ON RIGHTS-BASED SOLUTIONS

Many are drawn to the thought that a choice resulting in a person coming into existence can still wrong her—even if she would not have otherwise come into existence and has a life worth living—because the making of that choice violated her rights. Boonin helpfully distinguishes between direct and indirect versions of this argumentative strategy. The direct version says that Wilma choosing to conceive now rather than later wrongs Pebbles because that choice violates a right that Pebbles has at the time the choice is to be made. The indirect version says that Wilma's act of choosing to conceive Pebbles generates a new right in Pebbles that will foreseeably be violated.

The direct version requires that we posit the objection to Wilma conceiving Pebbles as Pebbles' claim-right against Wilma that she not do so. That there is such a right looks difficult to make sense of, let alone defend. How can a person who was not previously in existence and does not now exist be the bearer of a right that requires that the very choice that will bring her into existence not be made? The indirect version of the rights-based strategy looks comparatively more promising. Assuming that each person has a right not to exist as blind, and Pebbles can only exist as blind, conceiving Pebbles is to choose to bring someone into existence whose right cannot be honoured. More precisely, the proposal is that her only being able to exist as wronged, because her right cannot be honoured, is a decisive reason for Wilma to delay conceiving.

Boonin argues that indirectly appealing to Pebbles' right cannot preserve the intuitive conviction that Wilma ought not to conceive now because doing so would wrong Pebbles. Two arguments for this conclusion are, for my purposes, of particular importance. First, if Pebbles has a right not to be born blind, it follows that Wilma violates that right by bringing her into existence, whether or not Wilma is able to conceive a sighted child. That is implausible. Wilma arguably owes it to any child she chooses to conceive to take reasonable steps to ensure that the child is born without any significant impairments. But if, for reasons outside her control, she can only conceive children who will be born blind but have otherwise good lives, she does not, even defeasibly, wrong anyone when she decides to

conceive. The suggestion that she does is even less plausible than that of Wilma doing no wrong in bringing Pebbles rather than Rocks into existence.

The second argument is one not explicitly stated, but is one I take to underlie both Boonin's discussions of permissible rights-infringements and the 'rights waiver' argument (Boonin, 2014:116, 119-23). It goes as follows: if Pebbles has a right not to be born blind, it is reasonable to think that there must be some value for Pebbles in having the kind of protection the right in question affords her. But it is doubtful that the right does in fact have any value for her, which is a reason to think the right in question lacks a rationale, other than the positing of it appearing to offer grounds for resisting the non-identity problem.

Consider, first, Boonin's discussion of a rights-infringement case:

> "Suppose, for example, that you will soon die if I don't donate a kidney to you, and that I can either promise that I will give you a kidney and promise that I will paint your house, in which case I will keep the first promise but break the second promise, or that I can promise to do neither, in which case I will do neither. In this case, it seems clear that it would not be wrong for me to make both promises and then keep the first one, even though doing so would again generate in you a right that would later be violated. This suggests that even if the generation of a new right that will later be violated establishes a moral presumption against doing an act, the presumption can be overcome if there is a substantial enough difference in terms of the overall consequences between doing the act and not doing the act. Even if Wilma's act of conceiving Pebbles does prove to generate a new right in Pebbles that will later be violated, then, this will not be enough to show that her act was immoral….If it is not wrong to generate a right that will later be violated when doing so is necessary in order to save someone's life, then why should it be wrong to generate a right that will later be violated when doing so is necessary in order to give someone a life that would be worth saving in the first place?" (Boonin, 2014:116).

Proponents of a rights-based approach to morality will point out that the kind of 'overcoming' that this passage suggests is meant to be excluded by taking the reason the right provides seriously. As they see it, allowing the reason to respect a justified right to be overcome by substantially better overall consequences of violating it simply undermines the value of having the protection the right affords. But Boonin's point, I take it, is not that the presumption that the right be honored is simply 'overcome'. Rather, it is that specifying the contours of the right by reference to the interests of the

individual that the right is meant to protect shows that, in the promising case, there is no reason to hold that not keeping the promise to paint the house violates the individual's right, and that in Pebbles' case, there is no reason to think that her right is violated if she is born blind.[2] In general, if having the protection a right affords is not in the right-holder's interest, why posit such a right to begin with?

This point is even more crisply illustrated by the 'rights-waiver' argument, which goes as follows:

> "if Pebbles really does have a right not to be brought into existence as a blind person, for example, it is reasonable to suppose hypothetically that she would agree to waive her right against Wilma conceiving her if she could or to conclude that she actually will agree to waive her right after the fact if Wilma conceives her…blindness is presumably a bad condition to exist in, but it is not so bad that it is worse than not existing at all and so, on the whole….Pebbles will be glad that she was conceived despite her lack of vision. And if Pebbles would or will waive her right not to be conceived under the circumstances in which Wilma conceives her, then the fact that she has a right not to be conceived under those circumstances—if it is a fact—cannot make it morally wrong for Wilma to conceive her" (Boonin, 2014:120).

It is, I agree, plausible that Pebbles would waive any right she has not to be conceived because she will be blind. This is because the protection the right provides is not a right that protects her interests. Quite the opposite, it is an obstacle to her existing as a subject with interests and a life worth living. This fact is epistemically accessible to Wilma prior to conception. She is in a position to know that refusing to violate Pebbles' right not to be born blind cannot be justified on the grounds that she is respecting Pebbles' interests or status. From her point of view, allowing the right in question to play a role in her deliberations looks like a form of rights fetishism. But if the right in question does not provide Pebbles a form of protection she has reason to value, and there is no reason for Wilma to give it any role in her deliberations, why think that there is any such right to begin with?[3]

Advocates of rights-based solution to the non-identity problem ought, I believe, to agree: there is no such right. That is, Pebbles has no right to be

2 For useful discussion, see Oberdiek (2004).

3 I assume here that the wider society has no interests that might be compromised by people choosing to conceive children who will be born blind.

born sighted, such that she is wronged if she is born blind. What they should go on to say, however, is that it is not their view that a right—in particular, a claim-right correlated with another's duty—is properly understood as an entitlement to a certain state of affairs obtaining, such that one is wronged by that state failing to obtain. Individual rights are fundamentally relational in character: each marks out a respect in which an individual is entitled to have certain of her vital interests be recognized by others, by way of giving their protection priority over other considerations in the practical deliberations of the person against whom the right is held. Taken together, individual rights form a 'moral basic structure': a web of duties and permissions that govern how individuals are to relate to one another in a range of circumstances.

The interest individuals generally have in not having a certain type of state of affairs obtain partially fills out the content of a given individual right. But my right that you not make a certain choice is one that I have in virtue of the fact you stand in a position relative to me of someone whose choices can affect the interests of someone in my position, in ways that persons in my position generally have good reason to want to be protected against. The right provides this protection by requiring that one against whom I have a claim, in virtue of my right, conduct herself in a way that is consistent with prioritizing in her practical deliberations the securing of my interests.

For instance, if I promise to take you to your doctor's appointment, you have a claim-right against me that I do so. What that entitles you to is that in planning my day, I organize it in a way that is compatible with getting you to the appointment on time—even if doing so is not optimal from the point of view of my own interests—and that I make a reasonable effort to stick to the plan. This last qualification is important for understanding what counts as a violation of a right. I may not succeed in getting you to your appointment on time. Whether or not my failure counts as a violation of your right depends, however, on the circumstances which account for the failure. I wrong you if I'm late because I didn't bother taking into account the traffic around the hour of your appointment. But I don't wrong you if I'm late because someone ran a red light and slammed into my car. To violate your right, such that resentment and a demand for compensation are justified, it isn't enough that the right outcome fails to obtain. The failure has to be traceable to my failure to be guided by your interests in the way to which you are entitled by your right.

Now consider a different case, one in which I cut through your backyard as a shortcut on my walk to the coffee shop, reasonably assuming you will not mind. In so doing, I violate your property right. The interest the

property right protects is your interest in it being up to you to decide who gets to use your property. I may be right that you would not object to my cutting through your backyard, but the point of the property right is to deny others the authority to make a call concerning the use of your property. My taking the shortcut wrongs you.

That would not, of course, be true if I cut through your backyard to save your life, my own life, or someone else's life (cf. Boonin, 2014: 122). Rights are not absolute; they have contours that reflect a balance between the interests of the right holder in having a certain kind of protection and ways in which respecting an individual's right can be burdensome or costly for others.

This way of understanding rights claims is what I believe advocates of rights-based response to the non-identity problem have in mind. Consider, for example, Woodward's influential discussion of the non-identity problem as it arises in the context of Parfit's Risky Policy case. That case goes as follows:

> *The Risky Policy* as a community, we must choose between two energy policies. Both would be completely safe for at least three centuries, but one would have certain risks in the further future. This policy involves the burial of nuclear waste in areas where, in the next few centuries, there is no risk of an earthquake. But since this waste will remain radio-active for thousands of years, there will be risks in the distant future. If we choose this risky policy, the standard of living will be somewhat higher over the next century. We do choose this policy. As a result, there is a catastrophe many centuries later. Because of geological changes to the Earth's surface, an earthquake releases radiation, which kills thousands of people. Though they are killed by this catastrophe, these people will have had lives that are worth living. We can assume that this radiation affects only people who are born after its release, and that it gives them an incurable disease that will kill them at about the age of 40. This disease has no effects before it kills (Parfit, 1984:371-2).

Woodard, developing the intuitive idea that those killed in further future are wronged by the adoption of the risky policy, argues that it

> "seems plausible that the [people who are later killed by the leaking toxic waste], like other innocent people, have rights that others should not knowingly pursue policies that will kill, injure, or poison them or will create substantial risks of these results, at least when there is no

> weighty justification for such policies and alternative policies which involve no such risks are available" (Woodward, 1986:812).

Woodward's proposal, Boonin argues, cannot succeed, as the right in question is either too strong or too weak. Read simply as a right not to be killed, it is too strong (Boonin, 2014:118). Highways, for example, are designed according to safety standards that reduce, but do not eliminate, the risks associated with using them. Those risks are expected to eventuate in the death of certain number of drivers each year. Their deaths are correctly understood to be a consequence of past decisions about highway safety standards. But it is implausible to hold their right not to be killed to have been violated. Similarly, people are killed all the time as the result of past planning or design decisions that create unforeseeable risks that eventuate in freak accidents. No right not to be killed of those who end up dead is violated if the risk whose eventuation kills them was not reasonably foreseeable at the time it was created. Reading the right as a right to not be unjustifiably killed, on other hand, makes the right in question too weak. The adoption of the risky policy results in future deaths, but if the right is a right not to be unjustifiably killed, it will not speak to whether or not those deaths count as wrongful rights violations. They do only if adopting the risky policy was not justified. Settling that question is not a matter to which the right not to be unjustifiably killed can contribute.

We can make better sense of what Woodward is getting at, however, if we think of the right in question not as an entitlement of future people that a state of affairs not obtain in which they have been unjustifiably killed, but as a right that policy choices be made that give their vital interests appropriate weight in determining what risks of harm (or in some cases, virtually certain harms) they can be justifiably exposed to. So understood, those killed in the further future are not wronged because they have been killed (or when they are killed). What wrongs them, rather, is that at the time of deciding between the policy alternatives, a decision was made to gamble with their lives, for no compelling reason, by adopting a policy that created a non-trivial risk that what does in fact happen will happen. That they are killed alters the significance of that wrong in their lives, and in standard cases, would have implications for how we think about their claim to compensation. It does not bear on whether they are in fact wronged by the adoption of the risky policy.

Finally, to return to the question of how to think about the morality of Wilma's choice, the proposed way of thinking about rights-claims lacks the implausible implication that Wilma wrongs Pebbles by conceiving her,

even if she is only able to conceive children who will be born blind. Pebbles' right is not to a state of affairs obtaining in which she exists as sighted rather than in a 'harmed state' as blind. Rather, what her right entitles her to is Wilma giving appropriate weight, in her deliberations about when to conceive, to the interests of her future child in not being born blind. If that cannot be affected by when Wilma decides to conceive, Pebbles' right cannot be violated by that choice.

3. RIGHTS-BASED SOLUTIONS AND WIDE-PERSON-AFFECTING PRINCIPLES

The understanding of rights advanced in the last section—as claims individuals have against others that they regulate their conduct in a way consistent with the protection of one another's vital interests—is one I think of as being distinctively non-consequentialist in character. The proposal is that it takes us some way towards redeeming the promise of appealing to rights in non-identity cases.

It only goes some of the way because it appears to falter over the matter of the identity of the right holder. Recalling the distinction between direct and indirect versions of a rights-based approach, the way of appealing to rights in non-identity cases I am proposing is direct. It says that Wilma violates Pebbles' right when she chooses to conceive Pebbles rather than Rocks; an indirect argument would locate the objection to conceiving Pebbles in bringing into existence a person who will exist with an important right of hers violated. The indirect approach was rejected because, in positing a right not to be born in a certain state, it misconstrues the interpersonal role of rights. In particular, it leaves out a crucial dimension of rights as demands individuals make of another that they constrain their conduct in ways that embody recognition of one another's vital interests. A direct approach captures that aspect of rights, but it appears to problematically posit as a right-holder a person who does not, and will not, exist unless the very choice is made that respect for her right requires not be made. That is a deeply puzzling idea.

One way to think about the problem, prompted by the earlier discussion of rights, that makes it less puzzling goes as follows: what Wilma's child's right to be born sighted rather than blind amounts to is a requirement that she give a special weight, in her deliberations concerning whether and when to conceive, to the interests of the future child, choosing the course that looks to be reasonably protective of the child's interests. The 'interests of the child' in question are not, however, the interests of any particular

child, but those of a generically described child (or the interests' children typically have).

Boonin considers a suggestion along these lines under the heading of the 'actual and *de dicto*' people approach. It holds that Wilma choosing to bring Pebbles into existence wrongs Pebbles because doing so is unfair. We can see why it is unfair if we think about what principle governing fair procreative choice would be chosen from an impartial point of view, described by a veil of ignorance. The veil secures the impartiality of the choice made behind it by depriving all those concerned—a domain that includes Wilma, Pebbles, and Rocks—of any information relevant to determining the particular people they are. The question to be considered is what principle for the regulation of procreative choice is it rationale for them to adopt in light of its implications for both a woman and her child, understanding 'her child' in a *de dicto* rather than a *de re* sense. As Boonin explains,

> "the claim that Wilma wants to marry the tallest man in Bedrock, for example, could mean two different things: that there is a particular man that Wilma wants to marry who happens to be the tallest man in Bedrock (the *de re* reading) or that she wants to marry whatever man turns out to satisfy the description "the tallest man in Bedrock" (the *de dicto* reading)" (Boonin, 2014:31).

In asking what principle it would be rational to choose behind the veil, then, are the implications of different possible principles regulating procreative choice given the equal likelihood of turning out to be (a) a woman with a slight preference for conceiving a blind child now rather than sighted child in a few months' time, (b) the child she conceives, who could either be blind or sighted, or (c) someone for whom the principle has no implications. Reasoning this way supports the conclusion that the principle it is rational to accept is one that requires that Wilma delay conception because the worse-case scenario for the worst off person is not as bad as the worse-case scenario that imposes no such obligation (Boonin, 2014:135).

Boonin concludes that this line of argument does not succeed. Why that is so rests on a point that is evident when we compare the use of the veil of ignorance device in this case with how it is used by John Rawls and John Harsanyi, in the context of reasoning about principles for fair distribution. The veil they describe asks us to imagine a range of social positions, each of which describes *someone's* social circumstances. Reasoning behind the veil, what a person has to ask herself is how she would choose to have the

goods of cooperation distributed across the social positions, knowing that one of them characterizes not just some actual person's life, but *her* life.

The veil of ignorance being appealed to in defense of Wilma's obligation to delay conceiving is different, insofar as it asks us to consider the implications of the principle for different lives that may or not turn out to be actual, depending on what choice is made, rather than the lives that, outside the veil, are the lives of actual people. Taking into account, behind the veil, the implication of a principle for merely possible lives is unmotivated. It is easy to understand the motivation for including behind it the implications of a principle for the lives people will in fact be leading outside of it. The concern is to be able to arrive at a principle that will allow the circumstances of their lives to be justified to them as fair. But nothing analogous can be said in favor of taking into account merely possible lives. The best reason for doing so, that it allows us to avoid biting the non-identity bullet, is obviously question begging (Boonin, 2014:136).

There is, however, another way of thinking about what the *de dicto* appeal to the interests of Wilma's child is getting at that avoids the implausible appeal to the interests of 'possible people'. What derails the idea of appealing to the interests of the child *de dicto* is treating being born blind rather than sighted as an identity-fixing fact, such that Wilma's choice between her child between born blind or sighted is a choice between two possible lives, only one of which will be actual. The alternative reading says that the way to understand Wilma's choice is as a choice between her child being blind or sighted, where the life we are referring to is a life that will be actual, that of her child, whichever choice she makes (Boonin, 2014:138).

This reading succeeds, Boonin argues, only if we assume that whether a person is blind or sighted is not identity fixing. Doing so allows us to think of Pebbles and Rocks as representing two possibilities for one and the same person. As he puts it, this reading

> "produces the desired result….only because it views conceiving blind Pebbles rather than sighted Rocks as making the person who is conceived worse off than that very person would otherwise have been. This means that this version of the approach can justify the claim that Wilma's act is morally wrong only by treating the case as one that is not a non-identity case. But if the approach works only in cases that are not non-identity cases, then it cannot solve the non-identity problem" (Boonin, 2014:138).

It is, I agree, implausible to treat the choice between creating a blind or a

sighted child as not affecting the identity of the child, such that we can say that it is worse for a particular child to have been born blind than sighted. But this is not a claim that a defender of the alternative reading of the *de dicto* argument needs to make. What she should instead say is that the objection to Wilma's choice, from an impartial point of view, makes no appeal to the comparative claim that choosing to conceive Pebbles rather than Rocks is *worse* for her child. She should, rather, defend the principle requiring Wilma to choose to conceive Rocks rather than Pebbles as what Parfit calls a *wide-person affecting principle*. Such a principle

> "makes claims about the intrinsic goodness of different outcomes…. When we apply the Wide Principle, we can consider each outcome on its own, and ask how good or bad this outcome would be for people. An outcome's intrinsic goodness does not depend on its relation to other outcomes" (Parfit 2017:140).[4]

Narrow person-affecting principles, on the other hand, are not concerned with "intrinsic goodness or badness, [so] we cannot consider each outcome on its own. We cannot ask whether some outcome would be bet for people, or worse for people, since these relations are essentially comparative, holding only between different outcomes" (Parfit 2017:140).

Parfit characterizes wide and narrow person-affecting principles as applying to the assessment of outcomes. They can, however, be readily put to work in the context of thinking about the justification of rights by restating them in deontic rather than telic terms. So understood, they are principles that bear on the assessment of choices rather than outcomes. A narrow deontic principle makes claims about which choices are better or worse for people, while a wide deontic principle makes claims about the intrinsic goodness or badness of a choice for people.

The claim that 'Wilma's prospective child' has a right to Wilma choosing to conceive a sighted rather than a blind child is most plausibly understood in wide deontic person-affecting terms, eschewing any appeal to comparative claims about what is better or worse for the child. What the right entitles 'Wilma's child' (understood *de dicto*) to is that Wilma constrain her deliberations so as to make choices concerning her child that are, given the information available to her, defensible as consistent with a concern for what will be intrinsically good for her child. The right is therefore violated by Wilma making a choice that will foreseeably result in

4 Boonin is careful to distinguish the 'moderate principle' he ultimately endorses from the narrow person-affecting principle (Boonin, 2014:276).

her child being born blind when there is an adequate alternative option available, not simply by it being born blind. It has no bite if Wilma is only able to conceive blind children.

I cannot here try and address all the objections to the line of argument I have been developing; I will just consider two. The first points out that the way of thinking about the wrong Wilma does to Pebbles I am proposing leaves us unable to square Wilma's deliberative perspective in making her choice with that of the particular child created. Say Wilma goes ahead and conceives Pebbles. The suggestion is that in doing so, she wrongs Pebbles by violating her right. Because she is wronged, her experiencing certain reactive attitudes, like resentment or blame, associated with having been wronged would be completely justified. But in this case, it is hard to see why Pebbles should resent, or blame Wilma. Wilma's choice, after all, had no bad implications for Pebbles, and that it did not is not just a matter of luck. Wilma's choice, that is, is in no way an affront *to Pebbles*.

The idea that a person's objection to conduct that wrongs her must be personal gains support from the intuitive association of having been wronged with reactive attitudes like blame and resentment. What non-identity cases show us is that this association is potentially misleading, insofar as it leads us to conclude that the test for whether a person has been wronged is to ask ourselves what justifies the wronged *personally* complaining about the wrongdoer's conduct. Pebbles has no personal complaint to lodge against Wilma's choice because that choice could not express an attitude towards her in particular; at time of the choice, she was only available to Wilma in her deliberations under the description of 'my future child'.

What follows, however, is not that she cannot claim to have been wronged, but that the basis of her complaint is in an important respect less personal. Wilma's choice wrongs her child because it is not in conformity with her obligation, having decided to have a child, to make choices that reasonably prioritizes the protection of the child's interests. Pebbles can feel indignation towards Wilma for choosing not to wait two months before conceiving. And she can resent that choice when she thinks of herself under the more impersonal description of 'Wilma's child'. It is in virtue of it uniquely applying to her that she can identify herself as the person wronged by Wilma's choice. In another, more personal, respect, however, she cannot, as long as she has a particular life to which she is attached, resent Wilma making the choice she did. Which is not to say that Pebbles' waives her right that Wilma not make the choice she did; because the right is one that constrains Wilma's choices with respect to 'her future child', it is not a right that any particular future child who ends up coming into

existence is in a position to waive.

The second objection protests that locating the grounds of Pebbles' claim to be wronged in a more impersonal consideration lacks a rationale apart from that of wanting to evade the non-identity problem. A plausible attempt to solve the non-identity problem cannot employ the problem's evasion as part of the argument in support of it.[5]

The rationale for the move in question does not in fact appeal to the non-identity problem. What it relies on is a general feature of rights-claims, as I understand them. If you have a right against me that I not make a certain choice, this is in virtue of the character of the relation we stand in to one another, not in considerations having to do with us as particular individuals. To illustrate, consider the right to privacy. Bob has a right that Sue not read his e-mail without his permission. That is because the relation between them is one of strangers. Now, Bob does not in fact mind if strangers read his e-mail without seeking his consent (there is nothing in it he would not happily reveal if asked). He would even welcome someone doing so because it would be evidence of someone taking a special interest in his life. But that has no bearing on his right that Sue not read his email without his consent. That right is one he has against Sue in virtue of the fact that they are strangers to one another; the right to privacy is one that governs relations between strangers, protecting interests that those who are strangers to one another typically have. Sue may come to know, through a third party, that Bob would welcome a stranger breaking into his e-mail account and reading his messages. But that would not change the salient, comparatively impersonal, fact that their being strangers to one another is sufficient grounds for her reading his e-mail without his consent to constitute, and for her to conclude that her doing so would, wronging Bob by violating his right to privacy.

4. CONCLUSION

In this brief discussion, I have argued that the prospects for a rights-based solution to the non-identity problem remain good, despite Boonin's forceful arguments to the contrary. In arguing for this position, I have appealed to various aspects of non-consequentialist thinking about rights and their moral significance that I believe any successful rights-based solution to the non-identity problem needs to emphasize. Whether such a strategy can ultimately be vindicated can only be settled by further exploration of what the best version of such a solution looks like.

5 Boonin calls this the 'independence requirement' (Boonin, 2014:20).

BIBLIOGRAPHY

Boonin, D.,2014: *The Non-Identity Problem and the Ethics of Future People*, Oxford: Oxford University Press

Oberdiek, J.,2004: "Lost in moral space: On the infringing/violating distinction and its place in the theory of rights", *Law and Philosophy* 23: 325-346.

Parfit, D., 1984: *Reasons and Persons*, Oxford: Oxford University Press

—2017: "Future People, the Non-Identity Problem, and Person-Affecting Principles" *Philosophy and Public Affairs* 45 2:118-57.

Woodard, J., 1986: "The Non-Identity Problem" *Ethics* Vol. 96 4:804-831.

DOI: 10.31009/LEAP.2019.V7.05

Chasing the Elusive Wrongdoing Intuition

DAVID WASSERMAN
Clinical Center Department of Bioethics
National Institutes of Health, USA

ABSTRACT

In this paper, I offer a partial explanation of the tenacity of the wronging intuition in David Boonin's central non-identity case, Wilma. In Chapter 7 of *The Non-Identity Problem and the Ethics of Future People*, Boonin argues that the claims made against Wilma – that she wrongs someone or acts wrongfully without wronging anyone – are even *less* plausible than the "Implausible Conclusion" – that Wilma does nothing wrong. My arguments complement his, in explaining how the "Implausible Conclusion" is *more* plausible than it may appear. But some of the fault is Boonin's: the needlessly scant detail of *Wilma* provides a foothold for moral complaint. I will consider three ways in which Wilma's conduct, as described, could be regarded as wrongful. First, Wilma could be seen as breaching her duties as the steward of "the gift of life": her reasons for choosing to conceive a child now rather than in two months could be seen as slighting the beneficiary or devaluing the benefit. Second, shifting the focus from benefit to harm, Wilma could be seen as deviously evading her duty to protect her child from harm. Finally, Wilma could be seen as irrationally or impermissibly choosing more over less suffering. Whether or not these can be valid moral complaints, however, they apply to Wilma only if certain assumptions are made about her reasons for not delaying conception, assumptions permitted but not required by Boonin's description of the case.

Keywords: wronging, wrongdoing, procreative reasons, motivation, intuitions, insensitivity.

1. INTRODUCTION

On the second page of his comprehensive examination of the non-identity

problem, *The Non-Identity Problem and the Ethics of Future People*, David Boonin presents an example of the direct version of the problem that he will refer to throughout the book. I will quote it at length, because it will be equally central to my largely sympathetic commentary (the ellipses replace passages that do not concern the type of the condition described or Wilma's decision):

> "Wilma has decided to have a baby. She goes to the doctor for a check-up and the doctor tells her that there is some good news and some bad news. The bad news is that as things now stand, if Wilma conceives, her child will have a disability. The doctor cannot say precisely what the disability will be, but she can tell Wilma three things about it. First, it will be the kind of disability that clearly has a substantially negative impact on a person's quality of life . . . Second . . . this particular disability will clearly not be so serious as to render the child's life worse than no life at all. So while the disability will be considerably far from trivial, the child's life will nonetheless clearly be worth living. Finally, the disability will be irreversible. There will be no way to eliminate or mitigate its effects.
>
> The good news is that Wilma can prevent this from happening. If she takes a tiny pill for two months before conceiving, her child will be perfectly healthy. The pill has no side effects and will be paid for by her insurance company. Fully understanding all the facts about the situation, Wilma decides that having to take a pill once a day for two months is a bit too inconvenient and so chooses to throw the pills away and conceive at once. As a result of this choice, her child [Pebbles] is born with a significant and irreversible disability."

Boonin stipulates that the "significant and irreversible disability" is blindness but invites readers to substitute another condition if they don't believe blindness fits the bill. I doubt it does, but I will accept the stipulation for purposes of my argument, which does not concern the impact of particular disabilities on quality of life but the central question Boonin addresses: does Wilma do anything wrong? Boonin denies that she does, defending the Implausible Conclusion, that Wilma's act of conceiving Pebbles is not morally wrong.

Unlike Boonin, I am not ready to acquit Wilma of wrongdoing, or of wronging Pebbles. My uncertainty persists despite agreeing with Boonin that a woman who chooses to conceive a child with a significant disability now, rather than a child without one in two months, need not wrong the child she ends up having, nor commit any other wrong. I also agree with Boonin that the leading arguments for Wilma's wronging Pebbles are mistaken, some obviously so.

Part of the work in defending the Implausible Conclusion, however, lies in explaining the tenacity of the wrongdoing intuition, whose grip I have not fully broken. In invoking the Modesty Requirement against wrongdoing claims, Boonin is asserting that those claims have implications even less plausible than the Implausible Conclusion. The burden of proving immodesty can be reduced by showing the Implausible Conclusion to be more plausible than it may initially appear. Boonin seeks to do this in Chapter 7; again, I find little to disagree with there (although more than in Chapter 5). His strategy is to compare the case of Wilma and Pebbles *(Wilma)* to other cases of creating or extending worse-off lives. He claims that in these cases, the no-wrongdoing claim is plausible, and that if his arguments in preceding chapters have been successful, those cases are relevantly similar to *Wilma*. But I think he makes his task needlessly difficult by the way he describes his central case. *Wilma* is underspecified in several morally relevant respects; I will suggest that its scant details may pump the very intuition that Boonin devotes so much effort to undermining.

What do we know about Wilma's response to her doctor's news? She apparently accepts the pills initially, then discards them because she finds taking them "a bit too inconvenient." She seems to recognize that it is biologically impossible for any child conceived at least two months later to be the same—by standard identity criteria—as any child she could conceive now.[1] That's all we know. We aren't told what she thinks about the prospect of having a blind child, though it's safe to assume she believes there is some reason not to. Otherwise, why bother taking the pills from the doctor? We can also assume that she takes her inconvenience as outweighing that reason, whatever it may be; this is not a case of akrasia or impulse. Can we flesh out her attitudes in a way that would warrant the claim that she wrongs Pebbles or (what may or may not be equivalent) commits a wrong involving Pebbles? I will suggest that we can.

There are certainly ways of fleshing out *Wilma* that preclude this claim; my point is emphatically not that Wilma necessarily wrongs Pebbles, but that she might be reasonably construed as having done so in the case as described. I will, so to speak, assess that case in the light most favorable to the prosecution. In Section 2, I will take an approach similar to Boonin's in Chapter 7 for a different, if not opposing, purpose: I will compare *Wilma* to cases in which in which a donor chooses whom to benefit for reasons that wrong, or at least slight, the beneficiary. I will argue that those cases are

1 As Melinda Roberts suggests in email correspondence, Wilma might have the same child if she could preserve the gametes for two months and then employ IVF, but to preserve an essential feature of the case (and avoid debates about necessary conditions for identity), I'll assume that any child so conceived would also be blind.

relevantly similar to Wilma's, and that they suggest why she, too, might be regarded as slighting or wronging Pebbles in the case as described.

Of course, *Wilma* is just one of the standard hypotheticals used to illustrate the direct version of the non-identity problem, and it might seem that I have placed too much weight on its details—or lack of details. But Boonin adopts this case as his central example, which he refers back to throughout the book. It is also, as he notes, a slight variation on a case used by Derek Parfit (1984) to introduce the non-identity problem. Moreover, the resilient intuition that Wilma does something wrong may suggest that even if she can acquit herself of that charge, she has the burden of defending herself against the presumptive wrong.[2]

I will consider three ways in which Wilma's conduct, as described, could be regarded as wrongful or morally deficient. First, Wilma can be seen as breaching her duties as the steward of a distinctive kind of precious gift. Her reasons for choosing to conceive a child now rather than in two months can be seen as slighting the beneficiary or devaluing the benefit. Second, shifting the focus from benefit to harm, Wilma can be seen as deviously evading her duty to protect her child from harm, as Robert Noggle suggests in a recent article (Noggle, 2018). Finally, drawing on a suggestion made by Kieran Setiya, Wilma can be seen as irrationally or impermissibly choosing more over less suffering. Even if Wilma were guilty of any of these offenses, it would not account for the intuitive judgment that she does something as, or almost as, wrong, by having a blind child as she would by blinding an already existing fetus or child. But I'm convinced that this judgement is either deeply confused or (perhaps) worse, impersonally consequentialist.

The first claim of wrongdoing focuses on the good of the life that either Pebbles or Rocks would enjoy; it involves seeing Wilma, and prospective parents generally, as the stewards of a precious gift, one which the individual need not confer, but which can be conferred in ways that slight or disrespect the gift or its recipient. The second and third claim focus on harm rather than benefit: on the blindness that Pebbles experiences and that Wilma could have avoided by waiting to conceive. The second sees Wilma as acting wrongfully by creating an excuse for breaching her duty to protect her actual child from blindness – by rendering the fulfillment of that duty impossible. The third claim is that Wilma shows insensitivity to

2 This may suggest an asymmetry in direct nonidentity cases: that making a choice which leads to having a worse-off rather than a better-off child imposes a burden of justification to the resulting child that is not faced by a choice with the opposite result. Wilma as described does not clearly satisfy that burden, and while she could, it is not a burden she would face if she chose to wait. If a non-identity problem survives Boonin's sustained critique, it may lie in explaining this apparent asymmetry in the burden of justification.

the harm of blindness in the way she responds to the prospect of that harm, choosing to subject her child to it for trivial, selfish reasons.

I doubt that Wilma acts wrongly in either of the first two ways and remain uncertain about the third. I will defend the more modest claim that if she does, it is only in virtue of the reasons for which she chooses to conceive now. This modesty reflects real uncertainty on my part about how to evaluate procreative or gestational decisions made without the "due diligence" and child-centered motivation that I believe all children brought into a harsh, dangerous world are owed. Despite this uncertainty, I am convinced that parents who bring children into being with such reflection and motivation commit no wrong whatsoever if they have a child they expect to be substantially less well-off that another they could have had – as long as they reasonably expect the child to have a good life, are committed to securing that outcome, and do not expect to significantly harm any third party.

I am also uncertain about another critical point: whether any wrong on the part of Wilma-as-described can be regarded as a wrong to Pebbles, giving her a complaint, or is rather a "victimless wrong," about which Pebbles has no special standing to complain. With reference to Boonin's overall argument, I will assume that Wilma's act harms no one, and that the case against her rests on the denial of one of two premises leading to the Implausible Conclusion: P4, "If an act does not harm anyone, then the act does not wrong anyone" or P5, "If an act does not wrong anyone, then the act is not morally wrong." The wrongdoing claims I will consider sometimes pose more of a challenge to P4, sometimes to P5. It is, however, not always easy to distinguish these challenges, a difficulty I will discuss in examining all three claims, especially the first. But I will argue that if Pebbles does have a complaint, contra P4, it is not one that the great value she finds in her life compels her to waive. That is because Wilma could have conceived her for reasons that would have given Pebbles no grounds for complaint.

2. SLIGHTING THE RECIPIENT OR THE GIFT

I In one of the earliest attempts to find a wrong in direct non-identity cases, Greg Kavka argued that it lay in the abuse of the progenitor's reproductive powers. He understood the wrong of such abuse in terms of a modified categorical imperative, that would "forbid treating rational beings or their creation . . . as a means only" (110). Although his account has faced a great deal of criticism, there is a resilient appeal in the idea that the awesome

power to create life comes with significant moral constraints—if not, or not only, the one Kavka identified. I want to explore the constraints on reproductive power that may be imposed by regarding prospective parents as the stewards of the "gift of life" they confer by conception and gestation. Of course, they are not stewards in any ordinary sense, because, at least on the prevailing view, they have no duty to confer the gift at all. Rather, they assume that role only if they decide to bear a child. And, of course, they do not give the gift of life to pre-existing recipients; that gift, so to speak, creates its own recipients.

In order to treat prospective parents this way, we must distinguish the gift from its recipient. In doing so, however, we may appear to treat the gift as "mere existence." We seem to strip the gift of its distinctive characteristics by treating it as something that can have different possible recipients. Although the life given is itself a particular, in the sense that that Pebbles and Rocks would receive different gifts, because they would have numerically different lives, those gifts would not differ qualitatively because they have no qualities beyond conscious existence.

But this view of the gift as bare existence is impoverished. Prospective parents, as opposed to sperm-donors, not only bring children into existence; they bring them into their lives, families, communities, and their time and place, making the gift they give much less generic. The setting for the child's life is not something distinct from the gift; a network of intimate relationships and a rearing environment are an integral (if severable and modifiable) part of it. At the same time, the character and value of the life given will obviously depend on its recipient. A life with the potential for rapturous experience and great achievement could not be given to a recipient with the potential for nothing more than mild euphoria and mediocrity. (Assessing potential is notoriously difficult, conceptually and empirically, but I leave those issues aside.)[3]

3 These qualifications are critical if we are to frame the evaluation of Wilma's conduct in terms of the moral constraints on the donor of a precious gift. As I will discuss later, such donors can be regarded as indiscriminate or finicky in choosing the recipients. But it would be difficult to regard a donor as guilty of either excess when the value of the gift he appeared to under- or over-value varied substantially with the recipient. Similarly, it would be difficult to regard the donor as displaying an objectionable attitude toward a type of potential recipient if the type of potential recipient could only be identified in terms of the gift they could receive. For example, a female child might reasonably feel wronged by her parents if she learned that she had been gestated only because all of their male IVF embryos failed to implant. To regard her parents as having violated a moral constraint on the stewards of a precious gift—that they should not show sexual bias in bestowing it—it is necessary to regard the gift itself as gender-neutral, so to speak: the gift of life as a male is not one she could have received. (Perhaps "she" could have in the sense that a gamete or embryo with a Y chromosome could have been modified to have an X, which would be identity-preserving for non-essentialist about sex. But we can assume that was not a practical possibility in the case described.)

If we can make sense of the notion of prospective parents as stewards of a precious gift, perhaps we can regard wrongful procreative decisions as violations of the duties imposed by stewardship. In some cases, the violation will slight only the gift, not the recipient; in others, it will give the recipient a complaint. In non-procreative contexts, the distinction is clear. If an art collector carelessly failed to appreciate the immense value of a painting, regarding it as a pleasant decoration, but gave it to her favorite nephew, it's not clear the nephew would have a complaint. A recipient need not be slighted in being given a gift that the donor mistakenly regards as a trifle, unless the donor regards him as worthy only of trifles. In contrast, a collector who appropriately valued the painting would slight the recipient if he gave it to a despised nephew, just to see what he would do with a work of art the collector believed he lacked the capacity to appreciate. (He might also slight the painting in this case by placing it in such philistine hands.)

Admittedly, it may sometimes be hard to tell if the donor slights the gift or the recipient. When one casts one's pearls before swine, one may be devaluing the pearls, the recipients, or both. The biblical injunction, taken literally, suggests the former, but the way in which the donor dispenses her pearls, say by casting them into manure for the recipients to root out, could slight the recipients as well, by treating them like swine.

What, then, are the moral constraints on the stewards of the precious gift of life, and can Wilma be said to violate any of them? And when does the violation of those constraints wrong the recipient, as well as, or as opposed to, the gift? Clearly, if there is a victim of the wrong, it must be Pebbles, at least if we take the conventional view, as I will, that merely possible people like Rocks have no standing to complain about the decisions, however flawed, that deny them actualization.

One place to start would be to treat a procreative steward, like a fiduciary, as obligated to act only in the interests he was charged to serve—for the steward, the gift of life. A procreative steward would violate his duties if he chose how to dispose of the gift for which he was responsible on the basis of his personal convenience. It might be held that Wilma abused her stewardship by violating a duty not to choose the recipient of her gift on the basis of trivial, selfish reasons.

But this is far too demanding a standard for procreation. After all, some couples try to time pregnancy to spend the final months in the relative comfort of spring or fall, rather than the harsher temperatures of summer or winter. They do so recognizing, or able to recognize, that their timing is identity-affecting. They are permitted to time their pregnancies for their

own convenience, as long as they are willing and able to care for the child conceived at the chosen time. In doing so, they do not devalue the precious gift they bestow.

In contrast, Parfit's 14-year-old girl, who has a child before she is ready to raise one, may fail to display adequate appreciation of the gift she is conferring. (I emphasize "may" because so much depends on the attitudes and beliefs she has in deciding to conceive.) She may also wrong the child, I would argue, by imposing an unacceptable risk that she would be unable to fulfill her responsibilities to it.[4]

It is less clear, however, that Wilma displays a similar failure to appreciate the immense value of the gift in choosing, for her own convenience, to bestow it on a blind rather than a sighted child. Although she presumably prefers a sighted child and believes it would have a better life, she hardly regards a blind child as an unworthy recipient, since she believes it would also have a good life and seems willing and able to raise it. Given these beliefs and attitudes, she would hardly display deficient appreciation for the gift in bestowing it on a blind child if she could not have a sighted one. Indeed, she would not value the gift appropriately if, after deciding to have a child, she changed her mind upon learning that she could only have a blind one. As I argue later, she would fail to recognize that the immense value of the gift to a sighted *or* blind child dwarfed any difference in its comparative value.

But even if Wilma need does not show a deficient appreciation of the gift, it might be thought that she slights Pebbles by the way she confers it on her, overriding a preference for another type of recipient for a reason as trivial as her temporary convenience. Of course, giving a benefit to a less-preferred recipient need not slight the recipient. Someone offered a position after two more favored candidates had turned it down could have no complaint, even if he was hurt by the discovery. Getting a gift on the rebound does not itself seem to be a slight to anyone. In contrast, no one rejects Wilma's gift of life to a child she would conceive in two months; rather, she herself decides not to give the gift to a preferred recipient because of her own convenience. Pebbles might complain that if Wilma had really preferred a sighted child, she should not have overridden her preference for a trivial reason. But this seems a perverse complaint: the fact that Wilma did suggests that her preference against having a blind child was not very strong. And this seems to weaken, not strengthen, any claim of insult to Pebbles.

4 This claim raises questions, which I cannot address here, about when, or whether, one can violate the rights of a child who does not yet exist, and whose existence would not have been possible without a rights violation.

There are cases in which overriding one's initial preference for recipient-type does seem to wrong the recipient, but these are cases in which the reasons for disfavoring the actual recipient are disrespectful, and where overriding them seems to add further insult. Thus, a Black diner would be wronged if he were admitted to a segregated restaurant that took down its "Whites Only" sign when exclusion became "a bit too inconvenient" because of slow business. In a procreative context, similar offense could be taken by a female child whose parents had strongly preferred a male child, and who had been gestated only after all the male embryos in the IVF array had failed to implant. But claiming that kind of wrongdoing by Wilma would only be plausible if her initial willingness to take the drug and delay conception was based on prejudice or aversion towards blind people, rather than on the belief—which, recall, I've accepted for the sake of argument—that a sighted child would have a much better life than a blind child. If Wilma would have had the same preference against having any kind of child expected to be significantly less well-off, regardless of the source of its hardships, it would be difficult to find bias.[5] Wilma could only be faulted if, in deferring to her own convenience, she displayed a callous insensitivity to the suffering that accounted for this difference in expected well-being, a suggestion I consider in Section 4.

A final suggestion involves seeing the stewardship obligations of prospective parents in Aristotelian terms, as requiring them to avoid the extremes of indiscriminateness and finickyness in conferring precious gifts. This suggestion has some kinship to Rosalind Hursthouse's (1991) argument that abortion can be assessed in virtue-ethical terms, and that woman may reveal a kind of vice in aborting for reasons that reflect an insufficient appreciation of the value of human life or of parenthood. Unlike Husthouse's account, though, the approach I'm proposing seeks to find a wrong in the act (conceiving now), not, or not only, in a vice of the agent.

Prospective parents would arguably display indiscriminateness if they gave the gift of life to someone for whom they believed it would have only marginal value—someone whose life they expected to have so much pain or so little pleasure that it would barely be worth living. The steward of a precious gift should only give it to someone for whom it would be of significant value. To be willing to give the gift to anyone for whom it would not be bad, even someone who would have a life barely worth living, is to devalue it, to care too little for its robust appreciation. But Wilma does not even come close to that extreme: she expects Pebbles to have a life well

5 This wrongdoing claim was suggested, although not endorsed, by Melinda Roberts, email correspondence.

worth living, well above any reasonable threshold for adequate enjoyment. Clearly, she would have devalued the gift if she regarded a blind child as barely capable of benefitting from it: it's not the actual value of the gift, but what she believes it to be, that is relevant in assessing her stewardship. But there's no indication she holds this dubious view.[6]

On the other hand, it would be both presumptuous and finicky to limit the gift of life to the recipient for whom it was expected to have the most value: presumptuous in assuming that one could make more than a rough guess about the comparative value of existence for different people; finicky because the value of the gift for many or most possible recipients would be great enough to dwarf the differences among them. To insist that the gift of life go to the candidate who can most enjoy it is to treat that gift in an inappropriate way, like a Stradivarius that should only be given to the most talented violinist. It is appropriate to be highly discriminating in bestowing an exquisitely crafted musical instrument to the most talented player because respect for the instrument requires a recipient who can fully exploit its musical potential. Moreover, a less talented violinist, even if he would have enjoyed playing it more than the prodigy who gets it, can get great enjoyment from a lesser instrument. The contrast with the gift of life is almost too obvious to spell out—nothing else can be enjoyed and only a noxious perfectionism would demand that life to the recipient who can achieve the most with it, or enjoy it the most. But there is no indication that Wilma's initial preference to delay conception to have a sighted child reflected any such finickyness or perfectionism.

Although in the end it is hard to justify, the impression that Wilma slights either Pebbles or the benefit she bestows by bearing her arises only because of the manner in which Wilma responds to her doctor's advice. No such impression would have arisen had she declined to delay conception, not because it was inconvenient, but because she was willing and able to raise a blind child. Her willingness need not have reflected the expectation that a blind child would have a life just as good as, or on a par with, that of a sighted child. It might have instead reflected the belief that even if the difference in well-being was substantial, it would be dwarfed by the goodness of either life, as well as the recognition that each child had "but one life to live." Wilma would not have slighted either the gift or the recipient in conceiving a blind child with such attitudes and expectations. Rather, she would have achieved a morally attractive mean between the extremes of indifference and finickyness.

6 It may also be the duty of prospective parents as stewards to carefully appraise that value, but that is a distinct issue.

3. DEVIOUSLY EVADING AN OBLIGATION

Robert Noggle has recently offered an account of Wilma's wrongdoing as (primarily) victimless, which focuses on the harm of blindness, not the benefit of a rewarding life (Noggle, 2018). It begins with the claim that Wilma has a duty to protect her children from blindness and the recognition that in Pebbles' case, she is excused from fulfilling that duty by the impossibility of doing so. But she is not off the hook, since she has made it impossible to fulfill that duty: "[D]eliberately making it impossible to fulfill a moral obligation manifests a defective attitude toward morality – an attitude which sees moral obligations as something to be dodged whenever they are inconvenient. . . . [Acting] on this attitude is a wrong-making feature that is independent of any wrong that might be done to Pebbles" (1). Noggle identifies Wilma's wrong as the violation of this Principle of Deliberate Impossibility (PDI), placing her in a rogue's gallery of duty-dodgers who make it impossible for them to fulfill an onerous duty.

Noggle's spells out his argument in four premises and a conclusion (8):

> "1. Wilma is obligated to protect any actual child she has from blindness (Assumption).
>
> 2. It is morally wrong for Wilma to deliberately U at t1, if Wilma expects her U-ing to make it impossible to W at t2, where Wilma's W-ing at t2 would have been obligatory had Wilma's U-ing at t1 not made it impossible for Wilma to W at t2, and the impossibility of her W-ing is the direct cause of its being nonobligatory (PDI).
>
> 3. Wilma expects her forgoing the medicine and conceiving at t1 to make it the case that, at t2, it is impossible for her to protect an actual child of hers from blindness.
>
> 4. Protecting this actual child from blindness would have been obligatory for Wilma at t2 if her act (at t1) of conceiving a child with incurable blindness had not made it impossible (and its impossibility is the direct cause of its being nonobligatory).
>
> 5. TTherefore, it is wrong for Wilma to deliberately forgo the medication and conceive Pebbles."

The weak link in Noggle's argument is Step 4. My bracketed insertions specify its application to Pebbles as well as Wilma:

> "Protecting this actual child [Pebbles] from blindness would have been obligatory for Wilma at t2 [when she acquires responsibility for Pebbles'

> health and welfare] if her act of conceiving a child with incurable blindness had not made it impossible (and its impossibility is the direct cause of its being non-obligatory)".

The problem with this claim is that it's not the case that Wilma's act of conceiving Pebbles made it impossible for her to protect *Pebbles* from blindness, at t1 or t2, regardless of whether t1 = t2. It's not the case that had she not conceived Pebbles, it would have been obligatory for her to protect Pebbles from blindness, for the simple reason that there would have been no Pebbles to protect: it would have been impossible to protect Pebbles from blindness if she did not exist. It's true that if Wilma had not conceived an incurably blind child, but instead, say, the sighted Rocks, she would have been obligated to protect Rocks from blindness. But her actual child, per Step 1, is Pebbles, not Rocks. If Step 1 were modified with a definite description, like "her present and future children," then Noggle's argument would become just another version of the failed strategy of charging Wilma with violating a supposed de dicto duty to "her next child" – a strategy Boonin challenges in Chapter 2 of his book and which I address in the next section. Wilma could have fulfilled such a de dicto duty only by conceiving a different child than Pebbles, so it is hardly a duty whose breach would compel Wilma to offer an impossibility defense to the charge that she wronged her actual child Pebbles. It was never a duty she owed to Pebbles, and even if it somehow was, we can safely assume that Pebbles would have waived her right to its fulfillment.

Noggle also engages in a bit of questionable intuition pumping. The duty-evasions in the hypotheticals he presents to motivate PDI are objectionable in part because they involve harming or wronging third parties who exist independently of the duty evasion. Because Slippery Pete and Weaselly Willy (two of Noggle's duty evaders) have made themselves unable to take their turns teaching an 8:00 am course, the other faculty members have to do additional teaching. An independent wrong to third parties can also be found in the case of Clare, who, by developing the new elective course requested by her Chair, which makes it impossible for her to prevent this additional course from being taught by adjunct faculty in an exploitative arrangement. Clare thus contributes to, and is complicit it, the wrongful exploitation of adjunct faculty. There is no counterpart to this in Wilma's case – it's stipulated that no independently existing person suffers a harm or wrong as a result of her choice.

Finally, Noggle's application of PDI would require prospective parents to avoid having, or selecting against, very minor disabilities. If parents have a duty to protect their actual child against blindness, which they would deviously evade by having an incurably blind child, they also have a

duty to protect their actual child from very minor impairments, which they would deviously evade by having a child with very minor incurable disabilities. But many or perhaps most of those who fault Wilma would not fault, even to a lesser degree, prospective parents who declined to select against very mild asthma or astigmatism. Indeed, they might fault as finicky a prospective parent who did select against such minor impairments. Yet they would condemn parents who, because it was a bit to inconvenient, declined to take easy precautions to prevent their actual child from becoming very mildly asthmatic or astigmatic. Parents do have a duty to protect their actual children from very minor disabilities, but they would not wrongfully evade that duty by having a child who could not exist without such a minor disability.

Even if Wilma could be said to have deviously evaded her duty to protect her child from blindness, that could hardly be said of a woman who declines to delay conception just because she was ready and able to raise a congenitally blind child. Indeed, she might find it far more inconvenient to have a blind child than to delay conception for two months and have a sighted child. She might be completely willing to accept that additional inconvenience, expecting that any child she has now will have a life well worth living and recognizing that it would not exist if she delayed pregnancy. She might welcome, and actively seek, a way to prevent or treat that child's blindness. She could conceive now while being committed to having prenatal surgery to prevent or treat blindness, were such surgery to become available during her pregnancy – even if that surgery would be far more burdensome than a two-month delay in conceiving. There is nothing devious in her intentions or attitudes; no duty that she is shirking. The appearance, which Noggle seeks to account for, that Wilma fails to take her duty as a prospective parent seriously, may reflect not her evasion of a parental duty to prevent her actual child's hardships, but rather her failure to show an appropriate concern for those hardships.

4. DISPLAYING INSENSITIVITY

Many people would likely judge that Wilma acted wrongly, and wronged Pebbles, just because she was indifferent about having a blind or sighted child. That indifference might be regarded as the source of the wrong to Pebbles. The focus of this claim, like Noggles', is on the harm to which Wilma would subject or expose Pebbles, however unavoidably. If, as I've been assuming, life would be much harder for a blind child than a sighted one, indifference between the two future lives displays insensitivity to the greater hardship.

This complaint would seem even more plausible if Pebble's condition had a more clearly adverse impact on her well-being, e.g.., cystic fibrosis (CF) rather than blindness. The complaint would indict the mother for insensitivity to the substantial hardships of life with CF – the breathing difficulties, the disruption of activities, the increasing dependence on a difficult medical regimen, and, ultimately, premature death. The complaint does not assume that it would be wrong for a woman to have a child who lacked sight, or had CF, when she could, merely by waiting, have a different child who would have sight or not have CF. But it suggests that she must have a good reason for doing so.

In a paper published the year before *The Non-Identity Problem and the Ethics of Future People*, Kieran Setiya distills a complaint about insensitivity that is often made about hypothetical parents like Wilma:

> "[T]he fact that your child will suffer as a result of your action is a reason against it and a reason to prefer the situation in which you wait. If you have a child later, she will suffer too – everyone does – but most likely she will suffer less. Since there is no other reason to have a child now, and no other relevant difference between these options, you should wait to have a child, if you have a child at all" (Setiya, 2014, 291).

Now . Wilma does have a reason for her choice – her own convenience – but Setiya likely means a good, moral, or appropriate, reason (earlier on the same page, he stipulates that you have no 'urgent" reason.). Construing those terms in any plausible way, Wilma lacks such a reason. The question, however, in why she needs one for not waiting, but not for waiting, given that her choice will bring different children into existence. The answer Setiya gives abstracts the suffering from its bearer, and in doing so, echoes both de dicto and impersonal approaches to NIP.

The idea seems to be that just because Wilma could only apprehend any future child de dicto at the time she declined to select, she could only apprehend the harm to be selected against, and not the actual child who would bear it. Her refusal to delay conception could not be explained by her love or partiality for a particular child who would have a life worth living despite that harm. Given her epistemic and psychological limitations, Wilma's refusal to wait and thereby prevent the additional suffering can only be regarded as displaying insensitivity to it. The actual child wronged by that insensitivity has a complaint, even though she may be glad for the existence she could not have had without it.

I see this as the underlying logic of several recent arguments for wrongdoing. Perhaps the clearest example is Nicholas Vrousalis' (2013)

claim that worse-off children whose existence results from a failure to screen are "smuggled into existence," with a parental attitude that is disrespectful because of its failure to show due concern for the child's harm. This logic has affinities to contractualist approaches like Rivka Weinberg's (2016), which seek to take existence "off the scales," and those, like Rahul Kumar's (2003), which maintain that the relevant principles concern types of people, not tokens.[7] With existence off the scales, so is the distinct numerical identity of different future persons. All that's left is the simple choice between less and more suffering, about which it would be perverse to be indifferent. In much the same way, we could reasonably reject a principle that did not enjoin us to make our "future children" – a type – more rather than less happy.

I suspect that a similar logic lies behind the explanation prospective parents often give for their decision to select against disability: they do so for the welfare of "the child" because they want to prevent "their child" from having a harmful condition. These individuals are, at other moments, fully aware of the basic facts of reproductive biology, or of what Kavka (1982) called "the precariousness of existence." It is surely more charitable to assume that they find these facts of little or no relevance to their decisions than that they simply fail to grasp them. This is especially so when the distinct identities of the future children they could have are made salient by IVF selection. It hardly requires a high level of biological or philosophical literacy to recognize that a distinct child (or children) would develop from each embryo. But I doubt that the concern expressed about the welfare of "the child" is reduced much, if at all, by this recognition.

Clearly, Setiya himself recognizes that the child who suffers more would be a different person than the child who suffers less. His claim seems to be that this cannot supply an adequate reason for the woman to conceive now and bear the former. But this claim rests on an untenable assumption: that although the woman is, and knows she is, making a choice between two

7 Kumar seeks to find a wrong in Wilma's decision in its violation of a principle that no one, even Pebbles, could reasonably reject, given the level of generality at which it is framed. He argues that the appropriate principle will concern types of people, here parents and children. Such principles are familiar for other types, like drivers and pedestrians, types of which we can be tokens at different times. A principle prohibiting drunk driving could not be reasonably rejected, since it protects interests in life and limb over relatively trivial interests in convenience. In driving drunk, one puts pedestrians (as well as other drivers, passengers, etc.) – whoever they may be – at grave risk of injury. Kumar claim that by analogy, prospective parents put future children at risk by not avoiding the creation of children with additional burdens. But the analogy is flawed: while an act of drunk driving risks harm to particular individuals, even if they are now unknown, no one is subject to a risk of harm by a failure to prevent the creation of children with additional hardships, because that omission does not harm the child created. No possible harm, no risk.

different children (or non-overlapping sets of possible children who can become actual), she can only apprehend each of the future children she is choosing among as a generic future child – a "bare subject of predicates" (to use Bernard Williams' memorable phrase, Williams, 1973) who will either have or lack additional suffering. This might be true of Wilma: from what we are told, she knows nothing about the future child she would have now or the one she would have in two months except that the former will be blind, the latter sighted, and so she seems to have no good reason to have a child who will "suffer" more. But this characterization rings false for many women who would decline to delay – a point I will return to shortly.

Setiya suggests the kind of position I'm attributing to him in another paper, where he presents a case, inspired by Caspar Hare (2013), involving a choice between equally-priced concert and theater tickets. If you think you'd enjoy the play as much as the music, it would be rational to prefer either one. It would still be rational if the price of one but not the other went up a dollar. Before the "sweetening" provided by the extra dollar, the reasons favoring the two performances were not strictly equal, merely on a par -- the reasons for favoring each one were incommensurable.[8] Those reasons remain on a par even with the extra dollar. But if each ticket were held in an opaque box, with a dollar attached to only one, you should prefer that one. It's only when you can apprehend the alternatives as particulars – the concert and the play – that the extra dollar becomes what Frances Kamm (1987) has called "an irrelevant utility" (Setiya, 2019).

I suspect that something akin to this idea lies behind Setiya's claim about the two future children: they are like the opaque boxes, distinguishable only by the additional suffering that is attached to one. Given this profound limitation in how one can apprehend their lives, it would be immoral to select the life to which the additional suffering was attached. It would display an indifference or insensitivity to that suffering – the only discernable difference between the two lives. This might be a fair way to describe the way Wilma apprehends the lives of the children she chooses between. Although she expects that both would have good lives, those lives are opaque to her. But one of those lives comes with what Wilma herself regards as a significant harm – and that is the one she decides to have, for no good reason.

Moreover, since all prospective parents face these profound limitations in apprehending the particular lives their future children would lead, and

8 This incommensurability need not reflect anything as profound as "incommensurable values." The goodness of the two events, or artistic genres, differ in so many aesthetic respects that a full ranking is impossible.

the particular individuals they would become, it might seem that any prospective parent who declined to delay conception to avoid blindness would, in the absence of a good reason, show similar insensitivity toward the (stipulated) harm of blindness, no matter how willing and able they were to raise a blind child. But this conclusion exaggerates the opacity of future lives to prospective parents. While it is obviously not possible for them to apprehend those lives in their full particularity, it is possible for them to see those lives as more than receptacles for differing amounts of well-being and suffering. This is in part because, as I argued in Section 2, prospective parents are not bestowing "mere existence"; they are, in normal circumstances, bestowing a life with a particular family in a particular place and time. They can anticipate in some detail – if not always accurately -- how the lives of a blind or sighted child would go, and how they would go differently.

Seen this way, the limitations faced by such prospective parents differ mainly in degree from those faced by the person choosing between the play and concert. Those performances are future events, which can be apprehended only by imagining and extrapolating. The prospective parents can apprehend some aspects of an uncertain future life; the box is, so to speak, translucent. Like the person choosing between the play and concert, they can envision the future, but only "through a glass, darkly." The person choosing a ticket can imagine the two performances; the prospective parent can imagine the lives of their possible children in the environment they expect to rear one or the other. Of course, their speculation could be far off the mark – the life of either child will be full of unpredictable developments. But the person choosing a ticket is also speculating, and his speculation could also be off the mark. The person making the choice may never have been exposed to either group or work; indeed, this could be the group's or the work's debut. Even if he is acquainted with both, the performance he experiences may fall far above or below his expectations. The prospective parent's choice thus may be more like that between the tickets than the opaque boxes holding the tickets.[9]

Seeing the life of a blind child this way would not involve judging it to be on a par with that of a sighted child – the analogy to the tickets does not go that far. Rather, it involves seeing the life of a blind child as rich and rewarding enough to make the difference in expect well-being morally irrelevant. Prospective parents need not regard the lives of a blind and

9 The difference is, admittedly, significantly greater in some of Hare's own cases, where the choice is between two fully accessible objects. But while those objects can be, in one sense, fully apprehended by the chooser, he must speculate to some extent about how the acquisition of one or the other will matter to him in the future. Perhaps the most dazzling one will lost its luster, while the plainer one will grow on her.

sighted child as incommensurable in the same sense as the play and concert; they could regard blindness as a "bad difference" without regarding it as a reason for choosing one future child over another.[10] They can apprehend the lives of their children as far more than mere existence with or without sight. Their understanding, however imperfect, of how their future child's blindness could be part of a life well worth living would suffice to rebut a charge of insensitivity.

The difference in expected goodness, while greater between the blind and sighted child than between the tickets with and without a dollar attached, may be even less relevant to the choice in the former case. If we raise the sum attached to a hundred dollars, it may become relevant – the two tickets may no longer be on a par, and it would no longer be rational to choose either. In contrast, the disutility "attached to" one child is already substantial, and the two lives are not on a par prudentially. But they do not need to be. What makes them incommensurable is not their expected well-being, but their separateness – a separateness that prospective parents are capable of appreciating more fully than Setiya acknowledges.

Thus, Peter Herissone-Kelly (2007) suggests that prospective parents should decide when to conceive or which embryo to implant by adopting an "internal perspective," on which they "imaginatively inhabit" the life of each future child to see if each, from its own perspective, would have acceptable prospects. If it would, then they have no moral reason to select another child with prospects judged better from an external perspective. From the perspective of the first child (or set of possible future children who could be conceived now), the fact that a different child conceived two months later would face less hardship hardly makes its own prospects less acceptable. If, as assumed, a blind child's life would have acceptable prospects, the internal perspective would give prospective parents no reason to disfavor such a child. The fact that they could have had a sighted child does not affect the blind child's prospects, or her retrospective appraisal of her own life.

It may appear that a stubborn asymmetry remains. Even if an imaginative exercise like that prescribed by Herissone-Kelly justifies a prospective parent in refusing to select against a worst-off child, no such exercise may seem necessary for a refusal to select against a better-off child. Wilma doesn't appear to owe Rocks a reason for waiting two months to conceive; alternatively, an adequate reason can be found in the mere

10 But given her knowledge of, and experience with, blindness or CF, would it be acceptable for her to base her decision on whether to have a child with those conditions on her preference for giving birth in fall rather than winter? I have conflicting intuitions, which I hope to explore in the future.

fact that a child conceived after two months would not be blind. But perhaps that mere fact would not be enough for Rocks. If Wilma had been indifferent about whether she conceived a sighted or blind child, or even a child with or without CF, and decided when to conceive solely only the basis of her comfort or convenience, she could also face a complaint of insensitivity from a sighted child, or a child without CF. The latter child, for example, could plausibly complain, "I'm relieved that I don't have CF, but it's little thanks to you. You didn't care if you had a child without or without CF, and your only reason for having one without it was that you preferred a later birth!" His complaint would not be grounded merely in his mother's willingness to have a child with or without CF, but in the fact that her willingness reflected her lack of interest in, or concern about, the hardships of that condition. Had her willingness to have a child with or without CF arisen from the kind of exercise in n anticipatory empathy prescribed by Herissone-Kelly, he would have little or no ground for complaint.[11]

5. CONCLUSION

In Section 2, I suggested that prospective parents could be seen as the donors of a precious gift, required to choose the recipient responsibly, on pain of slighting the gift or the recipient. In Section 4, I suggested that prospective parents should not bring a child into existence without some sense of what their child's life would be like. Intentional but thoughtless conception, in contrast to merely accidental conception, may show a kind of disrespect for the future child – even if it does not result in, and does not even predict, any dereliction of parental duty. Although I doubted that Wilma could be seen as slighting either the gift or recipient, I found it plausible to claim that she displayed insensitivity to the (stipulated) hardship of blindness in choosing to have a blind child without seriously considering what life would be like for such a child. I also found that such insensitivity would give Pebbles, the blind child she ended up having, grounds for complaint.

11 If she had gone through that exercise, and judged the prospects of a life with or without CF or blindness to be acceptable, or if she had knowledge of, and experience with, CF or blindness, would it be permissible for her to base her decision on whether to have a child with those conditions on her personal preference for conceiving at a specific time of year? I have conflicting intuitions. Given the substantial differences in the lives of her possible children, and her life in raising them, there still seems something inappropriate about deciding which to conceive on the basis of her short-term comfort or convenience, rather than at random, or by "first come, first served." This is not to say that it would be wrong in any way for her to decide on the basis of substantial burdens or hardships she reasonably expected to face in raising one child or the other.

But if Pebbles does have a complaint, the goodness of her life would give her no reason to waive it. Her grievance would not be about her blindness, which is an unavoidable aspect of a life she values. Rather, it would concern Wilma's failure to bring her into existence with, in my earlier phrase, due diligence. Unlike preventing the existence of a blind child, that moral exercise would not have required Wilma to delay conception, and there are no grounds for regarding it as identity-affecting (Roberts, 2007). The ex ante odds of Pebbles coming into existence would have been the same if Wilma had declined to delay for reasons consistent with due regard for the benefit she was conferring, adequate respect for its recipient, and keen appreciation of the expected hardships of the life she was creating. I think Boonin would be mistaken if he denied, as he appears to do in his Appendix on the rights waiver objection, that Pebbles had a moral reason to press her grievance about Wilma's failings in any of these respects. To adopt the language of Section 2, Pebbles could value the precious gift of her life while disapproving of the reasons for which the donor conferred it on her. Her disapproval would be compatible with her deep appreciation of the gift..

BIBLIOGRAPHY

Benatar, D. and Wasserman, D., 2015. *Debating procreation: is it wrong to reproduce?* New York: Oxford University Press, Ch. 8.

Boonin, D., 2014. *The non-identity problem and the ethics of future people*. New York, Oxford University Press, USA.

Hare, C., 2013. *The limits of kindness*. Oxford: Oxford University Press, Chs. 3, 4.

Herissone-Kelly, P., 2007. Parental love and the ethics of sex selection. Cambridge Quarterly of Healthcare *Ethics*, 16(3): 326-335.

Hursthouse, R., 1991. Virtue theory and abortion. *Philosophy & Public Affairs*, : 223-246.

Kamm, F.M., 1987. The choice between people: "common sense morality and doctors. *Bioethics*, 1(3): 255-271.

Kavka, G.S., 1982. The paradox of future individuals. *Philosophy & Public Affairs*, : 93-112.

Noggle R. 2019. Impossible obligations and the non-identity problem. *Philosophical Studies*, 176(9):2371-90.

Kumar, R., 2003. Who can be wronged? *Philosophy & Public Affairs*, 31(2): 99-118.

Noggle, R., Impossible obligations and the non-identity problem. *Philosophical Studies*, 1-20. https://link.springer.com/article/10.1007/s11098-018-1130-7. Accessed July 5, 2018.

Parens, E. and Asch, A. eds., 2000. *Prenatal testing and disability rights*. Washington, DC: Georgetown University Press.

Parfit, D., 1984. *Reasons and persons*. OUP Oxford.

Roberts, M.A., 2007. The Non-identity fallacy: harm, probability and another look at Parfit's depletion example. *Utilitas*, 19(3): 267-311.

Setiya, K., 2018. Ignorance, beneficence, and rights. *Journal of Moral Philosophy*, published online, February 28, 2019, 1-19.

Setiya, K., 2014. The ethics of existence. *Philosophical Perspectives,* 28(1): 291-301.

Vrousalis, N., 2013. Smuggled into existence: Nonconsequentialism, procreation, and wrongful disability. *Ethical Theory and Moral Practice*, 16(3): 589-604.
Weinberg, R., 2016. *The risk of a lifetime: How, when, and why procreation may be permissible*. Oxford University Press.
Williams, Bernard, 1973, The idea of equality, in: B. Williams, *Problems of the Self*, Cambridge: Cambridge University Press, pp. 230-249.

DOI: 10.31009/LEAP.2019.V7.06

How Rule Consequentialism Avoids Boonin's Implausible Conclusion

TIM MULGAN
University of Auckland and University of St Andrews

ABSTRACT

David Boonin's Implausible Conclusion says a person (Wilma) who deliberately chooses to conceive a blind child when she could, at no significant cost to herself, have conceived a sighted child instead does nothing morally wrong. Boonin argues that we should accept his Implausible Conclusion, in part because every moral theory that avoids that conclusion has other implications that are even less palatable. I argue that Rule Consequentialism can avoid Boonin's Implausible Conclusion without leading to greater absurdities. My focus is Boonin's fifth premise in the non-identity argument, or the claim that "if an act does not wrong anyone, then the act is not wrong." I introduce an imaginary case where an act that wrongs no one seems to be very wrong. This cases involves a future refusal to avoid imminent human extinction by costlessly launching a Generation Starship programme whereby humanity would continue in space. I argue that this refusal is intuitively wrong, even though it wrongs no one. Any acceptable moral theory must accommodate this intuition. This requirement makes Rule Consequentialism more impersonal and demanding, which enables Rule Consequentialism to condemn Wilma's choice, solve the non-identity Problem, and avoid Boonin's Implausible Conclusion.

Keywords: David Boonin, Non-Identity Problem, Rule Consequentialism, Human Extinction, Derek Parfit.

1. INTRODUCTION

In his excellent, thorough, and thought-provoking book *The Non-Identity Problem and the Ethics of Future People*, David Boonin argues that we

should accept his Implausible Conclusion, in part because every moral theory that avoids that conclusion has other implications that are even less palatable. I argue that Rule Consequentialism can avoid Boonin's Implausible Conclusion without leading to greater absurdities. I cannot fully demonstrate such an ambitious claim in this short paper. My aim here is to sketch the resources available to Rule Consequentialism, highlight the challenges faced by its rivals, and motivate further exploration of Rule Consequentialist responses to Parfit's Non-Identity Problem (Parfit 1984: 351-379).

Boonin's Implausible Conclusion concerns the following case:

Wilma's Decision: Wilma deliberately chooses to conceive blind Pebbles when she could have, by taking a small pill once a day for two months, conceived sighted Rocks instead. Pebbles has a life worth living, but Rocks would have enjoyed a better life.

The Implausible Conclusion says that Wilma's act of conceiving Pebbles is not morally wrong. (Boonin 2014: 5 – all subsequent unattributed references are to this book). It is arrived at through the following line of reasoning:

P1: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been.

P2: If A's act harms B, then A's act makes B worse off than B would otherwise have been.

P3: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not harm anyone other than Pebbles.

P4: If an act does not harm anyone, then the act does not wrong anyone.

P5: If an act does not wrong anyone, then the act is not morally wrong.

C: Wilma's act of conceiving Pebbles is not morally wrong (27).

My focus is Boonin's fifth premise – the claim that if an act does not wrong anyone, then the act is not morally wrong. In his Chapter 6, Boonin critiques various proposals to 'identify a property that can make an act wrong even if the act does not wrong anyone' (149). These include 'solutions based on appeals to utilitarian and consequentialist principles, in both act- and rule- forms, and to Parfit's Principle Q, as well as solutions based on considerations of group rights and of virtue.' (149) While Boonin claims that each of these proposals ultimately fails, I will argue in this paper that

Rule Consequentialism has the resources to avoid the Implausible Conclusion and dissolve the non-identity problem.

2. A NIHILISTIC CONCLUSION

I focus on an imaginary case where an act that wrongs no one seems to be very wrong:

> *Generation Starship*: In some future century, scientists discover that deadly cosmic rays will hit the Earth in two hundred years, instantly and painlessly killing all living things and rendering the Earth uninhabitable. We cannot prevent this. However, we can avoid total human extinction by constructing interstellar 'generation starships' where a small population and their descendants will continue the human story in space.[1] Scientists also discover that ordinary humans could never survive in space. Only specially genetically engineered individuals can hope to survive the rigours of starship life. But these *interstellar people* could not survive on Earth. They can only flourish in space. Thanks to advances in technology since the twenty-first century, we know both that the interstellar people would enjoy very rich lives and that the cost to present people of launching the generation starship programme is comparatively low. Suppose we collectively decide not to launch the generation starship programme. (Perhaps Boonin's book has persuaded us that this would not be wrong.) When the cosmic rays strike, humanity goes extinct. Consciousness disappears from the cosmos, never to return.

This tale is very unrealistic. I explore the ethics of *plausible* generation starship cases more fully elsewhere (Mulgan forthcoming a, forthcoming b). Complicating factors include: (a) the construction of a successful generation starship programme using any currently imaginable technology would be very costly, very uncertain, and involve very high risks of very bad outcomes both for future interstellar people and for their descendants on some hostile planet; (b) generation starships would pose an existential threat to any indigenous life on other planets; (c) launching generation starships thus risks imposing on future colonists a morally unbearable choice been xenocide and survival; and (d) we could never be certain that cosmic rays were coming, that they couldn't be stopped, that our extinction wouldn't pave the way for either the evolution of superior *terrestrial* rational beings or the successful colonization of an empty Earth by

1 This scenario is familiar in speculative fiction – classic examples include Aldiss (1958) and Heinlein (1963). For scholarly discussion, see Caroti (2011).

superior *extra-terrestrials*, and so on.

In this paper, I set all these complications aside. I seek an intuitively clear case, not a feasible one. My tale's artificial features are designed to rule out two familiar moves in the non-identity literature. First, realistic extinction threats invariably involve clear harms to present people or to future people who would have existed whatever we do, such as people killed by meteor strikes, catastrophic climate change, pandemic, alien invasion, runaway AI or nanotechnology etc. By contrast, my tale involves *pure mere addition*: if we launch generation starships, then we produce a future whose inhabitants would never otherwise have existed at all. Second, I also stipulate that we *know* that if we create generation starships, then the future quality of life will be positive. This is a *very* controversial assumption, but I retain it here because (a) it is a common presupposition in current philosophical debates about extinction (e.g., Beckstead 2013, Kaczmarek 2017); and (b) if we question this assumption, then it is far from obvious that it would be wrong to embrace human extinction.

For this paper, I make two assumptions about my Generation Starship tale:

1. No one is wronged by this decision.
2. When we decide not to launch the generation starship programme and thereby bring the human story to an abrupt end, we behave very wrongly.

Both claims are controversial. But both are also plausible. A full defense of my first assumption – that no one is wronged – is beyond the scope of this paper. But such a defense could easily draw on Boonin's own arguments in earlier chapters. If we don't launch generation starships, then the interstellar people never exist. So they have not been wronged. And no one else is wronged either.[2]

My second assumption is that our behavior is wrong. Many people will insist that we behave very wrongly because it is *always* very wrong to fail to prevent imminent human extinction at comparatively little cost to oneself. Perhaps failing to avoid extinction is less wrong than deliberately causing

2 What about children born in the last generation before the cosmic rays hit? Their lives are cut tragically short. Don't we wrong them? We could retell the story to remove this anomaly. (If we stop reproducing 50 years before the cosmic rays hit and use robots to meet the needs of the last humans, then no life is tragically short.) Alternatively, we could agree with Boonin himself that these people are not wronged, because (for reasons familiar from Parfit's Depletion vs Conservation and Risky Policy tales – Parfit 1984: 361-4, 371-2) they wouldn't otherwise have existed at all (Boonin 2014: chapter five).

it. But it is still very bad. This reaction is unsurprisingly (more or less) universal among consequentialists. But it is also very common among non-consequentialists, including many theorists who do not think it is generally wrong to fail to produce good outcomes in non-identity cases. Consider three representative quotations:

> "Most of us believe that human extinction would be the worst of those possible tragedies that have more than a negligible probability of actually occurring" (McMahan 2013: 26).
>
> "The Earth may remain inhabitable for at least a billion years. … Our descendants might, I believe, make the further future very good. But that good future may also depend in part on us. If our selfish recklessness ends human history, we would be acting very wrongly. Such acts might be worse for no one; but, as I have argued, that fact could not justify these acts" (Parfit 2011: vol 2, 616).
>
> "Surely, if we developed a pill enabling each of us to live wonderful lives for 120 years it would be terrible for us to take the pill if the cost of doing so were the extinction of humanity" (Temkin 2012: 414[3]).

By analogy with Boonin's Implausible Conclusion, I will say that these philosophers all reject:

> The Nihilistic Conclusion: When we decide not to launch the generation starship programme and thereby bring the human story to an abrupt end, we do not do anything that is morally wrong.

The intuition that it is wrong to fail to avoid human extinction at comparatively little cost is, in some ways, an extreme example of the standard non-identity intuition that is Boonin's main target in his book. Boonin could therefore simply bite the bullet and insist that, counter-intuitive as it may seem, it is *not* wrong to fail to launch Starships and thereby bring about the premature end of the human story. If so, my new example would serve to highlight the counter-intuitiveness of his position.

I am not sure how Boonin himself will respond to this particular case. However, I am confident that many people who are fully persuaded by Boonin's defense of the Implausible Conclusion would still baulk at accepting my Nihilistic Conclusion. There does seem to be something special about extinction-avoidance. My question in this paper is whether Boonin could accommodate this asymmetric response. Can he distinguish my Starship case from his original Wilma case?

3 Temkin adds in a footnote that this claim 'is almost as obvious as any claim can be in the normative domain' (Temkin 2012: 414).

If my goal were simply to undermine Boonin's Premise Five, then a single case of wrongness where no one is wronged would suffice. I would then be simply begging the question against Boonin by assuming that my Starship provides such a case. But my goal here is more ambitious. I will argue that Boonin cannot treat extinction as a special case. If we reject my Nihilistic Conclusion, we are driven to reject his Implausible Conclusion as well.

Suppose we accept that our action in Starship *is* wrong. We must then explain *why* (and *when*) acts can be wrong even if they wrong no one. Boonin argues that those who reject his Implausible Conclusion must provide a moral theory (T) that avoids that conclusion *also meets the following three constraints*:

1. *Independence*: T must be justified independently of the fact that it avoids the Implausible Conclusion. (We can't introduce an ad hoc principle simply because it implies that Wilma's act is wrong.) (20)
2. *Robustness*: T must be strong enough to warrant rejecting any weakened versions of Boonin's premises that are still strong enough to generate the Implausible Conclusion. (22)
3. *Modesty requirement*: T must not be so robust that it generates implications even more implausible than the Implausible Conclusion. (22)

Analogously, if we are to reject the Nihilistic Conclusion, then we need a moral theory T* that avoids that conclusion while satisfying Boonin's three constraints *with respect to the Nihilistic Conclusion*. For instance, just as Boonin rejects ad hoc solutions tailored to avoid the Implausible Conclusion, we must also reject ad hoc solutions to the Nihilistic Conclusion. I will argue that the best candidate for T* is a version of Rule Consequentialism (hereafter RC) *that also avoids the Implausible Conclusion*. Therefore, rejecting the Nihilistic Conclusion leads us to reject the Implausible Conclusion.

3. THERE IS NO ALTERATIVE

The most obvious challenge for my argument is robustness. Generation Starship and Wilma's Decision are very different kinds of cases. Surely many possible moral principles will judge the former to be morally (very) wrong without finding any fault with the latter? After all, Wilma herself might heartily agree that *it is always wrong to fail to avoid human extinction*

at comparatively little cost.

My ambitious reply is that any theory that can avoid the Nihilistic Conclusion while meeting Boonin's independence and modesty constraints will also avoid the Implausible Conclusion. It is easy to construct rival theories that *seem* to avoid the Nihilistic Conclusion without having *any* implications for Wilma's case. But such rivals either (a) lack an independent rationale or (b) have other even more implausible implications.

The challenge for alternatives to RC is to say *why* our behavior is wrong *without merely redescribing it*. Merely asserting that 'it is wrong to fail to avoid imminent human extinction at little cost' doesn't *explain* anything.

An adequate refutation of all alternatives to RC is far beyond the scope of this paper. In this section, my aim is only to illustrate the challenges facing such alternatives, by using two representative rivals. The philosophical literature on human extinction offers three main explanations for the badness of human extinction:

1. *Utilitarian*: Human extinction would be bad because it does not maximize future human well-being (e.g., Beckstead 2013, Kaczmarek 2017).
2. *Individual Meaning*: Human extinction would be bad because, if there were no future people, our present projects would be drained of meaning (e.g., Scheffler 2013, 2018).
3. *Sui Generis*: Human extinction has a special sui generis badness over-and-above any impact on either the lives of individuals or total future well-being (e.g., Frick 2017, Temkin 2015).

RC exemplifies the first type of explanation. One prominent example of the second type is Samuel Scheffler's argument that, if you knew that the world would end thirty days after your death, then this would drain your life projects of much of their current meaning (Scheffler 2013, 2018). Scheffler's approach may seem ideal for Boonin's purposes. Extinction is bad because, despite initial appearances to the contrary, it *is* bad for individuals. We therefore have no counter-example to Boonin's argument. Unfortunately, Scheffler-style meaning is not sufficient to help Boonin. Even if imminent extinction has some negative impact on present individuals, that impact does not seem sufficient to explain all the badness of extinction. Most importantly, while considerations of meaning might explain why we *wouldn't* allow the human species to become extinct, it does nothing to explain why this would be *wrong*. For instance, in an earlier contribution to the debate, Brooke Trisel argues that *some* valuable projects are

sufficiently self-contained to be immune to extinction-driven meaningloss (Trisel 2004). Suppose we agree with Trisel. Therefore, instead of launching generation starships, we decide to simply exchange our future-oriented projects for self-contained ones (perhaps we inaugurate a new religion of purely self-contained contemplation of the cosmos). We might be making a prudential error – especially if Trisel turns out to be mistaken. But prudential errors of this sort are not normally regarded as moral errors.

My representative example of a sui generis explanation of the badness of extinction is suggested by Temkin (2015):

> *Temporal Consequentialism* (TC): An outcome's value depends, not only on the quality of individual lives, but also on how those lives are distributed through time. It is not always better if there are more happy people. But it *is* always better if there are *more time*s when happy humans live. Adding extra happy people at already inhabited times does not improve things, but adding them at otherwise empty times does. And failing to add such lives is prima facie wrong.

Unlike most non-consequentialist theories, TC rejects Boonin's P5. Failing to costlessly prevent imminent human extinction is wrong even if no one is wronged. Unlike standard utilitarian consequentialism, however, TC denies that failing to maximize aggregate happiness *at a particular time* is wrong. TC thus delivers the two verdicts we are seeking: it condemns our decision not to launch generation starships, but it says *nothing* against Wilma.

TC seems to fit the bill. Unfortunately, it has two failings. TC lacks an *independent* case and it is *immodest*. These failings are linked, because the principal argument *for* TC is presumably that it captures our considered moral judgements (Temkin 2012, 2015). Consider the following choice:

> *Leave or Remain*: Humanity faces an existential threat. We have only two options. We can either (a) maximize the quality of life on Earth for a short time, or (b) pour all our resources into a very Spartan short-lived generation starship programme. If we *Remain*, then 100 million people will enjoy good lives for one century, and then humanity becomes extinct. If we *Leave*, then during each of a thousand centuries, 100,000 people will enjoy less good lives, and then humanity becomes extinct. Temporally neutral consequentialists prefer Remain to Leave. The numbers are identical, and everyone is happier. Suppose also that this is a *Same People Choice* where both outcomes contain *the same individuals* (Parfit 1984: 356). (Perhaps Leave involves cryogenic

> storage in outer space, with 100,000 people thawed each century.) If we Leave, then every individual is worse-off than she would otherwise have been. Temporally-neutral *non-consequentialists* will thus also prefer Remain. By contrast, TC insists that Leave has one very significant advantage: *Humanity lives for an extra 999 centuries*! If the difference in individual quality of life between Remain and Leave is sufficiently small, then Leave must be *superior all-things-considered*.

TC's verdict is hard to believe. Suppose we choose Leave. We make 100 million people worse-off than they would otherwise have been. Could we then justify ourselves *to those people* simply by noting that their sacrifice *ensures that humanity survives longer*? I doubt they would be very impressed![4]

Of course, I have not refuted all alternatives to RC. But I have hopefully done enough to show that constructing an independent moderate robust theory T* is not straightforward. We cannot assume that the Implausible and Nihilistic Conclusions can be so easily separated.

4. RULE CONSEQUENTIALISM

For any consequentialist, our decision in Generation Starship is very

4 To further highlight TC's implausibility, consider a first-person analogue:

Duplication: You are about to be duplicated ten times. Each copy will live on a different planet. There will be no interaction between them. Does it matter *to you* whether those copies exist simultaneously or consecutively? Presumably not.

You might reply that you would rather not be duplicated at all. Your original self could then live on ten successive planets. This is because what matters to you is not how long you exist, but how your life's *narrative* goes. But, as the following tale reveals, this consideration counts against TC.

Time Travel: Each year, you have two options. You can step into either (a) a device that *only* transports you to another planet, or (b) a device that both transports you to another planet *and sends you one year into the past*. Either way, you will live in different places. If you *merely teleport*, then you also live through many different time periods. If you instead time travel, then you always live through the same time period. Both devices are perfectly reliable. But *mere teleportation* carries a significant cost. Life is getting progressively worse everywhere. If you merely teleport, your quality of life continually declines. If you time-travel, it remains constantly high.

I suggest that there is *no* self-interested reason to opt for mere teleportation. If time-travel offers a richer life, then it is irrational to refuse it. What matters *to you* is the internal narrative duration of your life, not its occupation of external time periods. Similarly, temporally-neutral consequentialists will insist that what matters morally *in Leave or Remain* is the internal perspective of the 100 million inhabitants, not how they happen to be spread through time.

wrong. The loss of all that future happiness is a moral catastrophe. Any plausible form of RC must avoid the Nihilistic Conclusion. The two crucial questions are:

1. How does RC deliver this result?
2. What (if anything) does this tell us about RC's response to Wilma's choice?

I will argue that RC's best response to extinction threats enables it to avoid the Implausible Conclusion. In this section, I present a very abbreviated independent case for RC. The main justification for this brevity is that Boonin himself does not question the independence of RC (175-8).

RC offers a moderate, liberal alternative to act utilitarianism. RC pictures morality as a collective enterprise, and evaluates moral codes by their collective impact on human well-being. RC's fundamental moral questions are: 'What if we did that?', and 'How should we live?' (Hooker 2000, Mulgan 2001: 53-103, 2006: 130-160, 2009, 2015, 2017.)

RC ranks competing moral codes by asking what would happen if everyone in the next generation *internalized* any given code. (This sets aside the cost of *changing* existing moral beliefs, but factors-in the difficulty of trying to internalize a very demanding ethic.) We then evaluate *acts* by asking what someone who had internalized the ideal code (call her 'the RC ideal agent') would feel free to do in that situation.

We cannot hope to specify the ideal code in any detail. At most, we can identify some of its general features. But often that is enough. We might, for instance, be confident that the ideal code does not support gratuitous torture, even if we don't know exactly what it says about other issues.

The independent case for RC has the following elements:

1. RC captures a perennially attractive picture of morality as a collective human enterprise passed on from one generation to the next.
2. RC follows the spirit of the classical utilitarians, especially Jeremy Bentham and J. S. Mill, who also began with simple moral principles and allowed complexity to emerge empirically in response to our evolving knowledge of human nature and the human situation.
3. Hooker argues that 'the best argument for rule consequentialism is that it does a better job than its rivals of matching and tying together our moral convictions' (Hooker 2000: 101). The ideal code must be internalized by a new generation *of human beings*. Human

nature is not infinitely plastic. Any plausible ideal code will include familiar moral dispositions such as honesty, generosity, promise-keeping, courage, murder-aversion, and so on. The RC ideal agent will not walk callously past children drowning in ponds, take pleasure in the sufferings of others, or reject the basic goods of human life.

4. However, RC does not merely recapitulate common-sense morality. It also improves it. Our actual moral practice is not optimal from an RC perspective, especially when it comes to its impact on the well-being of distant future people. RC will depart from commonsense morality by including a stronger reason to promote the good, a stronger prima facie obligation to confer benefits on others (and avoid harming them) when one can do so at comparatively little cost to oneself, a greater willingness to extend these obligations to cover future people even in Different People Choices, a greater general commitment to temporal impartiality, and so on (Mulgan 2006, 2015, 2017).

I conclude that invoking RC to avoid the Nihilistic and Implausible Conclusions is not ad hoc. And it is easy to see how RC must avoid the former. The moral code that maximises future well-being won't allow us to gratuitously fail to avoid imminent human extinction. The real question is whether or not RC also avoids the Implausible Conclusion.

5. HOW RULE CONSEQUENTIALISM AVOIDS THE IMPLAUSIBLE CONCLUSION

What does RC say about Wilma? Two RC commitments push in opposite directions here. On the one hand, RC is *consequentialist*. We choose an ideal code that best promotes future human well-being. The RC ideal agent surely accepts a more demanding and impersonal morality than we do. These features all suggest she will *not* condone Wilma's choice.

On the other hand, RC presents itself as a moderate liberal *alternative* to act consequentialism (Hooker 2000). An ideal code that is *too* demanding or impersonal could not be taught to human beings. Proponents of RC argue (i) that RC recognizes a broad sphere of personal moral freedom where agents are free to not maximally promote the good, (ii) that *reproductive* freedom is a central moral commitment that RC must respect, and (iii) that, despite its general commitment to impersonal demandingness,

the ideal code may also include person-affecting principles, priority to actual people over future people, and a rejection of procreative coercion. The RC ideal agent will not feel obliged to create the happiest children she possibly could. (Mulgan 2006: 172-3) This suggests she will not regard Wilma's choice as *wrong*. If sub-optimal procreation is permitted, then why shouldn't Wilma create Pebbles rather than Rocks?

As Boonin notes, earlier RC discussions of this very issue (including my own) leave open RC's final verdict on cases like Wilma's (178, citing Mulgan 2009: 131). However, I will now argue that, once it is updated to accommodate the full range of possible futures – *including futures like my Generation Starship tale where extinction is at stake* – RC will reject Wilma's choice. (The rest of this section abbreviates much longer discussions in Mulgan 2011, 2015, 2017.)

Wilma might reply that factoring these new futures into RC cannot affect *her* choice, for two obvious reasons. First, human extinction is not at stake in her decision. Second, even if human extinction were a relevant consideration, creating Pebbles seems no better or worse *from the perspective of extinction-avoidance* than creating Rocks. What could Wilma's decision to have a less-than-optimally-happy child possibly have to do with human extinction?

The answer is that, to *reliably* avoid imminent human extinction across the full range of possible futures, RC must abandon its commitment to a liberal ideal code that incorporates a permissive procreative ethic and a presumption of reproductive freedom. While such a code probably will be optimal in more favourable futures, it could lead to premature extinction if things go badly wrong. For instance, there are many possible futures where human life is very harsh and the survival of humanity demands (i) that very scarce resources must be carefully preserved for subsequent generations, (ii) that population size must be strictly monitored, and (iii) that people continue to reproduce even though they know both that their children's lives will be very grim and that bearing and raising children will reduce the quality of their own lives. One example would be a less rosy (but more realistic) generation starship scenario, where people endure very restricted and unpleasant conditions on starships in the hope that later generations will enjoy a much better life on a new planet. Future people committed to current ideas of reproductive freedom might reject such brutal starship lives out of hand, even if they were humanity's last best hope. This is not an acceptable result for RC. In futures where the survival of humanity depends on people having children even though they don't want to, the ideal moral agent must feel *obliged to reproduce*.

Of course, Wilma *does* reproduce. She chooses to create Pebbles rather than *either* creating Rocks *or having no child at all*. Wilma might insist that a moral code that *obliges* people to have children who are much worse-off than Pebbles can hardly object to her decision. My reply is that, once it factors in the possibility of restricted futures and extinction risks, RC must make broader changes to its ideal code. RC will now select a much less permissive ideal code that rejects any presumption of reproductive freedom and also includes a much stronger obligation to do whatever one can to enhance future well-being. When making a procreative choice, the RC ideal agent will not feel free to consult her own desires when this would lead to a sub-optimal result in terms of future well-being. She would not feel free to indulge her own preferences in Wilma's situation, and therefore she will regard Wilma's sub-optimal creation of Pebbles as wrong.

We can easily *imagine* more flexible moral codes with fine-grained rules that would specify the optimal pattern of behavior in any particular future situation.[5] These flexible codes would be designed to limit procreative freedom only in bleak futures, thereby avoiding extinction without compromising Wilma's decision. The crucial question is whether such logically possible codes would produce optimal results when internalized in a whole future generation. There are three general reasons to believe that the optimal RC code will not take this form. First, every increase in complexity makes it harder for a whole generation to successfully internalize a given code, thereby reducing its effectiveness. Second, extinction threats will take a great variety of forms, and a code whose stricter rules only come into play in 'extinction scenarios' might misfire very badly if agents were caught unawares. Third, the additional benefit of also safeguarding Wilma's freedom would not be sufficient to justify these costs – especially if, as I argue in the next section, an RC ideal code that condemns Wilma's choice can consistently sanction limited procreative freedom in other cases.

This is a very quick sketch of how RC might avoid the Implausible Conclusion. The argument rests on speculative empirical claims which are hard to verify. This is a general feature of arguments within RC, which RC's opponents often find unsatisfactory. My goal here is not to vindicate RC, but merely to demonstrate that, if RC can establish its general credentials as a plausible moderate alternative to Act Consequentialism that makes sense across a wide range of possible futures, then it can also reasonably hope to avoid the Implausible Conclusion.

5 I am grateful to an anonymous reviewer for pressing me on this point.

6. IS RULE CONSEQUENTIALISM IMMODEST?

RC responds to restricted futures and extinction risks with a more demanding and impersonal ideal code. This ideal code provides RC's standard of right and wrong. RC concludes that Wilma's choice is wrong, thus avoiding Boonin's Implausible Conclusion. But is the price too high? Does RC's new demanding ideal code have other implications that are even worse than the Implausible Conclusion? [6]

Boonin contrasts Wilma's choice with two others:

1. *Wilma*'s Choice*: Wilma* has a medical condition which means she can only have blind children. Wilma* decides to have a child, who enjoys a very good life despite being born blind. (112-3, 121-2. 'Wilma*' is my label. Boonin himself presents this as an alternative version of Wilma's Choice)
2. *Jane's Choice*: Jane already has two children. She knows that her third child would be as happy as Rocks and that her other children would be no worse-off. Yet Jane decides not to have a third child. (171-2)

Boonin argues plausibly (i) that we intuitively think that, while Wilma's choice is wrong, neither Wilma* nor Jane does *anything* wrong, and (ii) that these intuitions are at least as compelling as our rejection of the Implausible Conclusion. Our moral theory T must avoid the Implausible Conclusion *without implying that either Wilma* or Jane does anything wrong.*

Boonin then confronts RC with a dilemma: 'The problem … is that if autonomy is not valuable enough to grant Wilma the freedom to conceive Pebbles rather than Rocks, then it is not valuable enough to grant Jane the freedom to refrain from conceiving a third child in the case where she has a slight preference not to have another child.' (176)

RC values autonomy. Either that value is sufficient to permit Wilma's creation of Pebbles, or it is not. If it is, then RC cannot reject Boonin's Fifth Premise. If not, then RC cannot permit Jane *not* to have a third child. Boonin regards the latter as more absurd than the Implausible Conclusion.

6 RC is very controversial for reasons unrelated to future people, non-identity, or human extinction. The literature provides many general objections to RC. Boonin could borrow these objections and argue that RC is already too immodest. A full consideration of this move would take us too far afield. If RC fails on other grounds, then its ability to dissolve the non-identity problem is moot. However, in his book Boonin focuses on whether RC's response to non-identity is itself immodest, and I follow his lead here. (I am grateful to an anonymous referee for pressing me on this point.)

RC has two options: differentiate or bite the bullet. I shall now briefly argue that both are promising.

RC can *differentiate* between Wilma*, Jane, and Wilma. Two features of RC are salient here:

1. The ideal code must be *internalized*, not merely followed. RC does not simply ask what those its idealized agents will do. It asks instead about their reasons, motives, or character, and about why they act as they do.
2. Boonin asks whether Wilma's action is *wrong*, not whether it should be socially sanctioned or legally prohibited. For RC, this is a *first-personal* question: Would the RC ideal agent *feel free* to act as Wilma does? RC ideal agents might not feel free to do what Wilma does, even if they agree that her reproductive freedom should not be subject to any socially or legally enforced constraints.

Intuitively, Wilma's choice is wrong, not only because it is suboptimal, but also (indeed, primarily) because there is no good reason for it. RC captures this natural thought. Wilma violates a prima facie obligation to confer benefits and promote future well-being when she can do so at little cost to herself. The RC ideal agent would only feel free to behave like this if she had a strong reason to do so. But what *is* Wilma's reason? And would the RC ideal agent regard it as a *good* reason?

Boonin discusses a similar objection: namely, that Wilma's act is wrong *because it reflects a bad character* (184-8). Boonin replies that this begs the question against Wilma (187-8). Wilma's character is deficient *only if her act is wrong*. But what if Wilma is a philosopher whose reason for thinking her behavior is permissible is that it wrongs no one?

Boonin's target here is not RC. And RC's objection to Wilma is not circular in this way. RC defines good and bad motives or character in terms of the ideal code, which in turn is chosen because it best promotes future human well-being. The ideal code must protect the interests of *distant* future people in non-identity cases. So it must ensure that *earlier* future people *don't* generally regard actions as permissible simply because they wrong no one. The RC ideal agent would not feel free to do something just because she knew it would wrong no one. Wilma needs *another* reason. But what else can Wilma say for herself?

By contrast, the RC ideal agent will treat *Wilma** very differently. Recall that our ideal code must work effectively in broken futures where *our present quality of life* is no longer available. I argue elsewhere that this is best achieved if the threshold for acceptable procreation *is relativized to*

the agent's situation (Mulgan 2006, 2011, 2015). People living in a broken future will feel free to have children even though they cannot provide anything approaching Rocks's quality of life. Wilma*'s situation is broken at the *individual* level. She too can only offer her child a diminished life. The RC ideal agent might well feel free to procreate in Wilma*'s situation. Therefore, RC judges that Wilma*'s does nothing wrong. Wilma cannot offer a similar justification, because *her situation is not broken*. She *could* offer a future child a better life.

This reply differentiates Wilma from Wilma*. But *Jane's* situation is *not* broken. Like Wilma, Jane fails to create the happiest child she could. She also fails to maximize future well-being. How could the RC ideal agent approve Jane's choice and reject Wilma's?

One answer is that a verdict of wrongness is much more threatening to Jane's reproductive autonomy than to Wilma's. If we judge that *Jane* acts wrongly, then we must conclude that she is *obliged to have a child she doesn't want*. If we judge that Wilma acts wrongly, then we must conclude that she is *obliged to have a different child at comparatively little cost*. These two obligations are not equally burdensome. 'Because I don't want another child' is a deeper reason than 'Because I don't want *this* (sighted) child'. RC can explain this intuitive distinction. Only in extreme circumstances will the RC ideal agent feel obliged to have a child she doesn't want. As Jane's circumstances are not extreme, she does no wrong.

I conclude that RC can differentiate the three situations: Wilma, Wilma*, and Jane. Alternatively, RC could *bite the bullet*, and accept that Jane is not morally free to choose. If RC takes this route, it will say, not that Jane is obliged to have a third child, but rather that she is *obliged not to*! For any consequentialist, the striking feature about Jane's choice is that the resources Jane would (inevitably) devote to her extra child's welfare could almost certainly produce much greater benefits elsewhere. As she already has two children, Jane can hardly argue that she must procreate because she has a burning desire to experience life as a parent (Cf. Weinberg 2015). Unless human extinction is a *present* threat, the RC ideal agent might not feel free to have another child in Jane's case. (Of course, this might also mean that Wilma's *real* fault is that, instead of having Pebbles, she should have no child at all.)

Are these implications absurd? Is RC guilty of immodesty here? I am not sure. It seems much less absurd to claim that Jane is obliged not to have a third child *because there is so much good to be done in the world* than to claim that Jane is obliged to have a third child *because that child would have a good life*. The latter claim strikes many non-consequentialists as intrinsically absurd, while the former is (merely) extremely demanding.

Many people, of course, do find such extreme demands absurd. But RC can reply that, once we take on board the very real prospect of broken and empty futures, we must acknowledge a more demanding morality across the board. Extreme demandingness is no longer conclusive proof of immodesty.

I conclude, contra Boonin, that once we recognize that it must avoid the Nihilistic Conclusion, RC does have the resources to avoid the Implausible Conclusion and dissolve the non-identity problem.[7]

BIBLIOGRAPHY

Aldiss, B., 1958: *Non-Stop*, London: Faber and Faber.
Beckstead, N., 2013: *On the overwhelming importance of shaping the far future*, Rutgers PhD thesis.
Boonin, D., 2014: *The Non-Identity Problem and the Ethics of Future People*, Oxford: Oxford University Press.
Caroti, S., 2011: *The Generation Starship in Science Fiction: A critical history 1934-2001*, Jefferson, North Carolina: McFarland and Co.
Frick, J., 2017: "On the survival of humanity", *Canadian Journal of Philosophy* 47: 344-367.
Heinlein, R., 1963: *Orphans of the Sky*, London: Gollancz.
Hooker, B., 2000: *Ideal Code, Real World*, Oxford: Oxford University Press.
Kaczmarek, P., 2017: "How much is rule-consequentialism really willing to give up to save the future of humanity?", *Utilitas* 29: 239-249.
McMahan, J., 2013: "Causing People to Exist and Saving People's Lives", *Journal of Ethics* 17: 5-35.
Mulgan, T., 2001: *The Demands of Consequentialism*, Oxford: Oxford University Press.
—2006: *Future People*, Oxford: Oxford University Press.
—2009: "Rule Consequentialism and Non-Identity", in *Harming Future Persons*, ed. M. Roberts and D. Wasserman, 115-134, Dordrecht: Springer.
—2011: *Ethics for a Broken World*, Durham: Acumen.
—2015: "Utilitarianism for a Broken World", *Utilitas* 27: 92-114.
—2017: "How should utilitarians think about the future?", *Canadian Journal of Philosophy* 47: 290-312.
Mulgan, T, forthcoming a: "What exactly is wrong with human extinction?", in *Proceedings of the International Society for Utilitarian Studies*, ed. C. Schmidt-Petri and M. Schefczyk, Karlsruhe: KIT Scientific Publishing.
Mulgan, T, forthcoming b: Future Worlds: *The moral significance of Generation Starships*, draft manuscript.
Parfit, D., 1984: *Reasons and Persons*, Oxford: Oxford University Press.
—2011: *On What Matters*, Oxford: Oxford University Press.
Scheffler, S., 2013: *Death and the Afterlife*, Oxford: Oxford University Press.
—2018: *Why Worry about Future Generations?* Oxford: Oxford University Press.

7 I am very grateful to two anonymous referees for helpful and insightful comments on an earlier draft.

Temkin, L. S., 2012: *Rethinking the Good*, Oxford: Oxford University Press.
—2015: "Rationality with respect to people, places, and times", *Canadian Journal of Philosophy* 45: 576-608.
Trisel, B. A., 2004: "Human extinction and the value of our efforts", *The Philosophical Forum* 35: 371-391.
Weinberg, R., 2015: *The Risk of a Lifetime: How, When, and Why Procreation May Be Permissible*, Oxford: Oxford University Press.

DOI: 10.31009/LEAP.2019.V7.07

Why Wear Blinders? Boonin and the Narrow Approach to the Non-identity Problem*

MELINDA A. ROBERTS
The College of New Jersey

ABSTRACT

Boonin endorses reasoning that leads to what he calls the Implausible Conclusion regarding when future-directed choices that at first glance seem to impose substantial burdens on future people are permissible. According to the Implausible Conclusion, such choices – provided they serve as "but for" causes of the future person's coming into existence, and provided they don't harm still other people – turn out to be permissible. That's so, according to Boonin, since those same choices don't in fact impose the substantial burdens on future people that we at first glance think they do and on the assumption (an assumption not questioned here) that any plausible moral theory will be person-affecting in nature and thus limit wrongdoing to choices that harm existing or future people. The purpose of this paper is to argue that the reasoning Boonin urges us to accept fails to establish the Implausible Conclusion, on two closely related grounds. First, Boonin ignores some of the morally significant factual details of many of the cases he wants his Implausible Conclusion to hold for, including some of their critically important *modal* details. And, second, at least one principle that Boonin endorses – a principle that governs when a future person is harmed by a prior choice – is clearly false, even on the assumption (also not questioned here) that harming a person is a matter of making that person worse off. The combination of those two mistakes results in an inappropriately narrow approach to the non-identity problem. Jettisoning that approach means that we are free, thankfully, to say what we thought we should say about at least many versions of the non-identity problem to begin with: that the Implausible Conclusion is false, and that future-directed choices that seem at first glance to impose substantial burdens on future people are indeed often wrong.

* I am very grateful to the anonymous referees for their invaluable comments on a prior version of this paper.

Keywords: Risky policy case, counterfactual account of harm, probability and harm, medical treatment cases and harm, contrast between two- and three-option cases.

1. WHAT IS THE NON-IDENTITY PROBLEM?

1.1 *The Implausible Conclusion*. David Boonin's presentation of the non-identity problem is sufficiently clean and clear that any interested reader – philosopher or not – will easily grasp both the reasoning Boonin identifies as giving rise to the problem and how difficult it is to say just where that reasoning goes awry. Putting things that way, of course, may seem to suggest that reasoning does go awry. This is certainly the conventional view. But it's exactly this conventional view that Boonin contests. The main argument of his book is that that reasoning *doesn't* go awry. [1] According to Boonin, to solve the problem is to accept the reasoning. Specifically, it is to accept that the non-identity problem isn't a *problem* after all but rather a sound argument to a wide range of surprising results. Boonin collects those results under the heading the *Implausible Conclusion*, and then ices his cake with some reasons to think that those results are not really "so implausible after all." [2]

Thus we thought that it was a problem for a moral theory that it evaluates as morally permissible all sorts of choices that impose devastating burdens on future people. But that was a mistake. The better view, according to Boonin, is that choices that burden the future people they bring into existence – *existence-inducing choices* – are perfectly permissible provided that those future people's lives are on balance worth living and that still other people – whether existing or future – are not themselves also burdened by those choices. [3]

That conclusion is sweepingly exculpatory. That's so in virtue of the fact that bringing a new person into existence – causing a new person to exist – is not as hard to do as we might have thought. It's not necessarily a

1 Boonin (2014), chaps. 1 and 7.

2 Boonin (2014): 191.

3 "Existence-inducing choices," I should note, is my term, not Boonin's. I should note that another point taken for granted in discussions of the non-identity problem is that the proper evaluation of a given choice must take into account not just the burdens that come with that choice but also the benefits. Thus, to properly evaluate whether the choice to vaccinate a child is morally permissible, we must balance the burdens against the benefits. The pain the child experiences is a burden; that the child is then protected against, for example, measles is a benefit. By hypothesis in the non-identity cases, the benefits always do at least counterbalance the burdens and the life is, despite the burdens, on balance always clearly worth living.

matter of getting pregnant or impregnating someone else. Rather, as Derek Parfit and others have argued, choices that impose burdens on future people and seem to have nothing to do with reproduction, on closer inspection, often – indeed, *typically* – turn out to existence-inducing.[4]

How can that be? How, for example, can a choice that we make today stripping distant future generations of important natural resources also serve to bring just those same people into existence?[5] Here, we take for granted that any distinction in the gametes involved in the conception of any one person inevitably would lead to a distinction in the identity of the person conceived. Then, as Parfit has argued, even the slightest departure from a given sequence of acts and events – any little change that might affect the "timing and manner" of conception, even changes aimed solely at avoiding or mitigating a given burden on behalf of a given future person – will lead, or at least, as Gregory Kavka observed, *very probably* will lead, to that person's never coming into existence at all.[6] How many of us would have existed, as Parfit puts it, had "railways and motor cars" never been invented?[7] Had World War II not been fought? Had our parents – indeed our forebears; indeed the prehistorical sea-creatures from which we have ultimately descended – not had sex just when and how and with whom they in fact did?

Now, an *impersonal* approach to the evaluation of choice may deem the fact that a given choice brings a particular person into existence beside the point. That a choice that has burdened a person has also conferred on that same person a counterbalancing benefit – the gift of life itself – is not morally relevant. If an *alternate* choice would have avoided the burden by way of bringing a *distinct* person into existence – a person *non-identical* to the person burdened by the choice in fact made – then other things equal that is reason enough, under an impersonal approach, to deem the choice in fact made morally wrong.

But suppose that we find a person-*affecting*, or *person-based*, approach to the evaluation of choice more intuitive than an impersonal approach. Specifically, suppose we accept what is called the *person-affecting*, or *person-based, intuition*, according to which we do nothing wrong, other things equal, in failing to bring a person into existence to begin with. Suppose we accept that what is "bad" must be "bad for" someone – that is, that a choice can be morally wrong *only if* that choice makes things worse for – or *harms* – at least some existing or future person.

4 Parfit (1987), Part IV; Kavka (1982).
5 This is Parfit's Depletion case. See Parfit (1987): 361-363.
6 Kavka (1982): 100 n. 15.
7 Parfit (1987): 361.

It's then that we face the non-identity problem. It's then that we seem to find ourselves compelled to say that the existence-inducing choice under scrutiny – however burdensome, and provided just that the life is worth living and no one else is affected – is perfectly permissible.

Thus does Boonin endorse a line of reasoning that ends in a sweepingly exculpatory position in respect of the choices that we make today that will predictably burden – and perhaps severely burden – the future people they bring into existence. According to Boonin, we are, in other words, compelled to say that when it comes to future people, almost *anything goes.* That's the *Implausible Conclusion.* And that's the line of reasoning and the conclusion that Boonin thinks we must accept in order to solve the non-identity problem.[8]

1.2 *The Wilma and Pebbles case.* Boonin explores both *direct* versions of the non-identity problem, including the Wilma and Pebbles case, and *indirect* versions of the problem, including the Risky Policy case.[9] A *direct* version of the problem is one in which a "choice directly determines which particular person will exist after the choice is made."[10] In the indirect version of the problem, a "choice has consequences that initiate a complex chain of events that eventually have [a] decisive effect on which particular people exist . . . ."[11]

However, as Boonin understands things the underlying logic of the problem is the same in the Wilma and Pebbles case and in the Risky Policy case. Now, my own view is that we should closely question whether that's in fact so. It's not that there's a clear moral distinction to be made between Boonin's *direct* cases and his *indirect* cases but rather that not all non-identity cases are themselves of a piece. If that's so then – contrary to Boonin's view – the correct account of the Wilma and Pebbles case and the

8 As noted earlier, the icing on his case are the reasons he gives, toward the end of his book, that suggest that the Implausible Conclusion is not really so implausible after all. I won't try to evaluate those reasons here, other than just to note that he proceeds in part by analogy. According to Boonin, an agent is "within his rights" and does not "do anything positively immoral" when he or she fails to rescue a child from certain death by starvation in a case where, at little cost to the agent and at no cost to anyone else, the agent easily could have rescued the child. If we let the agent off the hook in this case – the famine relief case; Boonin (2014): 192-195 – why is it so hard to believe we should let the agent off the hook in the non-identity cases as well? The problem with Boonin's argument is that it's not at all clear that we do let the agent off the hook in the failure to rescue cases. The transitivity argument he presents later in the same chapter is no more convincing. Boonin (2014): 198-205.

9 Boonin (2014): 2-7. As Boonin acknowledges, his Wilma and Pebbles case parallels Parfit's two medical programs case. See Parfit (1987): 367-368. Boonin's Risky Policy case is a version of Parfit's risky policy case. Parfit (1987): 371-372.

10 Boonin (2014): 5.

11 Boonin (2014): 5.

correct account of Risky Policy may be quite distinct. I will return to that point in what follows.[12] For introductory purposes, however, to understand the non-identity problem *as Boonin himself understands the problem*, it's enough to consider just the Wilma and Pebbles case.

There, the child Pebbles is born "incurably blind" as a result of a condition that afflicts her mother, Wilma.[13] Pebbles' life is clearly worth living. Wilma could have taken pills for two months that would have cured the condition. She could have then conceived a child – Boonin calls him Rocks – who would have been fully sighted. The catch is that Pebbles and Rocks are distinct (non-identical) individuals. In no clear sense would it have benefited *Pebbles* for Wilma to have delayed conception and produced *Rocks* instead. By hypothesis, no one else will be affected by how Wilma makes her choice. If – citing the person-based intuition – we think that wrongdoing is a matter of burdening, or making things worse for, some existing or future person, we seem to find ourselves compelled to say that what Wilma has done, in not taking the pills and producing a blind child when she could have taken the pills and (later) produced a sighted child instead, is perfectly permissible.[14]

Is this a non-identity *problem*, a line of reasoning that generates a result we must somehow get around? Or is it instead a sound non-identity *argument*, an argument that shows that Wilma's choice is perfectly permissible?

Boonin takes the latter view. And the evaluation of permissibility – to my ears, in any case – in his Wilma and Pebbles case does not seem implausible. At least, to my ears, the example does not on its own establish that the person-based intuition is false. When a given choice *maximizes* wellbeing for all involved – for all, that is, the existing and future people who exist under that choice – is it really that clear that it's *morally wrong* to bring the imperfect child over the more perfect child into existence? I'm not sure that it is – and, if it's not, then we don't have a clear-cut counterexample establishing the person-based intuition itself as false.

But neither, of course, does any one case prove that any given principle is true. The difficulty for Boonin is that still other cases suggest that his approach to the non-identity problem – his focus on certain factual details of the various non-identity cases and disregard of others, in combination with his formulation of the person-based intuition itself – really does at some point go awry.

12 See parts 3 and 4 below.
13 Boonin (2014): 2-3.
14 Boonin (2014): 5.

Thus consider the Warren and Percy case. Warren wants to have a child, and he wants to use the technique of in vitro fertilization to achieve that end. With technological support provided by others, Warren has arranged things so that his own single sperm and a single donor egg are in close proximity to each other in a glass dish and just moments away from fertilization. During that interval, Warren introduces an iatrogenic chemical into the dish, a chemical that Warren understands will inevitably cause the resulting child – call him Percy – to suffer a certain impairment. Why? Perhaps Warren is eager to test "accepted wisdom" which has it that the chemical is harmful; perhaps Warren has agreed earlier on to produce a child for a couple who seeks to raise a child with exactly the impairment the chemical will cause; perhaps it's just that Warren has been paid $100 to introduce the chemical into the dish. Whatever the details, it's a stipulation of the case that, had Warren not introduced the chemical into the dish, Warren would have simply discarded the contents of the dish and Percy would never have existed at all. Fertilization proceeds, the embryo is then transferred to the uterus of a commercial surrogate, the child Percy eventually develops out of that embryo and is born. Percy's existence is predictably flawed but nonetheless clearly worth having. And it's an existence that Percy never would have had to begin with had Warren not chosen just as he did.

It's a difficulty for Boonin that his approach to the non-identity problem – that is, the very approach that generated the result that *Wilma's* choice is permissible – now seems to generate the result that *Warren's* choice is permissible. In contrast to the result that Wilma's choice is permissible, the idea that Warren's choice is permissible seems *irredeemably* implausible. It seems then that after all Boonin's approach to the non-identity problem – at either the factual level or the theoretical level or both – has gone awry.

1.3 *Plan for this paper.* We haven't yet laid out the principles that Boonin himself adopts for purposes of formulating the person-based intuition, the idea, that is, that what is "bad" must be "bad for" someone. That work, which we shall turn to in part 2 below, will help to show why Boonin thinks we are committed to the Implausible Conclusion.

I suggested above that we do, after all, have reason to think that Boonin's approach to the non-identity problem at some point goes awry. The main purpose of this paper, then, will be to explore just how that might have happened.

Part 3 below examines one possible explanation – that Boonin himself hasn't fully appreciated the moral significance of some of the factual details of his own cases. Now, that's not a mistake he seems to make in the Wilma and Pebbles case. But in other cases, including, as we shall see,

Risky Policy, certain factual details – and specifically *modal* details regarding what the agent's *available alternative* options are just prior to choice – are never mentioned and are certainly never explored. Yet – as I will argue – it is a mistake to put blinders on in respect of those details. It is a mistake to assume in advance that those details have *no moral relevance* at all to our evaluation of the choice under scrutiny.

Perhaps, however, Boonin's difficulty arises not out of a failure to appreciate fully the moral significance of the factual details of the various cases but rather out of the particular principles he sets forth for the purpose of articulating the person-based intuition. Part 4 below explores that alternate explanation for the difficulty Boonin finds himself in. There, I will argue that principles that are more sensitive to the factual details of the various non-identity cases – principles that recognize the moral significance of those details rather than simply setting them aside altogether – may enable us to make the distinctions we need to make among the various cases.

The two possible explanations are clearly linked. My argument thus will be that Boonin's understanding of the relevance of the facts of his own cases in combination with the principles he adopts for purposes of understanding the person-based intuition and applying that intuition to those cases reflect an *inappropriately narrow approach* to the problem at hand. Once we set that inappropriately narrow approach aside and instead expand the scope of our both our factual and our theoretical inquiry – once we toss off the blinders – two things happen. First, facts begin to appear on our moral radar that we had previously ignored or considered irrelevant. And, second, principles that previously seemed to capture the person-based intuition in perfectly adequate ways come to seem inappropriately narrow.

Boonin's analysis, in effect, blinds us both to the morally relevant factual details of our own cases and to alternate formulations of the person-based intuition. A more expansive view is in order – if, that is, the person-based approach is to have any hope of success at all.

We may still find ourselves on Boonin's side in the case of Wilma and Pebbles. But we may at the same time be able to open the door to a finding of wrongdoing in those versions of the non-identity problem in which the

choice under scrutiny seems to us clearly to be wrong. [15]

2. BOONIN'S ARGUMENT

According to the person-based intuition, what is "bad" is "bad for" some existing or future person. Boonin unpacks the overall structure of the intuition in the form of premises (4) and (5) of the non-identity argument he constructs for the Wilma and Pebbles case: if an act harms no one, it wrongs no one; and unless an act wrongs someone or another, it's not wrong.[16]

I have no particular quibble with Boonin's work in that connection – that is, with Boonin's premises (4) and (5). For my purposes here, I will focus on how he unpacks not the overall structure of the intuition but just the *necessary condition* that the intuition sets forth. When, on Boonin's view, is that necessary condition satisfied? When does a choice *harm* a given person?

That question is answered by Boonin's premises (1) and (2). Here, I

15 I do agree with Boonin that Wilma's choice is permissible. But I want to underline that I do not accept Boonin's view that almost anything goes when it comes the burdensome, existence-inducing, choices that agents may make. Contrary to the position Boonin attributes to me, I have consistently rejected that position – including its application in cases involving commercial surrogacy and human cloning, with what I meant to be a vengeance. Boonin (2014): 14 n. 21; Boonin (2014): 189-290 n.1; cf. Roberts (2009), Roberts (2007), Roberts (1998), Roberts (1996).

It's true that I have defended a person-based form of maximizing consequentialism. Such an approach has seemed attractive to me in part in virtue of the support it provides to Narveson's famous dictum that "[w]e are in favour of making people happy" but "neutral about making happy people." Narveson (1976): 73. (Narveson's position is helpful in answering the question whether, other things equal, we are obligated to bring ever more happy people into existence. Narveson thinks, and I think, we are not.) But my defense of the person-based approach largely hinges on a demonstration that, in the sorts of cases where it's very clear to us that what has been done is morally wrong, the reasoning behind the nonidentity problem itself indeed goes awry. On my view, Boonin's nonidentity argument, applied, e.g., to Parfit's Depletion and Risky Policy cases, Kavka's Slave Child and Pleasure Pill cases, cases involving historical injustices and, now, the Warren and Percy case, in fact fails.

Boonin is correct to include me in that very small class of philosophers who think that a person-based approach can be made to work. But I think it can be made to work because the approach can be articulated in a way that gives it the resources that it needs to avoid the result of permissibility in the nonidentity cases in which it seems very clear to us that what has been done is morally wrong. Thus Boonin is mistaken to include me in that still smaller class of philosophers who, when it comes to our burdensome, existence-inducing choices, and specifically those we can agree are clearly wrong, claim to accept the Implausible Conclusion.

16 Boonin (2014): 5.

generalize on the basis of the argument he makes in the Wilma and Pebbles case.

> (1) In the non-identity cases, making the choice under scrutiny c rather than an alternate choice c' doesn't make the person p – e.g., Pebbles – burdened by c worse off than p "would otherwise have been"; and
>
> (2) A choice c harms a person p only if c makes p worse off than p "would otherwise have been."[17]

An assumption of the argument – an assumption that comes to us in the form of Boonin's premise (3) – insures that we need not worry about anyone other than the future person. We are thus to assume in the case of Wilma and Pebbles that neither Wilma nor any other existing or future person is made worse off, or harmed, by Wilma's choice to bring the blind Pebbles into existence rather than the sighted Rocks.

The combination of these five premises do indeed seem to generate the conclusion that Boonin suggests: that Wilma's choice is morally permissible.

As noted, my interest here is in how Boonin formulates the necessary condition the intuition is meant to establish.[18] In part 4 below, I will argue that that formulation is inappropriately narrow.

The discussion there will focus on the conceptual component of the non-identity problem. But I'll start in part 3 below by taking a look at the factual component of the non-identity problem. Does Boonin leave some of the factual details of some of his own cases – details we can't assume in advance to have no moral relevance to our evaluation of the choices under scrutiny – unmentioned or unexplored?

17 Boonin (2014): 3-4.

18 I begin here – in part 3 – with the factual analysis. In part 4, I turn to the theoretical analysis. I will there argue that Boonin's understanding of how the person-based intuition's necessary condition operates – as reflected in his own formulation of that condition – is inappropriately narrow. It blinds itself to the morally critical factual details of the relevant cases. Perhaps it's that the failure to appreciate those details from the beginning leads Boonin to adopt a crude formulation of the condition. Or perhaps the explanation goes in the other direction: perhaps the problem is that the crude formulation never requires us to look closely at the morally relevant details of our own cases. No matter. If the person-based intuition is to be made plausible at all, then the necessary condition that it asserts must be formulated in a way that is sensitive to the morally relevant details of the various choices that it deems morally permissible.

3. THE NARROW VIEW OF THE FACTS OF THE NON-IDENTITY CASES

3.1 *Direct version*. Consider, again, the Warren and Percy case. In my original description of the case, I didn't explicitly note that, in addition to (1) the option of introducing the chemical into the dish and allowing fertilization to proceed and the new person, Percy, to come into existence and (2) the option of discarding the contents of the dish, Warren also had (3) the option of not introducing the chemical into the dish yet still allowing fertilization to proceed and Percy to come into existence. It's a stipulation of the case that, had Warren not chosen option (1), he *would* have chosen option (2). But it doesn't follow from that stipulation that option (3) is not a perfectly available alternate choice for Warren at the moment just prior to choice. Indeed, the most natural reading of the case is that that alternative existed for Warren but he simply declined to avail himself of it. Perhaps he declined because he then wouldn't have had the chance to challenge accepted wisdom as to just how dangerous the chemical really is; perhaps he wouldn't have gotten the $100. But none of that even begins to suggest that that third alternative option did not exist for Warren. At least: we have no reason to think it doesn't. And it's a choice that would have made things better for Percy than things in fact are – as, that is, things are for Percy, given Warren's choice of (1).

This is not to suggest that Warren had the *obligation* to choose (3) rather than (1). Indeed, any plausible formulation of the person-based intuition will generate the result that Warren had no such obligation.[19] It's just to say that the fact that, for all the reader is told, (3) is an alternative for Warren is a detail of the case. Moreover, it's a detail that we cannot assume in advance to be irrelevant to the evaluation of Warren's choice of (1).

Thus, there is a distinction between the Warren and Percy case and the Wilma and Pebbles case: in the Wilma and Pebbles case, as Boonin describes that case, it really doesn't seem that the agent has any third option, but in the Warren and Percy case, it seems the agent does have a third option. The one case is thus a two-option case, and the other a three-option case. We can't, in advance, take it for granted that that distinction itself does not mark a morally relevant distinction between the two cases.

3.2 *Indirect version*. As noted earlier, Boonin's presentation of the non-identity problem – in the form of the Wilma and Pebbles case – is clean and clear. The difficulty is that the many versions of the non-identity problem – those in which we want to say the choice is wrong but in which the most

19 See note 15 above.

plausible victim of the wrongdoing may seem not to have been made *worse off*, or *harmed*, at all – *aren't* clean and clear. Rather, they are complicated and a bit murky.

A good example of such a version of the non-identity problem is the Risky Policy case. Thus consider the "wealthy society" that is "running out of fossil fuels."[20] That society faces the choice between two sources of energy, one that would allow existing people to retain their standard of living and have no effect on future people and one that would somewhat improve the standard of living for existing people but also generate a "significant amount of toxic waste," an amount that would eventually "painlessly" kill "tens of thousands" of future people "once they reached the age of forty."[21] It's stipulated that "over time, the effects of [the] subtle differences" in when and with whom people conceive their children "will be enough to generate two entirely distinct sets of people: the set of people who will exist five hundred years from now if the safe policy is selected, and the completely different [non-identical] set of people who will exist five hundred years from now if the risky policy is selected."[22] It's thus a stipulation of the case that, if "the wealthy society had instead selected the safe policy, those innocent people would never have existed in the first place."[23]

On the basis of those facts, Boonin notes the following:

> *The No Worse Off Claim*. The "innocent people who are killed as a result of the leaking toxic waste are not made worse off by the wealthy society's choice of the risky policy" than they would have been had the wealthy society chosen the safe policy instead.[24]

The No Worse Off claim is actually just premise (1) of the non-identity argument that Boonin wants to say is sound. But our interest in the No Worse Off Claim at this juncture is not to ask whether it's true or not. Rather, what is of interest here is what the claim reveals regarding Boonin's view of the facts of his own case – specifically, that Boonin's view of the case is itself very narrow.

Consider any possible person – say, *Polly* – who is a member of the class of people who will come into existence and die painlessly at age forty if the risky policy is chosen and who would never have existed at all if the safe

20 Boonin (2014): 5.
21 Boonin (2014): 5-6.
22 Boonin (2014): 6.
23 Boonin (2014): 6.
24 Boonin (2014): 6. This assertion would fit the pattern Boonin provides for premise (1) of the non-identity argument. Boonin (2014): 3.

policy is chosen. We should agree with Boonin that it's better – or at least not worse – for Polly to come into existence and die painlessly at age forty than it is for Polly never to exist at all. And we should accept the counterfactual stipulation that, had the safe policy been selected instead of the risky policy, Polly *would* never have existed at all. And for the sake of clarity let's just suppose that the risky policy has in fact been selected and that Polly in fact qwill exist and will die painless at age forty.[25]

The easiest way to see that Boonin's presentation of the facts of the case is at least arguably overly narrow is to reformulate what we have just said in the language of possible worlds. Thus: in the actual world – call it *w1* – in which the risky policy is in fact chosen, Polly eventually exists but is doomed to die painlessly at age forty, and in the closest possible world where the safe policy is chosen instead – call it *w2* – Polly never exists at all. Even given that it's better – or at least not worse – for Polly that Polly exist and die painlessly at age forty than it is for Polly never to have existed at all, all that we can infer from what we have said so far is that the choice of the risky policy *as that policy is put into effect at w1* doesn't make things worse for Polly than the choice of the safe policy *as that policy is put into effect at w2*. We can't, that is, on the basis of what we have said so far rule out that there exists a world – call it *w3* – in which the safe policy – or at least a *safer* policy – is itself implemented *in some alternate* way that both brings Polly into existence and makes Polly better off than Polly is under the risky policy as that policy is implemented in w1.

Boonin's own presentation of the case has us take a narrow view; it has us put our blinders on. We are to attend exclusively to the risky policy at w1 and the safe policy at w2. We determine that the one is better – or at least as good – for Polly as the other.

But now let's remove the blinders. We then recognize that there's more than one way to implement the safe policy. The safe policy can also be implemented in the way it is implemented in w3.

This is not to add facts to Boonin's case. This is just how the world is. The "precariousness" of any particular person's coming into existence does not mean that there's *only one possible route* for that person ever to come into existence at all. In fact, the choice is made to adopt the risky policy – and as a result of that policy Polly eventually exists and dies painlessly at age forty. That is what happens at w1 – at, that is, the *actual* world. But it *could* have happened – specifically, it's physically and

25 Boonin's language on this point goes back and forth between a simple material conditional and the corresponding counterfactual. For purposes of evaluation, this distinction doesn't seem material.

metaphysically possible – that the agents who made the one choice at w1 have chosen the safe policy instead and then implemented that policy in a way brings Polly into existence and makes Polly better off than she is in the actual world. That possibility – that alternate outcome – is represented by w3.

And that's a factual detail that's left unmentioned but would seem to be implicit in the case. It's a *modal* detail, to be sure, involving what could have been. But it's still part of the case.

The upshot? While it's clearly true that the risky policy *at w1* makes things *better for Polly* (or at least not worse for her) than the safe policy *at w2*, it's also clearly true that the safe policy *at w3* makes things *still better* for Polly than the risky policy at w1.

But if that's so, then it would be a mistake – starting out – to assume in advance that that particular factual detail isn't relevant to our moral evaluation of the choice of the risky policy. It would be a mistake to assume in advance that we can complete the moral evaluation without taking that detail into account. Now, perhaps, consistent with all that we have said so far, that detail will turn out *not* to be critical to our evaluation of the case. My point here, however, is that it's a mistake to assume – starting out – that it's not relevant to our evaluation of the case.

Familiar "same people" cases help to underline that fact. Consider, for example, Medical Case I. Suppose that the doctor has the choice to treat you with medicine A, not to treat you at all or to treat you with medicine B. Suppose that the doctor chooses medicine A and that, while medicine A saves your life, the life you are left with is, though clearly worth living, substantially diminished. It's nice that the doctor has saved your life, given that (let's suppose) you'd have died had the doctor not treated you at all. But it's surely morally relevant to our evaluation of the doctor's choice to treat you with medicine A that medicine B would have *both* saved your life *and* restored you to full health.

This is not to contest the No Worse Off Claim as Boonin has articulated that claim in connection with Risky Policy. The claim is perfectly true.[26] The risky policy at w1 is better for Polly – or at least not worse for Polly – than the safe policy at w2. But the claim that the safe policy at w3 is better for Polly than the risky policy at w1 is also perfectly true. Any assumption that the first detail regarding the case but not the second will be morally

26 The language of possible worlds certainly helps to avoid any quibbles in this context. In contrast, any implication that as a matter of physical necessity p can't exist under the choice of the safe policy – that in all physically accessible possible worlds the choice of the safe policy, or at least a safer policy, results in p's never existing at all – would be highly problematic.

relevant to our evaluation of the choice of the risky policy would seem to be – and clearly is, on my view – illegitimate. Of course, a critic may argue that I have made things too complicated and that a stipulation or two will clean things up in a way that shows that we should have no objection to Boonin's – or anyone else's – narrow view regarding of the facts of Risky Policy. Indeed a number of superficially appealing strategies for justifying the narrow view are available. I will consider three such strategies here.

(i) *Counterfactual stipulation.* It might be argued that my insistence on a more detailed description of Risky Policy ignores a main stipulation of the case: that had the risky policy not been chosen, then the safe policy would have been selected in its place and Polly would never have existed at all. We can articulate that stipulation in the language of possible worlds. The risky policy is implemented in the actual world w1 and Polly eventually exists there. Moreover, at the closest possible world w2 at which the risky policy isn't implemented and the safe – or a safer – policy is implemented instead, that policy is implemented in such a way as to leave Polly out of existence altogether.

The counterfactual stipulation may seem to provide a rationale for Boonin's narrow view of the case. It may seem to ground the position that the fact that the safe policy as that policy is implemented at w3 is irrelevant to the moral analysis. In fact, however, counterfactuals can seem morally telling when they aren't. In Medical Case I, the doctor who treated you with medicine A may offer the defense that he has saved your life. And indeed he did. He may even convince us that the relevant counterfactual holds: that is, that he would have declined to treat you altogether (perhaps exiting the hospital in a moment of pique) and you would have died had he not treated you with medicine A. We might (momentarily) be duped into thinking that the doctor is, after all, both morally and legally off the hook.

But he's not. What he's said in his defense is perfectly consistent with the further point that treating you with medicine B would have restored you to full health and thus left you better off than you in fact are. That that option exists is a function of what the *world* – metaphysically, physically – is like; it's not a function of what the *doctor counterfactually* would or would not choose to do. It's stipulated that treating you with medicine B isn't an option the doctor *would* have chosen. But it doesn't follow from that that it's not an option available to the doctor. The world, not the doctor, determines what the doctor's options are.

We clearly think that the factual detail involving the option of treating with medicine B has moral salience in Medical Case I. There's no basis for concluding that the analogous factual detail in Risky Policy – the detail involving the many different ways in which the safe policy might itself be

implemented; the detail that makes it so that w3 is an alternate possible way in which the future might have unfolded – doesn't have moral salience in that case.

(ii) *Stipulation of physical necessity.* Alternatively, Boonin might have meant his reader to understand that it's a stipulation of the case that, as a matter of the physical law, the safe policy can be implemented in only one way, a way that inevitably and unavoidably leads not to w3 but rather to w2, where Polly never exists at all.

It's true we could devise a case in which the stipulation of physical necessity holds.[27] For example, we could add to the case that the toxic waste produced under the risky policy in w1 – exactly that degree of risk and exactly that degree of toxicity – is physically *necessary* to Polly's ever coming into existence at all given her genetic makeup.

Now, Boonin doesn't seem to have that particular story line in mind – nor did Parfit, in his presentation of his own version of Risky Policy. Unlike the counterfactual stipulation, the stipulation of physical necessity seems completely foreign to the case that Boonin presents. It's not one the reader can be expected to bring to the table. If it's to be included in the case at all, it should be included *explicitly*.

But if we drop the stipulation, we shall need to accept that the choice of the safe policy can, after all, be implemented in more than one way – that is, that as of that moment just prior to choice the future isn't carved out between two and only two paths forward. That's just the way the actual world happens to be, given the physical laws that in fact govern that world. The safe policy doesn't *as a matter of physical necessity* mean that Polly *can't* exist. We dwell in more possibility than *that*. Dropping the stipulation in effect returns us to the earlier account of the case sketched above.

Of course, nothing forces us to drop the stipulation. That is: we can always construct another version of the case, a version that – explicitly – includes the stipulation and thereby strips w3 out of the picture altogether.

It seems clear, however, that that revised case is very different from the case Boonin starts us off with. In Boonin's case, situated (sans stipulation) in the world that is in important respects very much like the actual world,

27 It's possible Boonin himself intends Risky Policy to include the stipulation of physical necessity. Moreover, as a referee for this paper points out, "there are no facts to the case" beyond the facts the author of the hypothetical describes; it's Boonin's case to detail as he pleases. At the same time, Boonin doesn't explicitly include the stipulation in his description of the case. And the case where the stipulation isn't included and the case where the stipulation is included are two distinct cases. In this part 3.3.2, I am considering how the case plays out both ways – first, in this subpart (ii), as it plays out with the stipulation in place and, second, in subpart (iii) below, as it plays out without the stipulation in place.

we think it's clearly wrong for agents to choose the risky policy. But in the revised case it's not at all clear the choice is wrong. In the revised case, the risky policy is more like a fertility treatment, a treatment that produces a certain adverse side effect for Polly but is physically necessary to Polly's ever coming into existence at all. (Perhaps the genetic materials that eventually develop into Polly are such that cell division in the early days after conception *cannot* happen in the absence of the toxic waste; or perhaps the physical laws that govern the actual world simply don't apply.) In that case, Polly *can't* exist and *not* die painlessly at forty, just as I can't exist and not die, I hope painlessly, within the next few decades.

It's just not clear that on those facts – all known to the agents and all matters in respect of which agents have certainty – the choice of the risky policy is wrong. After all, we accept that the choice of a fertility treatment, even one that has certain adverse side effects, may be morally permissible. We may also accept that the choice of the risky policy – made in some distant possible world in which human reproduction physically *requires* exposure to toxic waste in a way that in the actual world human reproduction physically *requires* exposure to certain hormones – is permissible as well.

(iii) *Relevance of probabilities*. A third strategy for justifying Boonin's narrow view regarding the facts of Risky Policy concedes that, technically, w3 exists as an available option but at the same time insists that the *probability* that w3 will unfold – that is, that Polly will exist – given the choice of the safe policy is *extraordinarily* low and thus can, and should, be disregarded altogether for purposes of evaluation.[28]

That probabilities may be relevant to how we evaluate choice is demonstrated in Medical Case II. In that case, the doctor is deciding between medicines A and B. The doctor is fully informed, let's suppose, as to the following facts. First, the probability that medicine A will restore the patient to one degree of good health or another within a relatively narrow range – as opposed, let's say, to killing the patient – *is truly significant*. Whatever it is, it's a probability that the doctor should take into account in making his choice. Second, the probability that medicine B will restore the patient to some *greater* degree of good health but still within that same narrow range is also truly significant. It, too, is a probability the doctor should take into account in making his choice. And third, the probability that medicine C will restore the patient to an even greater degree of good health but still within that same narrow range is *very low, practically nil*. In

28 As one referee for this paper points out, the idea that Boonin's evaluation of the choice in Risky Policy turns on the relevant probabilities "comes rather out of the blue." That's so. The effort here, then, should be understood as an effort to understand how Boonin's argument might be reconstructed.

those circumstances, certainly the doctor needs to compare the different outcomes made possible by medicine A on the one hand and medicine B on the other, along with the probabilities that accompany each, in order to make his decision. But does he really need also to compare A against C and B against C, given the probabilities? No. Rather, it seems perfectly appropriate in that case for the doctor to decide what to do on the basis of the comparison of A against B and on that basis alone.

Now, for purposes of completing that comparison, it will be important to have detailed information regarding the relevant possible outcomes and the relevant probabilities. If the outcome under medicine A would be *better* for the patient than the outcome under medicine B, and if the (significant) probability of success under medicine A is also *greater* than the probability of success under medicine B, then surely medicine A is the correct choice.

We should, in other words, clearly recognize that probabilities may well be relevant to moral evaluation. However, what is often forgotten in discussions of the non-identity problem is that it's notoriously easy to make mistakes in calculating the relevant probabilities.

Setting C aside in Medical Case II may make sense, given that the probability of success for the patient under C is – by hypothesis – very low and practically nil. It is much lower than the probabilities of success for the patient under A and B which are – by hypothesis – significant. *Probability differentiation* is, in other words, in effect a stipulation of the case.

But Risky Policy is very different. We don't see – at least, we are making things up if we think we *do* see – any analogous probability differentiation in the facts of Risky Policy.

Why is that? Isn't it just *part* of Risky Policy that the probability that Polly will exist under the choice of the risky policy is significant while the probability that Polly will exist under the choice of the safe policy is very low, practically nil?

The latter issue – just what the subject's chances of existence are under the safe policy – is occasionally explicitly resolved in various descriptions of various non-identity cases.[29] But the former issue – just what the subject's chances of existence are under the risky policy – almost never is. Rather, what is typically included in those descriptions is just that, given the choice under scrutiny, a particular person (e.g. Polly) *in fact* comes into existence and *in fact* suffers in a certain way. The question of the *probability* that Polly will exist under the choice of the risky policy is never even raised,

29 See Kavka (1982): 100 n. 15.

much less answered.

So let's raise it now. Consider Polly, who, by hypothesis, exists and suffers as a result of the choice of the risky policy. What were Polly's chances of coming into existence, calculated at the moment just prior to that choice, given the choice of the risky policy? To think that Polly's chances of coming into existence, as of that moment just prior to choice, were anything other than very low, practical nil – to think that they were *significant* – is to fail to appreciate all the many ways that agents could have, as of that moment just prior to choice, gone about implementing the safe policy. Just as the fact that there existed many, *many* ways of implementing the safe policy just prior to choice meant that Polly's chances of coming into existence under that choice are very low, so does the fact that there existed many, *many* ways of implementing the risky policy just prior to choice mean that Polly's chances of coming into existence under that choice were very low.

To think otherwise is to think of the risky policy as a sort of fertility treatment that Polly's forebears might have unwittingly undergone – to think of the risky policy as having a sort of existence-inducing *power* in respect of the particular person Polly's coming into existence.

Can't we just stipulate that that's so? Of course. But then what we are calling "Risky Policy" is no longer Risky Policy. "Risky Policy" is now a fertility treatment case. And we do accept some adverse side effect, in connection with fertility treatments, when the treatment itself makes it more likely that a couple will conceive a child whom they badly want. And we evaluate fertility treatment cases in light of that fact. But Risky Policy in its original form is *not* a fertility treatment case.[30]

I noted earlier that philosophers' descriptions of the various non-identity cases typically include the fact that, the choice under scrutiny having been made, a particular person, e.g. Polly, in fact exists and suffers. In that connection, it's important to underline that it's a mistake to judge how *probable* an outcome is given a particular choice on the basis of how the future *in fact* unfolds. Rather, the calculation of probability, for purposes of the moral evaluation, should be based on information available to the agents just prior to choice, the whole point of that calculation being to provide the agents with a means, prior to choice, for determining

30 When it comes, e.g., to cases involving historical injustice, it may be that the choice we all agree is wrong – e.g., the choice to order or to participate in the implementation or continuation of the Holocaust – in fact makes it *less* likely that particular offspring of the contemporaneous victims of that wrong choice will ever come into existence at all. The same point applies to at least some other versions of the non-identity problem, including Depletion and Risky Policy.

permissibility.[31]

My argument, then, is that there is no basis for thinking that the choice of the risky policy, as of that moment just prior to choice, generates any significant probability of Polly's coming into existence at all – and certainly no greater a probability than the choice of the safe policy would have generated for Polly. For both choices, there are just too many ways of implementing that choice – the risky policy on the one hand, the safe policy on the other – to give Polly *any significant chance of existing at all*.[32]

31 An assessment of the probabilities typically is brought to bear to explain why the choice that looks to be the morally correct choice just prior to choice can't subsequently be dismissed as morally wrong when the future unfolds in some way that no one could have predicted. (Suppose by some quirk the seatbelt I've asked you to put on ends up strangling you when we are hit by a Mack truck speeding down the highway in the wrong direction; we don't think on those facts my asking you to put your seatbelt on is morally wrong.) Or why agents are not thought to have done anything wrong in the case where a future that would have been better for a particular person existed as a physical possibility for the agents but the agents has no clue how to bring that physical possibility about. (I *can* save your life by performing open heart surgery on the subway, but I have no idea how to go about that; we think on those facts my instead calling 911 and performing CPR is a perfectly acceptable alternative.) An assessment of probabilities, in other words, typically arises once it's determined that an expected value calculation is in order. That means, in turn, that the probabilities we should be focusing on are those we calculate (1) just prior to choice – at no later or earlier point – and (2) just on the basis of information available to the agents at that moment.

32 A referee for this paper suggests that my line of reasoning in this connection may be challenged by an alternate way of coming at the calculation of the probabilities at issue. According to that alternate approach, we first recognize that some person or another will exist (and suffer) under the choice of the risky policy. We then simply ask how likely it is that that very same person, "whoever it is," would have existed under the choice of the safe policy. Of course, the answer to that question is that the chances of that person's existing under the safe policy are very low. The suggestion, as I understand the referee's comment, is that we are then to compare the fact of that person's (burdened) existence under the choice of the risky policy against the very low probability that that same person would exist (and have a relatively unburdened existence) under the choice of the safe policy. It may very well be that many philosophers have – assuming they have contemplated the various probabilities at all – understood the analysis to proceed in just that way. Stepping back, however, I think we can see that on this approach what we are asked to compare is the value the future that *in fact* unfolds for that person – arbitrarily, p – given the choice of the risky policy against how the future can be *expected* to unfold for that same person p given the choice of the safe policy. We are asked to compare, that is, an *actual* wellbeing level against an *expected* wellbeing level – a ex post fact against an ex ante projection. But the principle that would deem such comparisons relevant for determining the choice that is better for the subject will lead to inconsistency, potentially directing in, e.g., a medical case in which the patient's chances of survival are low whatever choice is made, that medicine A is better for the patient than medicine B (A producing more actual value for the patient than B produces expected value) and that medicine B is better for the patient than medicine A (B producing more actual value for the patient than A produces expected value). I take no position here on whether an actual value calculation of wellbeing or an expected value calculation is correct for purposes of moral evaluation. But it does seem clear that we can't mix the two. For further discussion, see Roberts (2009); Roberts (2007).

But that means that the probabilities don't, after all, provide a basis for Boonin's narrow view. We can't disregard the third option on the ground that it's highly improbable that Polly will ever exist at all under the choice of the safe policy. For it's just as highly improbable – as of that moment just prior to choice – that Polly will ever exist under the choice of the risky policy. We thus can't assume, on the basis of the probabilities, that the third option won't have relevance for purposes of our evaluation of the agents' choice of the risky policy.[33]

4. THE NARROW VIEW OF THE CONCEPTUAL COMPONENT OF THE NON-IDENTITY PROBLEM

We now turn to the conceptual component of the non-identity problem – in particular, how we should formulate the person-based intuition's necessary condition on wrongdoing. My argument here will be that Boonin's narrow view regarding how that condition is to be formulated is a mistake.

4.1 Concept of harm. Let's begin with a close look at Boonin's premise (2). Generalized:

> (2) A choice c harms a person p only if c makes p worse off than p "would otherwise have been" had c not been chosen.[34]

Boonin's reference to the concept of harm in his formulation of the person-based intuition seems to me to be unproblematic[35] – as long as we get the concept of *harm* right. Thus: the intuition itself would be that a choice is wrong only if it *harms* a person who does or will exist given that choice.[36]

Moreover, the comparative account of harm that Boonin proposes also seems to be on the right track. Pain itself, e.g., isn't enough to show harm (in, as Parfit put it, a morally relevant sense). A vaccination may inflict *pain* but shouldn't *harm*. Harm requires, rather, that the person harmed has been made *worse off*.

33 I have made a similar argument elsewhere. See Roberts (2009); Roberts (2007).

34 Boonin (2014), pp. 3-4. Boonin, referring to a "commonsense understanding" of harm, words the point this way: "If your act harms someone, then it makes that person worse off than they would have been had you not done the act." Boonin (2014): 3.

35 This we find in his premise (4).

36 Here, I follow Boonin is sharply distinguishing harming from wrongdoing. Harming is a matter of (comparatively) reducing a person's wellbeing – of a person's having less wellbeing in one possible future, or outcome, than in another. On that view, a person may be harmed without having been wronged. Philosophers who want to under harm itself to come with clear normative implications may want to understand my "harm," and Boonin's, to constitute something like a comparative loss of wellbeing rather than a bona fide harm.

The difficulty for Boonin is that the particular comparative account of harm that he offers seems overly narrow. It compares how things have unfolded for a particular person under a particular choice against how things would have unfolded had that choice not been made. It elevates, in other words, the counterfactual stipulation – in, e.g., the Warren and Percy case and the Risky Policy case – to the status of primary conceptual significance. And that's surely a mistake.

Thus Boonin writes: "What seems to be a commonsense understanding of what is required in order for an act to harm someone" is the following:

> if your act harms someone, then it makes that person worse off than they would have been *had you not done the act*. That this is a widely accepted view seems to be confirmed by the fact that if a person is accused of harming someone, it is standardly taken as a sufficient rebuttal to the claim if they can establish that they have not made their alleged victim worse off than they would otherwise have been ["had you not done the act"].[37]

To see that this account of harm won't do, let's go back to Medical Case I.[38] It's part of that case that, had the doctor not treated you with medicine A, he'd not have treated you at all (perhaps exiting the hospital in a moment of pique). His defense is that he has not made you, the alleged victim, worse off than you "otherwise would have been" – indeed, that he's saved your

37 Boonin (2014): 3. Emphasis, and square-bracketed material, added.

38 Boonin's account of harm is thus very different from my own. Thus Boonin's account is unable to identify harm in either Medical Case I or the Warren and Percy case. In both those cases (as a referee for this paper has pointed out) the agent in a sense makes it the case that the impaired child must suffer the impairment or never exist at all. On Boonin's account, that is enough to show that the child isn't harmed. On my account, that the agent would decline to make a third choice that would have made things still better for the child is irrelevant to the analysis – a red herring. It's what the agent could have done – physically, metaphysically – that is salient to the case. The world, not the agent, determines what the agent's options are. In both these cases, that that third choice existed as an alternative for the agent shows that the one choice harms the child.

Of course, it's also part of the person-based approach that it's permissible never to bring the child into existence to begin with. Perhaps that choice itself can appropriately be said to harm the child or at least to cause the child to sustain a loss of wellbeing. I would agree – but also say that, under a person-based approach, the sorts of harms or losses we sustain by virtue of our never having existed are morally without any significance at all. The principle that losses should be distinguished in that way I call the Loss Distinction Thesis (or, in earlier work, Variabilism). Roberts (2011a), (2011b).

It should be noted that cases involving causal over-determination and preemption challenge Boonin's account of harm as well. While I don't survey such cases here, any plausible account should be able to identify harm in those cases. Thus on my view the bare fact that a group of agents had the option of creating more wellbeing for a given person p may be enough to show, depending on the details of the case, that the choice of any one member of that group by virtue of his or her "participation" in the group harms p. Roberts (2006).

life. And it's perfectly true that had he not "done the act" – had he not treated you with medicine A – you'd have died. Thus, he has not *harmed you*.

But how plausible is the doctor's defense? *Moral* law aside, *human* law would certainly scoff at such a defense and insist rather that a mistake has been made in identifying the appropriate *baseline* (a concept Boonin, and before him Feinberg, freely relies on). For purposes of the legal analysis, the baseline isn't what the doctor *would* have done to you – that is, still worse – had he not done what he did. Rather, it's what the doctor *could* have done for you given the options available to him at the critical time. Thus, the difference between your diminished life under medicine A, and the life you would have had under medicine B, a treatment that would have restored you to full health, provides the appropriate measure of harm in Medical Case I.

It's true that the choice under which you have the diminished life is better for you than the choice in which you simply die. But that truth is just a red herring: it tells us nothing about the harm that is done to you by the choice to treat with medicine A at all.

The case of Warren and Percy makes the same point. It's a stipulation of the case that had Warren not introduced the chemical into the dish, thereby causing Percy to be born with an impairment, Warren would have simply tossed the contents of the dish. What Warren in fact does nonetheless *harms* Percy. And what tells us that that's so is that Warren has a third option: he could have refrained from introducing the chemical into the dish and nonetheless allowed fertilization to proceed and Percy to be born. If we fail to attend to that factual detail of the case – or if our formulation of the relevant person-based necessary condition on wrongdoing sets that particular factual detail of the case aside as irrelevant to the moral evaluation – then that evaluation itself will be unreliable.

There's no reason to think that that more expansive concept of harm isn't the correct concept for purposes of formulating our person-based principles as well. Yes, the principles will be a (tiny) bit more intricate. But that's a small price to pay to avoid a principle that generates clearly false results across a wide range of cases.

The proposal here, then, is that Boonin's overly narrow analysis of harm – what we see in his premise (2) – be rejected, in favor of a more expansive understanding of harm. Thus:

> Expansive premise (2): A choice c at a world w harms a person p only if there is some choice c' that exists as an option for the agent (or agents) and some alternate accessible world w' such that c at w is worse for p

than c' at w' is.

That is: a choice harms a person p only if that choice makes things worse for p than at least some available alternate choice makes things for p – only if, that is, agents could have done more for p than they in fact have. If c is maximizing for p, in other words, it doesn't harm p.

Now, moving from harming p to wronging p – or to wrongdoing – isn't an easy task; all kinds of tradeoffs need to be taken into account. But at least on this conception of harm, the door is wide open to a finding of harm – which means the person-based intuition is wide open to a finding of wrongdoing – in all the right cases: in, e.g., the Warren and Pebbles case and the Risky Policy case. We still find ourselves in the controversial position – with Boonin – of declaring Wilma's choice perfectly permissible. But perhaps that's not such a bad place to be. (How could we live without Pebbles?)

4.2 *The red herring.* Let's now turn to Boonin's premise (1):

(1) In the non-identity cases, making the choice under scrutiny c rather than an alternate choice c' doesn't make the person p – e.g., Pebbles – burdened by c *worse off* than p "would otherwise have been" under c'.

And let's just note that, once we restore the counterfactual stipulation itself to its rightful place – once we focus, not on what the agents would have done, had they not done what they in fact did, but rather on what agents could have done instead – we can see that the comparison that premise (1) asks us to make – the comparison that launches Boonin's own version of the non-identity problem – may be (depending on the factual details of the particular case) just a red herring.

Let's step back. In the Wilma and Pebbles case, the comparison of how things have turned out for Pebbles and how they would have turned out had Wilma not chosen as she did is – if misleading – relevant. It's relevant, in virtue of the fact that in that case Wilma just doesn't have the option of making things still better for Pebbles than they are under Wilma's choice not to delay the pregnancy. As argued in part 2 above, we need to *say* that: we can't *assume* that no further alternatives exist, or that if they do exist they aren't relevant to the moral evaluation. But in this particular case that happens to be so.

But in cases where still better alternatives do lurk – such as in the Warren and Percy case and in the Risky Policy case – (1) itself is just a red herring. It's always nice to hear that one option is better for a person than another, but until we know whether there's a better option still we are in no position to evaluate the one option.

5. CONCLUSION

I have argued that Boonin has been mistaken to take a narrow view of the cases that give rise to the non-identity problem. That narrow view causes him to overlook certain factual details – including critically important *modal* details – that are part of the cases that he himself develops. It may have led him as well to think that how we formulate the person-based necessary condition on wrongdoing is a simpler matter than it in fact is.

A way of summing up my argument is to say that Boonin's treatment of the most challenging non-identity cases – those (unlike the Wilma and Pebbles case) in which the choice under scrutiny is clearly wrong – mistakenly describes those *three*-option cases as *two*-option cases. He then – mistakenly again, in my view – tailors his principles accordingly.

While I have distanced myself from the conclusions Boonin reaches in many of those cases – and in particular from the Implausible Conclusion – I should underline again that I have a great deal more sympathy for Boonin's position as it applies in the bona fide two-option case, including, e.g., the Wilma and Pebbles case. Such cases aren't really so rare. Imagine a couple who decides to proceed with a pregnancy even though genetic testing has revealed that the resulting child will be impaired in some way; and imagine, as well, that the case is one in which no one other than the very child whose existence is at stake will be affected by the couple's choice. When such a couple decides to proceed, I am – like Boonin – committed to the position that the choice is morally permissible.

BIBLIOGRAPHY

Boonin, D., 2014: *The Non-Identity Problem and the Ethics of Future People*, Oxford: Oxford University Press.

Kavka, G., 1981: "The Paradox of Future Individuals," Philosophy & Public Affairs 11: 93-112

Narveson, J., 1976: "Moral Problems of Population," in *Ethics and Population*, ed. Michael D. Bayles, 59-80, Cambridge, Mass.: Schenkman Publishing Company.

Parfit, D., 1987: Reasons and Persons, Oxford: Oxford University Press

Roberts, M., 1998: *Child Versus Childmaker: Future Persons and Present Duties in Ethics and the Law*, Lanham, Maryland: Rowman and Littlefield.
—1996: "Human Cloning: A Case of No Harm Done?" *Jo. of Medicine and Philosophy* 21.5: 537-54.
—2006: "Supernumerary Pregnancy, the Problem of Collective Harm and the Nonidentity Problem," *Journal of Law, Medicine & Ethics* 34.4: 776-792.
—2007: "The Nonidentity Fallacy: Harm, Probability and Another Look at Parfit's Depletion Example," *Utilitas* 19: 267-311.
—2009: "The Nonidentity Problem and the Two Envelope Problem," in *Harming Future Persons*, eds. M. Roberts and D. Wasserman, 201-228, Dordrecht: Springer..
Roberts, M., 2011a: "The Asymmetry: A Solution," *Theoria* 77: 333-367.
Roberts, M., 2011b: "An Asymmetry in the Ethics of Procreation," *Philosophy Compass* 6/11: 765-776.

DOI: 10.31009/LEAP.2018.V7.08

Solving the Non-Identity Problem: A Reply to Gardner, Kumar, Malek, Mulgan, Roberts and Wasserman

DAVID BOONIN
University of Colorado Boulder

ABSTRACT

The non-identity problem arises from the fact that five premises that seem plausible yield a conclusion that seems implausible. Most proposed solutions to this problem rest on rejecting one of the argument's premises. In *The Non-Identity Problem and the Ethics of Future People* (2014), I argue against such solutions and propose that accepting the argument's conclusion ultimately provides a more satisfactory response to the problem. In this paper, I reply to some of the main claims made in response to the arguments in my book by the contributors to this symposium. In section 1, I respond to Janet Malek's critique of my discussion of the argument's first premise. In section 2, I consider Molly Gardner's proposal that we reject the argument's second premise in favor of what she calls the existence account of harm, identify some problems with that account, and argue that even if we accept it, this would fall short of solving the non-identity problem. In section 3, I respond to Rahul Kumar's rights-based solution, which rests on rejecting the fourth premise of the non-identity argument, by identifying a trilemma for that solution and arguing that none of the three options it identifies offers a promising route to a successful solution. In section 4, I comment on David Wasserman's discussion of three arguments that might be thought to point to a satisfactory solution by rejecting the fourth or fifth premise of the argument. In section 5, I consider and respond to Tim Mulgan's rule-consequentialist solution, which involves rejecting the fifth premise of the non-identity argument. And in section 6, I respond to Melinda Roberts' claim that even if my approach to the non-identity problem is successful in two-option cases, it cannot be successfully extended to three-option cases.

Keywords: non-identity problem, future generations, procreative ethics, harm, rights.

1. INTRODUCTION

I am grateful to all the contributors for their careful and challenging responses to the arguments in my book. While I can't address everything they say here, I will try to offer some thoughts about at least some of the main points that each of them raises. I will begin by considering whether my treatment of the case of Wilma can survive the scrutiny aimed at my discussions of P1 (Malek), P2 (Gardner), P4 (Kumar, Wasserman), and P5 (Mulgan, Wasserman) of the non-identity argument and conclude by considering whether, even if my treatment of the case of Wilma is successful, there is a problem with my extending it to the case of Risky Policy (Roberts).

1. CHALLENGING P1: A REPLY TO MALEK

P1 of the non-identity argument maintains that Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been. Janet Malek suggests that "Breaking P1 into two sub-premises can . . . offer conceptual clarity about the relevant objections and rebuttals" by distinguishing between those objections that involve challenging the first sub-premise and those that involve challenging the second. The two sub-premises she identifies are (1) The Non-Identity Premise, which says that if Wilma had conceived two months later than she actually did, a different child would have been brought into being and Pebbles never would have existed, and (2) The Life Worth Living Premise, which says that Pebbles is not made worse off than she otherwise would have been by Wilma's decision to conceive if her life is worth living (2019: 18). Strictly speaking, P1 does not follow from these two "sub-premises". What follows is the conditional claim that if Pebbles has a life that is worth living and if having a life worth living is not worse than never existing, then P1 is true. As a result, the sense in which Malek's typology amounts to breaking P1 into two sub-premises is not entirely clear. Her distinction may nonetheless prove useful, though, so I will follow it here in organizing my response to her discussion of the five objections to P1 that I consider in chapter 2 of my book.

Malek relates three of these objections to the Life Worth Living Premise:

the Incoherence Objection, the Equivocation Objection, and the Asymmetry Objection. She characterizes my discussion of these three objections as "relatively convincing in demonstrating that these objections are not sufficient to solve the non-identity problem" (2019: 19) and says that my "arguments about their limitations are correct" (2019: 21). Since she doesn't try to defend any of these objections to P1, I won't say anything against them here. But Malek also says that "Boonin's analyses of these three objections reveal a fundamental conceptual flaw in his account" (2019: 21), and I will try to say something about that concern. The heart of Malek's complaint here seems to be that these three objections "address the question of whether a life is worth living, therefore comparing existence and non-existence, which is *not* the question at the heart of the non-identity problem as described in Wilma's case. Boonin therefore tackles three arguments that are conceptually irrelevant to the problem he wants to solve" (2019: 21).

I find this claim about my account unconvincing because I don't think we can know what question lies at the heart of the non-identity problem in the case of Wilma until we settle on a solution. Consider, for example, the solution that Gardner defends elsewhere in this symposium (2019). On that account, Wilma's act harms Pebbles even if it's true that Wilma's act doesn't make Pebbles worse off than she would otherwise have been. If Gardner's argument is successful, then whether Pebbles is the same person as Rocks is irrelevant to solving the problem and heart of the problem lies in our failure to correctly understand what it is to harm someone. Or suppose Kumar is correct, and that Wilma's act violates someone's rights even if it doesn't harm anyone (2019). If Kumar's argument is successful, then whether Wilma harms Pebbles is irrelevant to solving the problem regardless of whether Pebbles and Rocks are the same person, and the heart of the problem lies in our failure to see the many ways in which an act can wrong someone without harming them. Similarly, it seems to me, one cannot say in advance whether any of the three objections to P1 that Malek dismisses as irrelevant will succeed in explaining the wrongness of Wilma's act. One has to critically evaluate them first. Perhaps it will turn out, for example, that it's incoherent to say that Wilma's act does not make Pebbles worse off than she would otherwise have been but not incoherent to say that Wilma's act does make Pebbles worse off than she would otherwise have been. Or maybe the Asymmetry Objection can show in a satisfying way that Wilma would harm Pebbles more by conceiving her than she would harm Rocks by conceiving him and that this would suffice to show that it was wrong for her to conceive the former child rather than the latter. If arguments such as these turn out to succeed in showing that Wilma's act is wrong, then it will turn out that the relationship between existing and

not existing does indeed lie at the heart of the non-identity problem in the case of Wilma. So, I agree with Malek that these objections are ultimately unsuccessful, but I'm not convinced that the task of critically evaluating them is irrelevant to solving the non-identity problem in the case of Wilma.

While I disagree with Malek about the relevance of the first three objections, I do agree that they are not the strongest objections to P1. The most promising, I think, is one of the two she relates to the Non-Identity Premise, the *De Re*/*De Dicto* Objection as developed by Caspar Hare (2007). The other objection she discusses in that context, the Metaphysical Objection, strikes me as considerably less significant. This is not for the reasons I tentatively endorse in the book and that Malek discusses in her paper, but because of what I go on in the book to refer to as the "more fundamental problem" with the objection (2014: 41-42), a problem that Malek does not seem to address at all. The problem is that virtually everyone who thinks that Wilma's act is wrong thinks it's wrong even if they don't endorse (or even know about) the kind of metaphysical picture such as David Lewis's on which it turns out that "Pebbles" and "Rocks" literally refer to one and the same person. What they think, as I put it in the book, is that "it is wrong for Wilma to conceive a blind child now rather than a sighted child later even if we assume that the sighted child really would be a different person" (2014: 41). And since the Metaphysical Objection claims that the sighted child would not be a different person, it can clearly do nothing to vindicate this conditional judgment.

The *De Re*/*De Dicto* Objection is much more promising in this respect because it presupposes that Pebbles and Rocks are two distinct people. I therefore turn to Malek's critique of my responses to Hare. Hare's argument is grounded in the case of Tess, whose job involves safety regulations relating to automobile collisions (2007: 516). Tess implements some new regulations and one result is that the severity of injuries suffered in car accidents is significantly reduced. But another result is that, because the regulations affect various decisions that drivers make, different accidents occur than would have occurred had the regulations not been implemented, and different drivers are injured as a result. Because of Tess's regulations, for example, when Jill is in an accident, she suffers a relatively minor injury, but if the regulations had not been implemented, Jill would not have been in an accident at all and Jack would instead have been in an accident, which would have caused him a more serious injury than the one that Jill suffered. Now suppose after Jill's accident, we ask whether Tess's regulations made things better for the "accident victim". Well, they certainly didn't make things better for the "accident victim" in the *de re* sense of that term. The accident victim in the *de re* sense is Jill and Jill is worse off because of the

regulations. Had the regulations not been implemented, Jill wouldn't have been in an accident in the first place. But the regulations did make things better for the "accident victim" in the *de dicto* sense of that term: the injury suffered by "whoever ends up in an accident" was less severe than the injury suffered by "whoever ends up in an accident" would have been had the regulations not been implemented. Hare counts on the reader to agree that, in the case of Tess, her obligation is to accident victims in the *de dicto* sense, not in the *de re* sense, and this seems quite plausible.

But, of course, the case of Tess can help us solve the non-identity problem only if it turns out to be relevantly similar to the case of Wilma. Hare argues that it is. His argument rests on two claims (2007: 519). First, the reason that Tess has obligations to accident victims in the *de dicto* sense is that (a) Tess has a special role that gives her special duties to a particular class of people, namely "accident victims," and (b) because which particular people will end up as "accident victims" depends on what choice she makes, she can't help "accident victims" in the *de re* sense; she can only help them in the *de dicto* sense. Second, both of these features of the Tess case that explain why Tess has *de dicto* obligations apply to the case of Wilma: (a) Wilma has a special role that gives her special duties to "her child," and (b) because which particular person will end up as "her child" depends on what choice she makes, she can't help "her child" in the *de re* sense; she can only help "her child" in the *de dicto* sense. If Hare is right that the best explanation of why Tess has *de dicto* obligations entails that Wilma does too, then we can solve the non-identity problem by saying that Wilma's choice made "her child" worse off than "her child" would otherwise have been in the *de dicto* sense.

In my book, I raise two main objections to Hare's solution (2014: 34-8). First, I argue that the *de dicto* obligation that Tess has is best understood as a duty to minimize the harm that accrues to whoever turns out to be an accident victim rather than to maximize the well-being of whoever turns out to be an accident victim: the regulations are justified because as a result of implementing them, Jill suffers less harm than Jack would have suffered. If that's right, then even if Hare is correct that Wilma has the same kind of duties that Tess has, this won't solve the problem because no matter which child Wilma conceives, that child won't be harmed by being conceived. Second, I argue that Hare is mistaken in claiming that Tess has the obligations she has in virtue of (a) and (b) noted above and argue instead that she has the obligations she has because she explicitly consented to take them on when she agreed to take her job. If that's right, then it doesn't matter what kind of obligations Tess has. We can't move from the claim that Tess has *de dicto* obligations to accident victims to the

claim that Wilma has parallel *de dicto* obligations to her child because unlike Tess, Wilma never explicitly agreed to take on such obligations.

In pressing my first objection to Hare's argument, I point out that if we thought Tess's obligation really was to make the outcomes better for accident victims de dicto, rather than to minimize the harm they incur, then we'd have to think that Tess should try to redirect accidents from people who are less healthy to people who are more healthy, even if the more healthy people would suffer larger losses than would the less healthy people, just as long as the healthier people would still be better off after their more serious accidents than the less healthy people would have been after the less serious accidents they would have had. Assuming that implication is unacceptable, we should agree that Tess's obligations are really to minimize harm to whoever ends up in an accident, not to maximize the well-being of whoever ends up in an accident. Malek objects to my *reductio* against Hare here by saying that there is an important difference between the case of Tess and the case of Wilma (2019: 23). But whether there is such a difference is irrelevant to my use of the example. The point of my example is to show that the kind of obligation that Tess has (minimizing harm to accident victims, not maximizing the well-being of accident victims) is one such that, even if Hare is right to say that Wilma has the same kind of obligation that Tess has, cannot show that Wilma's act is wrong (because Wilma's act does not fail to minimize harm to her child). My example shows this regardless of whether there are differences between Tess and Wilma. And Hare is the one arguing that the case of Wilma is analogous to the case of Tess, not me. So Malek's point that there are differences between Tess and Wilma makes things worse for Hare, not for me.

Malek also suggests that my response to Hare here in part depends on circular reasoning (2019: 23): I argue that if I'm correct that Tess's obligation is really to minimize harm to whoever ends up being in an accident, then Wilma's act doesn't violate that obligation even if Hare is correct that Wilma has the same kind of obligation that Tess has. After all, if Wilma conceives Pebbles she doesn't harm Pebbles and if she conceives Rocks, she doesn't harm Rocks. So she can minimize the harm that she causes to whoever ends up being her child by conceiving either child. Malek objects that this part of my argument "only goes through if its conclusion is assumed: that Pebbles and Rocks should be referred to as two different possible children" (2019: 24). I certainly agree that my argument here depends on the assumption that Pebbles and Rocks are two different possible people. But I see nothing circular or objectionable about assuming this. If we thought that Pebbles and Rocks were literally one and the same

person, after all, there would be no need to appeal to *de dicto* harm in the first place. We would simply say that Wilma's act makes that particular person worse off than that very same particular person would otherwise have been. The entire point of Hare's argument is that in the case of Tess, there really are different people who might end up as the accident victims, and that Tess has an obligation to "the accident victim" in the *de dicto* sense, regardless of which of those particular people end up fitting that description. Hare agrees that Pebbles and Rocks are accurately described as two different, distinct possible children, and he wants to argue that Wilma has a *de dicto* obligation to "her child" in the *de dicto* sense, regardless of which of those particular people ends up fitting that description, because he thinks this follows from what he has to say about what he takes to be the parallel case of Tess. In following Hare in assuming that Pebbles and Rocks are different possible people, then, I'm simply meeting Hare on his own terms and discussing his objection in the only context in which it can sensibly arise.

Finally, in pressing my second objection to Hare, I in part point out that if Hare were correct that the best explanation of why Tess has the obligation she has is because of (a) and (b) noted above, then in deciding which people to take on as patients in my new medical practice, I should make the choice that would be better for "my patients" in the *de dicto* sense, and so select healthier people to have as patients, because in the case of a doctor who does not yet have particular patients, the parallel versions of (a) and (b) are also true: as a doctor I have special obligations to "my patients" and because which particular people will end up as "my patients" depends on what choice I make, I can't help "my patients" in the *de re* sense; I can only help "my patients" in the *de dicto* sense. Assuming the implication that I should therefore choose healthier people to have as my patients is unacceptable, we should agree that Tess's obligations are best explained in some other way, and I go on to suggest that the best explanation is the simplest one: Tess explicitly agreed to take on those obligations when she took the job. Malek again objects to my *reductio* against Hare by saying that there is an important difference between the case of Tess and the case of Wilma (2019: 23). But whether that is so is again irrelevant to my use of the example. The point of my example is to show that Hare's explanation of the source of Tess's obligations has unacceptable implications. This has nothing to do with whether Hare is justified in trying to move from the case of Tess to the case of Wilma. And, again, it is Hare who is claiming that the case of Wilma is relevantly similar to the case of Tess, not me. So Malek's point that there are differences between the two cases again makes things worse for Hare, not for me. Regardless of whether Malek is right to suggest that some of the objections to P1 that I consider in my book are literally

irrelevant to solving the non-identity problem, then, I continue to believe that none of the objections to P1 are strong enough to solve it.

2. CHALLENGING P2: A REPLY TO GARDNER

Molly Gardner's important work on the existence account of harming (2015) was published after I finished writing my book, so I am particularly grateful for the opportunity to engage with it here. In several respects, Gardner's account represents a significant advance over other harm-based approaches to the non-identity problem, but I will focus here on explaining why I think it nonetheless falls short of providing a satisfactory solution.

Gardner formulates the existence account of harming as follows (2019: 34):

> *Harming* (def.): An event, E, harms an individual, S, if and only if E causes a state of affairs that is a harm for S.
>
> *Harm* (def.): A state of affairs, T, is a harm for an individual, S, if and only if
>
> (i) There is an essential component of T that is a condition with respect to which S can be intrinsically better or worse off; and
>
> (ii) If S existed and T had not obtained, then S would be better off with respect to that condition.

As Gardner explains, "the account asks us to compare the world containing both the victim and the alleged harm to a world where the victim exists without the alleged harm; we then check to see whether, in the latter world, the victim is better off" (2019: 34). If the answer is yes, then the event harmed the victim.

The existence account of harming seems to solve the non-identity problem by showing that Wilma's act of conceiving Pebbles really does harm Pebbles. If we compare a world where Pebbles exists and is blind to a world where Pebbles exists and is not blind, we find (I am assuming) that in the latter world, Pebbles is better off. This establishes that the state of affairs in which Pebbles is blind is a harm for Pebbles. Wilma's act caused that state of affairs, and so Wilma's act caused a harm for Pebbles. An event harms someone if it causes a harm for them. And so, according to the existence account, Wilma's act harmed Pebbles. Gardner maintains that this solution satisfies both the robustness requirement and the modesty requirement, and since it seems clear that it also satisfies the independence requirement, this would mean that it successfully solves the problem.

A reason for rejecting a premise of the non-identity argument satisfies the robustness requirement if it is strong enough to warrant rejecting any weakened version of the premise that would still be strong enough to generate the conclusion that Wilma's act is not wrong. I do not think Gardner's solution satisfies this requirement. This is because if we replace P2 with the existence account of harming, we can simply revise P4 in a way that will leave it just as plausible as the original version of P4 and that will still suffice to entail that Wilma's act is not wrong. Even if Gardner's account can help to show that Wilma's act harms Pebbles, that is, it cannot help to show that Wilma's act wrongs Pebbles, and so cannot help to solve the non-identity problem.

The reason for this is as follows. If P2 is true, it's easy to account for the moral relevance of harm. Suppose, for example, that Pebbles was not born blind and that Wilma harmed Pebbles in the P2 sense of harm by poking Pebbles' eyes out shortly after Pebbles was born, thereby making Pebbles worse off than she would have been had Wilma not poked her eyes out. Clearly Pebbles would later have good reason to feel resentment about Wilma's act, to feel disrespected by it, to wish Wilma hadn't done it, to feel entitled to an apology and compensation from Wilma, to think others would have been justified in dissuading Wilma from doing it, and so on. But if the existence account of harm is true, then the mere fact that A has harmed B does not, in itself, seem to be morally relevant. In the case where Wilma conceives Pebbles, for example, the existence account justifies Pebbles in saying something like this to Wilma: "your act harmed me because even though it produced the best result for me that you could possibly have brought about, it nonetheless made things worse for me than they would be in an alternative world that you could not possibly have brought about, one in which I exist and am not blind." To me, at least, it seems clear that this fact about Wilma's act does not give Pebbles a good reason to feel resentment about Wilma's act, to feel disrespected by it, to wish Wilma hadn't done it, to feel entitled to an apology or compensation from Wilma, to think others would have been justified in dissuading Wilma from doing it, or to have any other response we would think justified in a case where Wilma had wronged Pebbles.

If this is correct, then we can simply accept the existence account of harm in place of P2 and replace P4, which says that if an act does not harm anyone, then the act does not wrong anyone, with P4*: if an act does not harm anyone by making them worse off than they would otherwise have been, then the act does not wrong anyone. The result would then be that Wilma's act does harm Pebbles, but not in a way that could make her act wrong. In order for the existence account argument against P2 to satisfy

the robustness requirement, then, it would have to provide a reason for rejecting P4* -- a reason, that is, for thinking that harming a person in the existence sense of harm is the kind of thing that can wrong a person even in cases where the act does not make that person worse off than they would otherwise have been and even in cases where the act leaves them as well off as the person who did the act could possibly have left them. I have difficulty seeing what kind of reason that would be, and Gardner does not provide one. She provides reasons to think Wilma's act harms Pebbles but no explanation of why harming Pebbles in this sense would make Wilma's act wrong.

A solution to the non-identity problem fails to satisfy the modesty requirement if it has implications that are less plausible than the claim that Wilma's act is not wrong. Gardner identifies two somewhat abstract implications of her solution that might be thought to pose problems in this respect and argues that these implications are not that implausible (2019: 37). I have nothing to say about her discussion there, but I do think the existence account of harm has more specific and concrete implications that render her solution unable to satisfy the modesty requirement.

Consider, for example, the case of Mort. Mort is mortal. If he weren't mortal, he would be immortal. Let's assume that would be better for him. In that case, the existence account of harm entails that Mort's parents harmed him by conceiving him. That account, remember, "asks us to compare the world containing both the victim and the alleged harm to a world where the victim exists without the alleged harm; we then check to see whether, in the latter world, the victim is better off." If the answer is yes, then the event harmed the victim. If we compare a world where Mort exists and is mortal to a world where Mort exists and is not mortal, we find that in the latter world, Mort is better off. This establishes that the state of affairs in which Mort is mortal is a harm for Mort. Mort's parents' act of conceiving Mort caused this state of affairs, and so their act caused a harm for Mort. An event harms someone if it causes a harm for them. And so, according to the existence account, Mort's parents harmed Mort by conceiving him. But, of course, all of us are mortal. So the existence account entails that every act of conception harms the person who is conceived. This strikes me as much less plausible than the claim that Wilma's act is not wrong. In addition, if harming Pebbles in the existence sense of harm really were enough to make it wrong for Wilma to conceive Pebbles, as would be necessary for the account to satisfy the robustness requirement, we would have to say not just that every act of conception harms the person who is conceived, but that every act of conception is wrong. That is an even stronger reason to think the solution fails the modesty requirement.

And Mort is just the tip of the iceberg. Candice was born lacking an immunity to cancer. Xavier was born without X-ray vision. Mindy can't read other people's minds. Assuming all of them would be better off if they existed and did not have these conditions, the existence account entails that their parents harmed them (and, if the account is to satisfy the robustness requirement, did something wrong) by conceiving them. And these cases would also generalize to all acts of conception. Perhaps an argument could be given for the claim that lacking vision is a "condition" while lacking X-ray vision (or being mortal) is not. But I find it hard to picture what such an argument would look like and even harder to see why appealing to it would not be ad hoc. If there is a sense in which lacking vision is a "condition" but lacking X-ray vision is something else, why would that distinction be relevant to the question of whether the state of affairs in which the person existed in that state was a harm to that person? Of course, we could point out that it was impossible for their parents to conceive Mort, Candice, Xavier, and Mindy without these conditions. But it was equally impossible for Wilma to conceive Pebbles without Pebbles being blind. And if a defender of the existence account of harm were to bite the bullet and agree that everyone is harmed by being conceived, this would just reinforce the concern that harm in this sense is morally irrelevant.

A second kind of problem arises if we consider the case of Betty. Betty has a slight preference for having a child and knows that if she ever conceives a child, the child will be incurably blind. She goes ahead and conceives Bamm-Bamm and, as a result, Bamm-Bamm is incurably blind. The existence account of harming entails that Betty harms Bamm-Bamm by conceiving him for the same reason it entails that Wilma harms Pebbles by conceiving her. If harming Pebbles in this sense really does wrong Pebbles, as must be the case if the existence account is to satisfy the robustness requirement, then harming Bamm-Bamm in this sense must wrong Bamm-Bamm. But the claim that it's wrong for Betty to conceive a blind child when any child she conceives will be incurably blind strikes me as far less plausible then the claim that it isn't wrong for Wilma to conceive Pebbles. This provides a second kind of reason to think that a solution based on the existence account of harming will fail to satisfy the modesty requirement.

Of course, as Gardner notes toward the end of her discussion, if we accept P2 and reject the existence account of harming, we will have to accept other implications that may themselves seem implausible. Gardner focuses, in particular, on P2's implication that Shooter does not harm Victim in *Preemption* and that Angry Alastair does not harm you when he slaps you in the face (2019: 42). I agree that these implications initially

seem implausible, but they strike me as considerably less implausible than the implications the existence account has in cases like Mort or Betty. So even if we look at things just in terms of initial plausibility, I think P2 already comes out well ahead. In addition, I think the strategies I deploy in the book to show that these implications of P2 are not really as implausible as they may at first seem can survive the challenges that Gardner provides here.

Suppose you are wondering, for example, whether your intuition in the *Trolley* case really is, as I assume, that pulling the switch doesn't harm Philip or is instead, as Gardner suggests, that you have a vanishingly weak reason not to pull it (2019: 43). In that case, you can change the story to one where a tree falls on the switch and activates it and ask if you think the tree's falling on the switch harmed Philip by causing Philip to smash into the second brick wall rather than the first. To me, at least, it remains just as clear that the event that switches the tracks does not harm Philip. And this intuition can't be debunked in the way that Gardner attempts to debunk the intuition in the original version of the case since trees do not act on reasons. P2 has no difficulty accounting for the fact that Victim ended up in the hospital, either, because P2 need not deny that Shooter caused Victim to be in a harmed state. Unlike the existence account of harming, P2 does not maintain that an act's causing a person to be in a harmed state is sufficient to make it the case that the act harms the person. If I push you out of the way of a bus that is about to hit you and you skin your knees as a result, P2 can agree that my act causes the harmed state that you find yourself in while maintaining that my act did not harm you since it did not make you worse off than you would have been had I not pushed you out of the way of the bus. The same would go for the case of Shooter's act. And if, as in the case of Angry Alastair, the police officer is going to cause something bad to happen to you either way, then P2 will say that his decision to save you $3,000 by stealing $5,000 from you rather than causing you $8,000 in damages does not harm you, though, as in the case of my pushing you out of the way of the bus, his act does cause you to be in a harmed state (and, in this case, wrongfully so, since I presumably have the right to push you out of the way of the bus but he has no right to steal the $5,000 from you; P2 entails that the police officer's choice to cause you a smaller loss rather than a larger loss does not harm you, not that his causing you the smaller loss doesn't wrong you). None of this is to say that P2 immediately produces perfectly commonsense results in all cases or that the existence account of harming does not have some advantages of its own. But on the whole, for the reasons I have tried to briefly articulate here, I am not convinced that a satisfactory solution to the non-identity problem can be found by replacing the former account of harm with the latter.

3. CHALLENGING P4: A REPLY TO KUMAR

In order to prove successful, a direct rights-based solution to the non-identity problem must provide a satisfactory answer to two questions: (1) what right is violated by the acts that take place in cases like that of Wilma or the Risky Policy? and (2) who is the bearer of the right that is violated in such cases? Put in terms of the case of Wilma, Rahul Kumar's answer to the second question is that the bearer of the right is not Pebbles herself, or potential or possible Pebbles before she was conceived, but rather "Wilma's child" in the *de dicto* sense, the sense in which Wilma's prospective child simply refers to whatever child Wilma ends up conceiving (2019: 59). And Kumar's answer to the first question is that the right that is violated in non-identity cases is not a right that a particular state of affairs not obtain, like the state of affairs in which Pebbles is blind, but rather a right that one's vital interests be recognized and given appropriate weight by others in their practical deliberations (2019: 56). Wilma, on this account, does not violate the right of Pebbles, in particular, not to be, or to be born, blind, but rather violates the right of "her child" more generally to have its vital interests recognized and given appropriate weight by Wilma when she decides when to conceive. By deciding to conceive blind Pebbles when she could instead have easily conceived sighted Rocks, that is, Wilma did not put sufficient weight on the effect that her choice would have on the vital interests of "her child" in the *de dicto* sense. This is why her act was rights violating and therefore wrong. This strikes me as an intriguing combination of answers to the two questions, one that results in a distinctive kind of rights-based solution that differs in interesting ways from those I discuss in my book. Rather than commenting on its merits relative to other rights-based approaches and to other approaches that also appeal to the *de re*/*de dicto* distinction (like Hare's), though, I will focus here on raising one concern about whether this approach is likely to produce a satisfactory solution to the problem.

The concern is about whether Wilma's child in the *de dicto* sense is the kind of thing that can be a bearer of rights and it can perhaps best be put in the form of a trilemma. Suppose, for example, that Wilma is a US citizen and that prior to her deciding when to conceive, I say "Wilma's child in the *de dicto* sense has a right to a US passport." One thing I might mean by this is that whatever particular child Wilma ends up conceiving, that particular child will have a right to a US passport. If she conceives Pebbles, then Pebbles will have such a right and if she conceives Rocks, then Rocks will have such a right. In this sense, there is nothing problematic about the claim that Wilma's child in the *de dicto* sense can be a bearer of rights. But in this sense, Wilma's conception decision does not seem to violate the

right that Kumar identifies. If she conceives Pebbles, then Pebbles will have a right that Wilma put appropriate weight on Pebbles' vital interests and if she conceives Rocks, then Rocks will have a right that Wilma put appropriate weight on Rocks' vital interests. But there is no reason to think that by conceiving Pebbles, Wilma failed to put appropriate weight on Pebbles' vital interests: Wilma made things as good for Pebbles as she could make them. And so there is no reason to think that Wilma's act was wrong on this first understanding of what it means to say that Wilma's child in the *de dicto* sense has rights.

Now suppose when I say that Wilma's child in the *de dicto* sense has a right to a US passport I mean something quite different. Suppose I mean that "Wilma's child" refers to a more abstract or general subject that is distinct from the particular child who fits that description, namely Pebbles, and that this more abstract or general subject– the position that is filled by some particular child -- is itself something that has a right to a US passport. If this is what I mean, my claim seems clearly false. When "Wilma's child" is understood in this way, as the general class or category that is filled by some particular child or other, it doesn't seem like the kind of thing that could have such a right. And it's not just because it doesn't have a name and can't be photographed. It doesn't seem like the kind of thing that could have any rights. It doesn't have any of the properties that we associate with rights-bearing subjects: it isn't sentient, or rational, or autonomous, it doesn't have preferences, or desires, or interests. What would it mean to say it had a right to free speech? Or privacy, or a fair trial, or an equal consideration of its interests? But it seems that this is what it would have to mean for Wilma's child in the *de dicto* sense to have rights if Wilma's act were to be capable of violating the right to have one's vital interests given the weight they deserve. This is because Wilma's act does not fail to promote the vital interests of "her child" in the first sense, but might in some way be taken to do so in the second sense: her act fills the position "Wilma's child" with a child who has significantly less well-being than the child who would otherwise have filled the position would have had. So it seems that in the sense that "Wilma's child" understood *de dicto* can have rights, Wilma does not violate them and in the sense that Wilma might be understood as violating rights, "Wilma's child" could not have them.

But perhaps there is a third option that can help Kumar overcome this concern. Perhaps he can say that "Wilma's child" has the right in the first *de dicto* sense but that it is a right with respect to the *interests* of "Wilma's child" in the second *de dicto* sense. Annotating one summary he provides of his view (2019: 59; the parenthetical words are his while the bracketed

words are mine), we could put the suggestion like this:

"What the right entitles 'Wilma's child' (understood de dicto) [in the first sense, as referring to whatever particular child Wilma ends up conceiving] to is that Wilma constrain her deliberations so as to make choices concerning her child [in the second sense, referring to something distinct from the particular child she ends up conceiving and more abstract and general: the position that is filled by some particular child] that are, given the information available to her, defensible as consistent with a concern for what will be intrinsically good for her child [in the second sense]".

But this strikes me as objectionably ad hoc. Why would the particular child that Wilma ends up conceiving have a right with respect to how Wilma treats "her child" understood as something distinct from that particular child?

Kumar responds to a somewhat similar worry toward the end of his paper by citing the case of Bob and Sue (2019: 61), but regardless of whether his example helps to diffuse the worry he is concerned with there, it doesn't seem to help with the worry I'm concerned with here. Bob's right that Sue not read his e-mails, Kumar says, is not grounded in particular facts about Bob or Sue but rather in the nature of the relation in which they stand to each other, that of strangers. This might help to show how Pebbles could have some rights against Wilma that are grounded in the fact that she is Wilma's daughter rather than in particular facts about Pebbles herself. But it doesn't help explain how Pebbles, the specific child that Wilma ended up conceiving, could have the kind of right against Wilma that involves having a claim that Wilma treat "her child" in the second *de dicto* sense in a certain way. In order for the case of Bob and Sue to help support that claim, we would have to think not just that Bob has a right against Sue because of the stranger relationship between them, but that Bob has a right that Sue not read the e-mails of "whoever ends up fitting the description 'stranger to Sue who has not given Sue permission to read their e-mails'." And that seems quite implausible. If Sue reads some other stranger's e-mails, that doesn't violate Bob's rights.

There is a further problem with this third option, the one on which "Wilma's child" understood *de dicto* in the first sense has a right that Wilma give appropriate weight to the vital interests of "Wilma's child" understood *de dicto* in the second sense when making her conception decision. The problem is that even if we can make sense of the claim that "Wilma's child" has this right, and even if Wilma's conceiving now rather than two months from now would violate this right, it seems reasonable for Wilma to assume that "her child" would hypothetically waive the right if she could or will

later waive it when she can. This is because if Wilma does not give significant weight to the vital interests of "her child" understood *de dicto* in the second sense, the result will not be worse for "her child" understood *de dicto* in the first sense. Whatever child she conceives today, that is, will clearly have a life worth living and would not exist at all if Wilma delayed conception by two months. Since "her child" in this sense would, or will, waive the right we are assuming it to have, the fact that "her child" has this right, if it is a fact, cannot make Wilma's act wrong.

Kumar provides a brief response to this kind of rights waiver objection. After acknowledging that when Pebbles thinks of herself as the particular person she is, she will not resent Wilma for making the choice she made, he adds: "Which is not to say that Pebbles' waives her right that Wilma not make the choice she did; because the right is one that constrains Wilma's choices with respect to 'her future child', it is not a right that any particular future child who ends up coming into existence is in a position to waive" (2019: 60-61).

Kumar's response here seems to depend on the claim that if a right constrains someone's treatment of X, then only X can waive the right. Why should we accept this claim? One plausible answer to that question arises from two further claims: (1) you can only waive a right if you are the one that has the right, and (2) if a right constrains someone's treatment of X, then only X can have that right. If only X can have a right that constrains people's treatment of X, and if you can waive a right only if you are the one that has the right, then if a right constrains someone's treatment of X, only X can waive that right. And if that's so, then "Wilma's child," understood as whatever particular child Wilma ends up conceiving, can't waive the right that "Wilma's child," understood as something other than that particular child, have its interests given due consideration by Wilma when she decides when to conceive. But Kumar cannot appeal to this argument to support the claim that his response to the rights waiver objection seems to depend on because this argument depends on the claim that if a right constrains someone's treatment of X, then only X can have that right. And if that's true, then "Wilma's child" understood *de dicto* in the first sense, can't have a right that constrains Wilma's treatment of "Wilma's child" understood *de dicto* in the second sense. If, on the other hand, it really is possible for "Wilma's child" understood *de dicto* in the first sense to have a right that constrains her treatment of "Wilma's child" understood *de dicto* in the second sense, then it is not clear why "Wilma's child" understood *de dicto* in the first sense can't waive that right.

While the answers Kumar suggests to the questions of whose rights are violated in non-identity cases and what rights those are may in some

respects improve on those provided by the rights-based solutions I discuss in my book, then, I'm not yet convinced that they suffice to ground a satisfactory solution to the non-identity problem.

4. CHALLENGING P4 AND P5: A REPLY TO WASSERMAN

David Wasserman's commentary on my book is largely sympathetic so it will come as little surprise that my response to his commentary is largely sympathetic, too. Rather than looking for minor points where I might be able to quibble with him, then, I will focus here on the three main arguments for the wrongness of Wilma's conduct that he discusses and add a few comments of my own.

The first kind of argument appeals to the claim that Wilma is the steward of a precious gift, the gift of life, and that by choosing to conceive Pebbles when she could instead have easily conceived Rocks, she violates a moral duty that is imposed by her stewardship of this great gift. This could be either because her decision disrespects the recipient of the gift in a way that gives the recipient a legitimate cause for complaint or because it fails to value the gift itself in a way that the gift should be valued, and so an argument of this sort might be developed as an attack on either P4 or P5 of the non-identity argument. Wasserman considers various versions of this argument and concludes that they fail to identify a wrong-making feature of an act even if the act has it, that there is no reason to think Wilma's act has the feature in the first place, or possibly both.

I have little to add to Wasserman's critique of this first kind of argument. But I would like to note one potentially important respect in which his discussion might be used to strengthen the case for the position in take in my book. Toward the end of the first part of his paper, Wasserman says that potential parents would "arguably display" an objectionable form of "indiscriminateness if they gave the gift of life to someone for whom they believed it would . . . barely be worth living. The steward of a precious gift should only give it to someone for whom it would be of significant value" (2019: 72). As Wasserman correctly notes, this kind of principle can't justify the claim that Wilma's act is wrong because the case of Wilma involves conceiving a child whose life is clearly well worth living. But, if accepted, the principle that Wasserman expresses some sympathy for here would justify the claim that Wilma's act would be wrong in a variant of the case where the doctor told her that if she conceived now her child would have a life that was just barely worth living.

Here is why this is a potentially important result. The non-identity

argument that seems to show that Wilma's act is not wrong in the original version of the story seems just as clearly to show that her act would not be wrong in this variant of the case. In biting the bullet in the original version of the story, then, I seem to be committed to biting the bullet in the variant version. And even if the strategies I use in my book succeed in making the former bullet seem not so difficult to bite (2014: 192-209), my attempt to do something similar for biting the bullet in the latter case (2014: 225-34) might not be as successful. It would therefore be helpful if, as a back-up plan, I could provide reasons to think that biting the bullet in the former case doesn't really require biting it in the latter.

I briefly mention a few such reasons in the book (2014: 226-27), but Wasserman's suggestion here provides a potentially valuable supplement. And the principle it appeals to might be supported by noting that it has significant plausibility in other cases involving the gift of life. It seems plausible, for example, to say that you are not morally obligated to donate a kidney to save the life of a stranger and plausible to say that if you do decide to donate a kidney, you may donate it to save the life of a blind person rather than the life of an otherwise comparable sighted person who would have a significantly higher overall quality of life. But it may well also seem quite plausible to say that while it isn't wrong to decline to donate a kidney at all, it would be wrong to donate a kidney to a stranger whose life is just barely worth living when you could instead at virtually no extra cost donate it to other needy people whose lives would be far, far better for them. In this respect, Wasserman's critique of the first argument against the position I take in my book also offers unexpected additional support for that position.

The second argument that Wasserman considers comes from a recent paper by Robert Noggle (2018). It's grounded in the claim that it's wrong for a person to do an act that deliberately makes it impossible for them to fulfill one of their moral obligations. Wilma has a moral obligation to protect any actual child of hers from blindness. By conceiving now rather than first taking the pill for two months, she makes it the case that she has an actual child that she cannot possibly protect from blindness. Therefore, Wilma's act is wrong. This argument, too, might be developed as an attack on either P4 or P5, depending on whether the conclusion is that Wilma's act wrongs Pebbles or that Wilma's act is simply wrong. Either way, though, as Wasserman correctly points out, the argument is unsuccessful because Wilma's act does not make it impossible for her to fulfill an obligation she would otherwise have had. Prior to deciding when to conceive, it was already impossible for Wilma to protect Pebbles from blindness. Wasserman goes on to note a few other problems with this argument, but I think it's worth adding a few more that he doesn't discuss.

First, if Noggle's argument shows that Wilma's act is wrong, then it also shows that it's wrong for Betty to conceive Bamm-Bamm when she knows that any child she conceives will be blind. In whatever sense Noggle could claim that Wilma is responsible for making it impossible for her to protect Pebbles from blindness, Betty is just as responsible for making it impossible for her to protect Bamm-Bamm from blindness. But, as I noted in the context of Gardner's solution, the implication that it's wrong for Betty to conceive a blind child when she knows that any child she conceives will be blind renders a proposed solution to the non-identity problem unable to satisfy the modesty requirement.

Second, and perhaps more importantly, a parent's obligation to protect their actual children from blindness can be overridden. And one scenario in which the obligation is clearly overridden is that in which a child's losing their vision is necessary for them to continue living. If a child will die without life-saving surgery that would cost them their vision, for example, it would clearly not be wrong for the parents to authorize the surgery. And one very plausible way to account for this is to say that if the child understood the situation, they would consent to the act that would cause their lack of vision or that later, when they do understand the situation, they will give their subsequent consent to that act. But if we can say this about the surgery case, it's not clear why we can't also say it about the case of Wilma: even if Noggle were right to think that Wilma's act makes it impossible for her to fulfill her obligation to protect Pebbles from blindness, this would still not be enough to make the act wrong because Pebbles would, or will, consent to relieve Wilma of this obligation[1].

The third kind of argument that Wasserman considers maintains that Wilma's act is wrong because it involves insensitivity to the harm of blindness. This is the most developed and detailed part of Wasserman's paper, but his considered position on this kind of argument is somewhat less clear. In the introduction, he says of the three kinds of argument he considers "I doubt that Wilma acts wrongly in either of the first two ways and remain uncertain about the third" (2019: 67). In the section of the paper he devotes to the argument, he considers a variety of ways that someone might defend some version of the argument and responds in ways that I, at least, took to be good reasons not to accept any of them. But in the conclusion, when he looks back on that section, he says in part that "I found it plausible to claim that [Wilma] displayed insensitivity to the

1 At the risk of engaging in just one minor quibble with Wasserman: for the same reason, I think Wasserman is mistaken, in the context of his discussion of the first argument, to suggest that the 14-year-old girl's act of conceiving is wrong in Parfit's case because it imposes an unacceptable risk that she would later be unable to fulfill her obligations to the child (Wasserman 2019: 70).

(stipulated) hardship of blindness in choosing to have a blind child without seriously considering what life would be like for such a child" (2019: 82). And he adds that "I also found that such insensitivity would give Pebbles, the blind child she ended up having, grounds for complaint" (2019: 82), which means this third kind of argument would best be understood as aiming at P4. While Wasserman at times seems close to neutral about this kind of argument, at times thoughtfully skeptical, and at other times more sympathetic, I will simply conclude with a few thoughts about this kind of argument to the extent that he seems favorably inclined toward it.

First, I don't think the fact that Wilma makes her choice based on convenience means she is not sufficiently sensitive to the harm of blindness. Suppose you can save one but only one of two drowning children, that it is a bit more convenient for you to save the one who is a bit closer to you, and that the two children are otherwise comparable except for the fact that the one who is a bit closer to you is blind. You can be fully sensitive to the respects in which the blind child's blindness has a negative impact on his well-being while still recognizing that his life is well worth living and while thinking that it isn't morally wrong to let convenience be the tie breaker in determining which life you will save. There may be other differences between life creation and life extension cases, but I don't see why deciding based on convenience in the former case would have to involve this kind of insensitivity if deciding based on convenience in the latter case would not.

Second, suppose that Wilma's act really was based on her being insensitive to the harm of blindness. Suppose, for the sake of the argument, that she didn't give a moment's thought to the kind of life her blind child would end up leading. Even if this is so, this would not show that her act was wrong, only that her motive was objectionable. Suppose, for example, that a man agreed to donate a kidney to his young daughter, who would have died without it, only because he knew his wife would be angry at him if he refused. Clearly his child would later have good reason to complain about his lack of concern for her well-being. But, just as clearly, this does not mean that it was wrong for him to donate the kidney to her.

Finally, even if we thought insensitivity of this sort could suffice to make the act itself wrong, and not just make the motive for doing the act objectionable, this would still not be enough to solve the non-identity problem because, if necessary, we could simply stipulate that Wilma does not act with this kind of insensitivity. Perhaps she had read my book carefully and was convinced by it, and so decided to conceive now only because she could see that despite whatever hardships would ensue for Pebbles, Pebbles would not be harmed or wronged by her doing so. In the

end, then, I don't think any of the arguments Wasserman discusses provide a promising route to solving the non-identity problem by attacking either P4 or P5, and I am, if anything, more skeptical of some of them than he is.

5. CHALLENGING P5: A REPLY TO MULGAN

Tim Mulgan's challenging paper can be summarized in terms of three claims: (1) that I can't account for the wrongness of failing to prevent extinction in his Starship case (2019: 86) without also conceding the wrongness of Wilma's act, (2) that Rule Consequentialism (RC) can account for the wrongness of failing to prevent extinction in his Starship case, and (3) that the way that RC does this enables it to show that Wilma's act is wrong without committing itself to any unacceptably immodest further conclusions. The second claim clearly seems right, but I'm not convinced of the first or the third.

Let's start with the first. There is at least one argument for the wrongness of failing to prevent extinction in the Starship case that strikes me as plausible and that does not entail that Wilma's act is wrong. The argument is grounded in two claims. The first is that if you can prevent the destruction of something of great value at a relatively little cost, it is wrong for you not to do so. The second is that the human species has great value. If both these claims are true, then it follows straightforwardly that it would be wrong not to prevent the extinction of the human species in the Starship case. And since choosing to conceive Pebbles rather than Rocks does not involve Wilma failing to prevent the destruction of something of great value – indeed, by conceiving Pebbles rather than Rocks, she does not fail to prevent the destruction of anything at all – this argument for the wrongness of failing to prevent extinction in the Starship case does not entail that Wilma does something wrong by conceiving Pebbles rather than Rocks.

Mulgan might object that this argument violates my independence requirement. "[J]ust as Boonin rejects ad hoc solutions tailored to avoid the Implausible Conclusion [that Wilma's act is not wrong]," he writes, "we must also reject ad hoc solutions to the Nihilistic Conclusion [that failing to prevent extinction in the Starship case is not wrong]" (2019: 89). But this argument does not violate that requirement. Put in terms of the case of Wilma, the independence requirement says that we can't justify accepting a claim simply because accepting it produces the result that Wilma's act is wrong. When applied to the case of Starship, then, the requirement says that we can't justify accepting a claim simply because accepting it produces the result that failing to prevent extinction in the Starship case is wrong. But each of the two claims the argument depends on seem reasonable

independent of the fact that, when combined, they produce this result. The first seems to best explain the judgments we would make in a wide range of cases where something of great value can be easily saved from destruction, and the second seems to best explain the judgments we would make in a wide range of cases where we are asked to judge the value of the human species against the value of something else.

It's true, of course, that the first claim this argument appeals to is incompatible with P5 of the non-identity argument, since P5 says that if an act wrongs no one then the act is not wrong. But in order to solve the non-identity problem, it's not enough to show that one of the premises of the non-identity argument is false. One must show that the reason it's false can be applied to cases like that of Wilma. And if P5 is false because an act can be wrong without wronging anyone when it fails to easily prevent the destruction of something of great value, this won't help to show that Wilma's act is wrong. Again, Wilma's act does not involve the failure to prevent the destruction of anything, let alone the destruction of something of great value. So I don't think the claim that failing to prevent extinction in the Starship case would be wrong poses a problem for the claim that Wilma's act is not wrong.

What about Rule Consequentialism? RC clearly has no problem accounting for the wrongness of failing to prevent extinction in the Starship case. But I'm not convinced by Mulgan's claim that RC can show that Wilma's act is wrong without committing itself to any unacceptably immodest further conclusions. Either RC permits Wilma to conceive Pebbles or, if it doesn't, it forbids too much to satisfy the modesty requirement. If we set aside doomsday scenarios like the Starship case, Mulgan seems to concede that there's no particular reason to think RC would deem Wilma's act to be wrong. An ideal code capable of being internalized by people can't be too demanding or impersonal, and a code that isn't too demanding or impersonal will permit a variety of sub-optimal choices. Wilma's would plausibly be among them. But, he argues, once the threat of extinction is added to the equation, "RC must abandon its commitment to a liberal ideal code that incorporates a permissive procreative ethic and a presumption of reproductive freedom" (2019: 95). The result, instead, is that "RC will now select a much less permissive ideal code that rejects any presumption of reproductive freedom and also includes a much stronger obligation to do whatever one can to enhance future well-being. When making a procreative choice, the RC ideal agent will not feel free to consult her own desires when this would lead to a sub-optimal result in terms of future well-being" (2019: 96). Clearly such an agent would not feel free to make the choice that Wilma makes.

Mulgan considers the possibility that the ideal code would be more fine-grained than this, leaving people free to make the choice Wilma makes in a world like ours while restricting their freedom in much bleaker scenarios. But he argues that internalizing these more flexible versions of the code across entire generations would produce suboptimal results for three reasons: making the code more complex makes it harder to internalize, we might not be good enough at recognizing when an extinction threat is upon us and thus disastrously fail to shift behavior toward the more restrictive rules when we were supposed to, and the extra benefit in terms of increased freedom for us is not enough to outweigh these potential costs to future generations (2019: 96). As with many attempts to apply consequentialist reasoning to complex concrete questions, it's hard to know just what to say about the claim that the ideal code would have to be as rigid and illiberal as Mulgan claims it would be in this respect. But let's suppose that he's right.

If he is right, there is no reason to limit the code's demandingness to procreative decisions. In the face of a genuine threat of extinction, virtually every kind of decision a person makes could have potentially disastrous consequences for the survival of the species: decisions about what and how much to consume, what and how much to produce, what and how much to learn, who to interact with, how much time to set aside for leisure, and so on. If the factors that Mulgan appeals to really suffice to show that an ideal code would not allow Wilma to "feel free to consult her own desires when this would lead to a sub-optimal result in terms of future well-being" when making procreative decisions, then, they will also suffice to show that an ideal code would not leave her free to consult her own desires when this would lead to a sub-optimal result in terms of future well-being when making consumption decisions, production decisions, educational decisions, socializing decisions, leisure time decisions, and so on. Indeed, it becomes difficult to see how the ideal code could leave people free to consult their own desires when making any decision at all if their choice would lead to a sub-optimal result in terms of future well-being. For all intents and purposes, people would not feel free to do anything just because they wanted to do it, if doing it would produce anything but optimal results for future well-being. The "moderate, liberal alternative" that RC was supposed to offer in place of act-consequentialism seems to have been left far behind.

This result seems worrisome enough even if it's limited to cases like Wilma's, where people have a relatively modest desire to act sub-optimally. What about cases where they would be more significantly burdened by not feeling free to act sub-optimally? Mulgan treats my case of Jane as an

example of this and suggests that RC could appeal to the difference in burdens in the two cases to explain why Jane should feel free to act sub-optimally while Wilma should not. Because of the more significant burden to Jane, he suggests, "Only in extreme circumstances will the RC ideal agent feel obliged to have a child she doesn't want. As Jane's circumstances are not extreme, she does no wrong" (2019: 99). But this suggestion is difficult to square with Mulgan's earlier claim that because of the risk of extinction, RC's ideal code "rejects any presumption of reproductive freedom." If there is no presumption of reproductive freedom, why should the cost to Jane in reproductive freedom count against requiring her to make the optimal procreative choice? And, perhaps more importantly, if Jane can be relied on to recognize whether she is in extreme circumstances, why can't Wilma? If Jane can successfully internalize a code that permits her to act sub-optimally in non-extreme circumstances while requiring her to act optimally in extreme circumstances, why can't Wilma? If Mulgan's argument succeeds in showing that the ideal code would leave Wilma feeling unfree to make the sub-optimal procreative choice that she makes, then, not only does it also show that we should virtually never feel free to act sub-optimally just because we want to, but it will require us to make optimal choices across a broad range of domains even in cases where we have more than just a modest desire not to do so. The implications of RC have become extremely illiberal by this point.

Let's suppose I'm right about this. Is that a problem? It depends on what we're trying to do. If we are trying to arrive at the correct moral theory, it need not be a problem. Perhaps the correct moral theory has a number of implications that will strike most people as extremely implausible. If there is a good enough reason to think that RC is true and a good enough reason to think that RC's ideal code should leave Wilma feeling unfree to conceive Pebbles rather than Rocks, then there is a good reason to think that the correct moral theory tells us to virtually never feel free to act sub-optimally just because we want to or even if we have a significant interest in doing so. I think that implication will strike most people as highly implausible, but if there's a good enough reason to accept it, then we should accept it despite its being so implausible.

But if what we are trying to do is solve the non-identity problem, then I think the extremely illiberal implications of the version of RC that we would have to accept in order to show that Wilma's act is wrong really are a problem. The non-identity problem arises in the first place, after all, only because the conclusion of the non-identity argument strikes so many people as implausible. The claim that Wilma's act isn't wrong is not self-contradictory or empirically false; it just seems implausible. The problem

that stands in need of resolution is that of avoiding being committed to an implausible claim. If the means of avoiding the implausible claim commit us to an even more implausible claim, then, this does not really solve the problem. And if the only way to show that Wilma's act is wrong really does commit us to the claim that we should almost never feel free to act sub-optimally in terms of future well-being just because we want to or even if we have a significant interest in doing so, it seems to me that the means of avoiding the claim that Wilma's act is not wrong commits us to an even more implausible claim. Not only is there a plausible explanation of the wrongness of preventing extinction in the Starship case that does not commit us to saying that Wilma's act is wrong, then, but the extremely illiberal version of RC that seems required to say the acts are wrong in both cases seems considerably less plausible than simply allowing that Wilma's act is not morally wrong.

6. FROM TWO-OPTION CASES TO THREE-OPTION CASES: A REPLY TO ROBERTS

I have focused to this point on the case of Wilma. What about the case of Risky Policy? I think the non-identity problem as it arises in the former case is fundamentally the same as the problem as it arises in the latter case. If my response to the problem works well in the one, then it works well in the other. But Melinda Roberts thinks there's an important difference between the two. Wilma, she argues, is a genuine two-option case, but Risky Policy is really a three-option case. And this means, according to Roberts, that even if my treatment of Wilma proves satisfactory, it can't be successfully extended to the case of Risky Policy. Extending my analysis of the former case to the latter, she argues, generates problems for my defense of P2 of the non-identity argument and for my strategy of trying to make it seem sufficiently plausible that we ultimately accept the conclusion of that argument. The first concern has to do with a claim about harm while the second concern has to do with a claim about wrongness. I will respond to these two concerns in turn. In doing so, I will set aside the question of whether Roberts is right to think the Risky Policy case really should be understood as a three-option case and focus my response on two of the three-option cases she uses in developing her arguments: Medical Case I and the case of Warren and Percy.

Let's begin with a three-option case that Roberts thinks poses a problem for P2, the case she calls Medical Case I (2019: 114). In this case, you are dying, and your doctor has three options: (1) do nothing, in which case you will die, (2) give you medicine A, in which case you will survive and partly

recover with a life clearly worth living but at a significantly reduced level of well-being, or (3) give you medicine B, in which case you will survive and fully recover and so enjoy a significantly higher level of well-being than you'll enjoy if the doctor gives you medicine A. In the case as Roberts describes it, the doctor gives you medicine A. But it turns out that if he had not given you medicine A, he would not have treated you at all.

P2 endorses the counterfactual account of harm: if A's act harms B, then A's act makes B worse off than B would otherwise have been. Roberts thinks Medical Case I poses a problem for this account of harm because it entails that the doctor's act does not harm you and because she thinks this result is implausible (2019: 123). P2 entails that the doctor's act does not harm you because his act does not make you worse off than you would otherwise have been. Because of his act, you survive and partly recover with a life clearly worth living. If he had not done this act, he would instead have done nothing, in which case you would have died. Surviving and making a partial recovery is not worse for you than dying. And so, according to P2, the doctor's act of giving you medicine A does not harm you. The question is whether this implication poses a problem for P2.

I don't think it does for two reasons. First, if we preserve the structure of the case but remove it from the medical context, I doubt people will think the act harms you. Suppose I'm considering doing you no favor, a small favor, or a big favor. I decide to do you a small favor and had I decided not to do you the small favor, I would have done you no favor. Do I harm you by doing you a small favor in this case? I doubt people will think I do. So if P2 entails that I don't harm you in this case, that shouldn't be a problem for P2. The medical case is different from this case only because the doctor is obligated to give you the best available treatment while I am not obligated to do you the biggest available favor. So we can explain away our initial belief that the doctor harms you when he gives you medicine A, assuming we had that belief in the first place, by saying that what we were really responding to is the fact that the doctor wrongs you by making you worse off than he is obligated to make you. And saying that is consistent with accepting P2. His act of giving you medicine A doesn't harm you. It benefits you. But because it doesn't benefit you as much as he is obligated to benefit you, he wrongs you nonetheless by doing it.

But suppose you think it really is a problem for P2 if it can't account for the claim that the doctor harms you by giving you medicine A. In that case, we can vindicate P2 by explaining how it can account for this claim. We can do this by breaking the three-option case down into a pair of two-option cases. Doing so will also help to show why I don't think three-option cases introduce any distinctive problems that can't be successfully

addressed in terms of my treatment of two-option cases.

Here's what I have in mind. First, the doctor chooses between giving you medicine B and not giving you medicine B. From these two options, he chooses not to give you medicine B. That decision makes you worse off than you would otherwise have been. Had he not made this decision, the best outcome would still be available to you. Because he made this decision, the best option is no longer available to you. So P2 says that he harms you by taking that option off the table. And that implication seems correct. Having decided that he's not going to give you medicine B, he now makes a second decision, choosing between giving you medicine A and doing nothing. He decides to give you medicine A. Because of his prior decision, it's his fault that if he hadn't given you medicine A, he would have done nothing. And because it's true that if he hadn't given you medicine A he would have done nothing, it's also true that this second choice – the choice to give you medicine A rather than to give you nothing – does not itself make you worse off than you would otherwise have been. So P2 entails that he does not harm you by giving you medicine A rather than doing nothing. But this implication, too, seems correct. The doctor harms you by taking option (3) off the table but does not harm you by choosing option (2) over option (1) given that option (3) is off the table. That's what P2 says and that seems right. So I don't think three-option cases pose a problem for the account of harm I endorse in the book.

What about my account of wrongness? Roberts presses this problem in the context of the case of Warren and Percy (2019: 107). A sperm and an egg are in a glass dish, moments away from fertilization. Warren has three options: (1) throw away the contents of the dish, in which case Percy will never exist, (2) put a chemical in the dish and leave it there, in which case Percy will exist with some significant impairment but with a life clearly worth living, (3) leave the dish alone, in which case Percy will exist without the impairment and have a considerably better life. Warren chooses the second option. If he hadn't, he would have chosen the first. Roberts argues that the approach to the non-identity problem that leads me to accept the conclusion that Wilma's act is not wrong also commits me to accepting the conclusion that what Warren does isn't wrong. And while, unlike many philosophers, Roberts is sympathetic to my accepting the claim about Wilma, she takes it as clear that we shouldn't accept the parallel claim about Warren. I think that's right, but I also think there's nothing about my treatment of the Wilma case that prevents me from saying so. And so, with respect to wrongfulness as with respect to harmfulness, I don't think three-option cases introduce any problems that can't be addressed in terms of the way I address two-option cases like that of Wilma.

We can again see this by analyzing the three-option case in terms of a pair of two-option cases and then looking at the question of whether Warren's behavior is wrong. First, Warren chooses between leaving the dish alone and not leaving the dish alone. He decides not to leave the dish alone. Next, having decided that he's going to tamper with the dish, he chooses between two ways of tampering with it: tossing out its contents, or pouring the chemical into it. From these two options, he chooses to pour the chemical into the dish. If he hadn't poured the chemical into the dish, he would have thrown away the contents of the dish. As a result of Warren's choices, Percy comes to exist with the significant impairment.

What does my treatment of the two-option case of Wilma commit me to saying about the three-option case of Warren? Consider first his decision to tamper with the dish rather than to leave it alone. Assuming that tampering with the dish doesn't wrong anyone else, such as the sperm or egg donor or those providing the technical support, my treatment of the case of Wilma doesn't yield a determinate answer to the question of whether this choice is wrong. If choosing to tamper with the dish leads him to toss out the contents, then Percy will never exist and so no one will have been wronged or harmed by his decision. If, on the other hand, his decision to tamper with the dish leads him to pour the chemical into it, then it will turn out that his decision to tamper with the dish did harm and wrong someone by making Percy worse off than he would otherwise have been. This implication is a bit complicated, but not in a way that should make it seem problematic. You can't decide whether someone acts wrongly by planting a bomb in an abandoned building they own, for example, without knowing whether they subsequently lure someone into the building before the bomb explodes. So I don't think my analysis generates any problems in assessing the wrongness of Warren's first choice. That choice is wrong if it leads him to subsequently put the chemical in the dish and not wrong if it doesn't.

Now let's focus on the second choice. It is now the case that Warren will either pour the chemical into the dish or throw away the contents of the dish. He chooses to pour the chemical into the dish rather than throw away the contents of the dish. My analysis commits me to saying that he does not do anything wrong by making this second choice. That result also seems right to me. And as long as we keep in mind that he has already taken the option of leaving the dish alone off the table, I think it will seem right to you, too. Suppose, for example, that the only way he could prevent the contents of the dish from falling into a garbage can was by pouring the chemical into the dish. Choosing to pour the chemical into the dish in that case would not seem wrong. If it's his own decision to tamper with the dish rather than contingent circumstances that limit him to these two options,

that doesn't seem to change that fact that choosing to pour the chemical rather than to throw away the contents doesn't by itself wrong or harm anyone. So is Warren's behavior in this case wrong? Yes. His decision to tamper with the dish ends up harming Percy. If his decision to tamper with the dish had instead led him to throw out its contents, then it would not have been wrong. These results seem right to me and are entailed by breaking the case down into a pair of two-option cases and then treating the two-option cases in the way that I treat the two-option case of Wilma in my book. It still seems to me, then, that if my treatment of the two-option case of Wilma in the book is satisfactory, it can be successfully extended to the case of Risky Policy even if Roberts is correct that that case must be understood as a three-option case.

BIBLIOGRAPHY

Boonin, D., 2014: *The Non-Identity Problem and the Ethics of Future People*, Oxford: Oxford University Press.

Gardner, M., 2015: "A Harm-Based Solution to the Non-Identity Problem. *Ergo* 2: 427-44.

—2019: "David Boonin and the Non-Identity Argument: Rejecting the Second Premise", *Law, Ethics and Philosophy* 7: 29-47.

Hare, C., 2007: "Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?", *Ethics* 117: 498-523.

Kumar, R., 2019: "Rights, Wronging, and the Snares of Non-Identity", *Law, Ethics and Philosophy* 7: 48-62.

Malek, J., 2019: "The Non-Identity Non-Problem: Constructing Continuity of Identity among Possible Persons", *Law, Ethics and Philosophy* 7: 15-28.

Mulgan, T., 2019: "How Rule Consequentialism Avoids Boonin's Implausible Conclusion", *Law, Ethics and Philosophy* 7: 84-101.

Noggle, Robert. 2018. "Impossible Obligations and the Non-Identity Problem", *Philosophical Studies*.

Roberts, M., 2019: "Boonin on the Nonidentity Problem: Too Narrow on the Facts? Too Narrow on the Theory?", *Law, Ethics and Philosophy* 7: 102-126.

Wasserman, D., 2019: "Chasing the Elusive Wrongdoing Intuition", *Law, Ethics and Philosophy* 7: 63-83.

Symposium on Antony Duff's *The Realm of Criminal Law*

DOI: 10.31009/LEAP.2019.V7.09

The Public Realms: On How to Think About Public Wrongs*

MICHELLE MADDEN DEMPSEY
Villanova University

ABSTRACT

This paper presents a way of thinking about public wrongs, offered as a friendly amendment to the view articulated by Antony Duff in his recent book, *The Realm of Criminal Law* (2018). The view defended here is distinct from Duff's in three ways. First, it denies a sharp distinction between public and private accountability, preferring instead to view matters as scalar. Second, it asks whether conduct is wrong, before asking whether it is a *public* wrong. Third, it views public accountability as taking place within various public realms, rather than conceiving of the public realm as primarily associated with the state. Along the way, this paper presents a brief sketch of public and private accountability and illustrates both the structural and practical implications of these two different ways of thinking about public wrongs.

Keywords: Public wrongs, Private wrongs, Criminalization, Accountability

1. INTRODUCTION

A central theme of Antony Duff's scholarship is that criminal law is best understood as a communicative enterprise through which the members of a political community call one another to account for (alleged) wrongdoing (See especially, [Duff 1986, Duff 2001, Duff 2007, Duff 2018])[1]. By declaring

* This paper was presented (via skype) at conference on Antony Duff's *The Realm of Criminal Law,* hosted by University of Pompeu Fabra, Barcelona, on 8 February 2019. My sincere thanks to the organizers, especially Gustavo Beade, José Luis Martí, and Roberto Gargarella - and to the participants, especially Antony Duff, whose model of scholarly excellence has long challenged and inspired me.

1 For ease in what follows, I will omit the word "alleged." It is true, of course, that in calling to account, we might get things wrong in terms of identifying the correct norms that should guide our conduct, supposing that it is our business to call someone to account for their wrongdoing, or simply be incorrect about the facts of what happened (e.g. where the person called to account is not the person who committed the wrong).

certain kinds of conduct to be criminal wrongs, enforcing such criminal laws, and imposing criminal punishments, criminalization partly constitutes the civil order of that political community (Duff 2018: 182, 203). So far, I agree entirely. Indeed, I might even go further to say that criminalization partly constitutes the character of that political community (Dempsey 2009): a point to which we will return in what follows.

This way of viewing criminalization makes salient the fact that a political community does at least three things through its criminal law. First, it reminds members of the community about the norms that (it claims) should guide their conduct. Second, it clarifies the scope and substance of those norms in cases of disagreement or confusion. Third (and importantly to what follows), it declares that certain wrongs are the "criminal law's business" (Duff 2018: 79; quoting Wolfenden 1957: para. 61). That is, by criminalizing a particular wrong (Φ-ing), the political community thereby claims that it is justified in declaring Φ-ing to be a criminal wrong, calling Φ-ers to account through its criminal process, and subjecting Φ-ers to criminal punishment.

In his most recent book, *The Realm of Criminal Law*, Duff puts the third point as follows: criminalization "make[s] clear that these are the polity's *public* norms, as distinct from whatever private norms citizens or groups of citizens might espouse outside the civic realm" (2018: 207). At this point, I find myself in disagreement with Duff, on three related fronts. First, while I agree that some aspects of living in community are more-or-less public and some are more-or-less private, I worry that Duff draws too sharp a distinction between the public and private spheres. On my view, the distinction between the public and private is not a matter of drawing lines between a polity's norms (norms which Duff calls *public* in the quote above) and the norms of citizens or groups of citizens (which Duff calls *private* in the quote above). Rather, the distinction between public and private is more a matter of a sliding scale that admits many gray areas.[2] Second, when we think about calling one another to account for wrongdoing, we should not seek to draw a distinction between the nature of the norms themselves (either public norms or private norms, as Duff

2 On my reading of Duff, he seems to prefer drawing rather sharp lines between the public and private realms, rather than thinking in terms of degree. I believe that Duff's account can accommodate degrees, as his discussion of "pure crimes" versus "pure torts" at p. 294 suggests. So perhaps my friendly amendment here is simply a plea for further clarification from Duff on this point if I have misread him. My thanks to an anonymous reviewer for pushing me on this point. In any event, however, on my view, a tort should be understood as a public wrong, insofar as the public justifiably responds to the wrong by providing a forum in which the victim can seek redress. It is an indirect public response to the wrong, but a response (by the public) nonetheless.

would say). Rather, we should understand the distinction between public and private in this context as one that turns on the *justifiability of calling one another to account for wrongdoing in a more-or-less public or private ways*. Third, in order to determine whether a more-or-less public or private calling to account is justifiable when it comes to any particular conduct, we must first understand whether the conduct is wrong and, if so, why.

In what follows, I will flesh out this way of thinking about public wrongs in friendly contrast to Duff's account. As we will see, my way of thinking about public wrongs fits well within Duff's broader project of understanding the criminal law as a matter of calling wrongdoers to account, and it emphasizes the ways in which a community's practices of calling to account, in part, constitute the character of that community. The essay proceeds in three steps. First, I briefly explain more-or-less private and public ways of calling wrongdoers to account, and highlight two important considerations that inform whether any such calling-to-account is justifiable: (1) whether and how the wrongdoer will otherwise be called to account; and (2) how practices of accountability are constitutive of the character of the person or group that does (or fails to do) the calling.[3] Second, I explain why we must first understand whether the conduct in question is wrong and, if so, why it is wrong, before asking whether a wrong is more-or-less public or private. Third, I apply this way of thinking about public wrongs to the project of identifying various public realms and attempt to illustrate the practical difference between my account and Duff's.

2. CALLING TO ACCOUNT: IN PUBLIC AND PRIVATE

Some forms of calling others to account are more private, whereas others are more public. Three hallmarks of a public calling to account are: (1) that it is done by someone in a *representative capacity*; (2) who is acting *in the interests of* a larger group of people; and (3) that it is done in a way that is generally *observable to others*.[4] A calling to account that is done by the direct victim of the wrong, on her own behalf, in a way that is not observable to others is what I will call *comprehensively* private. If this kind of response is the only justifiable response to the wrong, then it is a comprehensively private wrong. If the direct victim calls out the wrongdoer as a representative

3 By "practices of accountability," I mean to refer to whether and how a person or group either does, or does not, call wrongdoers to account.

4 Compare Duff (2018: 24, 113, 118, 229) on calling to account as a representative (my first hallmark); acting "in the interest" of the polity (my second hallmark); and (163-164; 293) on public and private spaces (my third hallmark).

of a larger group or in the interests of others (say, a rape victim calling her rapist to account by and for women generally), the calling to account is thereby somewhat more public. So, too, if she calls him to account on social media, it is all the more public. On the other end of the spectrum, a calling to account is *comprehensively* public when it is done by someone who both *represents* and is *acting in the interests of* the entire political community ("The People," "The State," "The Commonwealth," etc.), and it is done in a way that is transparently *observable* to the entire political community. If this kind of response to a wrong is justifiable, then it is a comprehensively public wrong. In between these extremes, there are more-or-less private and more-or-less public ways of calling others to account – and, thus, more-or-less private and more-or-less public wrongs.

This way of thinking about public wrongs differs from Duff's because it focuses our attention on the justification of various possible responses to the wrong, rather than the nature of the norms at issue. For Duff, when we think about calling one another to account for wrongdoing, the distinction between public and private wrongs turns on the supposed public or private nature of the norms at issue (Duff 2018: 207), or whether the wrong was committed in a public or private 'normative space' (Duff 2018: 165). On my account, the distinction between public and private (scalar though it is), turns on the justifiability of different ways the wrongdoer might be called to account. Specifically, it depends on who can justifiably call him to account; in whose interests can he justifiably be called to account; and what degree of observability is justifiable in calling him to account. As noted, where the only justifiable response to a wrong is one that comes from the direct victim of the wrong, acting solely in her own interests, and is done in a way that is not observable others (aside from the wrongdoer and the victim), then that wrong is properly understood as a comprehensive private wrong.[5] However, if a more public response (e.g. calling out the wrongdoer on Twitter) is justifiable, then the wrong is properly understood as a more public wrong. If it would be justifiable for the state, acting through its official representatives (police, prosecutors, judges, jailors) to call the wrongdoer to account in a way that is transparently observable to the entire political community, then the wrong is properly understood as a comprehensively public wrong. On this account, the same wrongful conduct can be (more-or-less) private, or (more-or-less) public, depending on the justifiability of the various ways in which the wrongdoer might be called to account for that wrong.

5 I'm concerned here only with wrongs that have a direct victim, other than the wrongdoer. I agree with Duff that if no one can justifiably call me to account for my wrong, it is a private wrong (Duff 2018: 164).

The justifiability of calling a wrongdoer to account turns on many considerations – far too many to unpack here. (Indeed, giving a full accounting of every salient consideration strikes me as impossible.) However, here are two important considerations in deciding whether any given calling to account is justified. First, we should consider whether and how the wrongdoer will otherwise be called to account. If the calling to account under consideration is the only way the wrongdoer will be held to account (if it is a "last resort" for accountability), then that fact weighs in favor of its justifiability. Second, we should consider whether and how the particular calling to account under consideration is part of a broader pattern of accountability practices, and whether/how those accountability practices are constitutive of the character of the person or group that does (or fails to do) the calling. The first consideration speaks to the idea of criminal law as a "last resort": if there are robust and effective means for calling wrongdoers to account *outside* the criminal law, then that fact relieves some of the pressure on the criminal law to do the work of calling them to account. All things being equal, the fact that there are other ways to call a wrongdoer to account makes the case for criminalization less justifiable.[6] The second consideration speaks to fact that political communities shape their characters through their accountability practices. By calling to account (or failing to call to account) those who commit certain kinds of wrongs (say, racist hate crimes), a political community constitutes its character as more or less racist, ceteris paribus. So too, as I have argued elsewhere, a community that uses its criminal law to target particular kinds of domestic violence can reconstitute its own character in valuable and virtuous ways by becoming less patriarchal (Dempsey 2009). All things being equal, if a political community uses its criminal law to cultivate valuable character traits through calling wrongdoers to account for certain kinds of wrongdoing, this makes the case for criminalization more justifiable than it otherwise would be.

3. PUBLIC WRONGS: WHICH QUESTION SHOULD WE ASK FIRST?

When it comes to identifying public wrongs, Duff insists that we must address the question, "Is it a public matter?" before we consider, "Is it wrong?" As he explains, "An account of crimes as public wrongs must begin not with wrongs, but with the public" (Duff 2018: 165). While I agree there

6 I revisit this point below in section 3.

is a methodological priority regarding the order in which we ask those two questions, I think it goes the other way. In this section, I attempt to explain why.[7]

Duff's worry seems to be that if we start by asking "Is it wrong?" we risk falling into a "Moorean legal moralism" (2018: 100). In that dystopian world, legislators will sit around debating the moral permissibility of my cancelling lunch plans with a friend for no good reason, or the white lie you told your cousin to get out of helping her move house. Clearly, these wrongs are *not* the business of the criminal law – and so Duff wants to ensure that they never even become candidates for criminalization. I agree. It *would* be disconcerting if legislators or other legal officials took seriously the question of whether such wrongs should be criminalized. Normally, they should deliberate as if these matters were not the criminal law's business (Dempsey 2011; Duff 2018: 93-95). Still, I doubt we need to prioritize the question, "Is it a public matter?" before the question, "Is it wrong?" in order to accomplish that end.

Indeed, in order to answer the question "Is it a public matter?" - framed as a *normative* question (that is, as a question of whether something *should be* regarded as a public matter) - we need to know what "it" is, so that we can consider the reasons that might weigh in favor (or against) the public taking an interest in it. That is, we need to ask the kinds of questions (with some modification) that Duff flags up on page 297:

> We must ask whether it is important, and how important it is, to mark and respond to the wrong as a wrong, by seeking to call its perpetrator to account; whether it is important, and how important it is, that the calling to account be collective—that the perpetrator be called to answer by and to [some group of people, rather than only the]… direct victim; whether it is important, and how important it is, to seek an authoritative, categorical judgment of the alleged wrongdoer's guilt or innocence; and whether it is important, and how important it is, to provide for punishment to be imposed on the wrongdoer … (Duff 2018: 297).

I think these are (many of) the right kinds of questions to be asking when

7 My point here is not to claim that Duff denies that public wrongs are wrongs. Certainly, he is clear that public wrongs are wrongs. My point is methodological: when we think about what counts as a public wrong, should we start by thinking about whether it is public, or whether it is wrong? As the quote above illustrates, Duff's methodological priority is to begin with the question of whether it is public. My methodological priority is to begin with whether it is wrong. Duff and I both agree that that when it comes to criminalization, we should not go directly from the premise "X is wrong," to the conclusion, "X is a candidate for criminalization." That is, neither one of us fall into what he calls "Moorean legal moralism." My thanks to an anonymous reviewer for pushing me to clarify this point.

we think about what properly falls within the criminal law's business – and what wrongs count as public wrongs more generally. But, of course, we can't even begin to ask these questions unless we know something about whether the conduct is wrong and, if so, why it is wrong. If we adopt Duff's methodological priorities, the answer to each question would simply be, "I don't know – I'm not even sure it's a wrong yet. I'm still focused on whether it's a public matter."

My worry about Duff's ordering of the questions is two-fold. First, if we start by asking, "Is it a public matter?" we risk missing those things that *should* be understood as public matters, but currently are viewed as private matters. For example, we might skip past a wrong involving sexual cajoling, which results in nonetheless consensual sex, because we properly judge that the conduct is not a public matter suitable for criminalization.[8] If we skip past that wrong, we miss the opportunity to recognize the ways in which *other* public responses – albeit non-criminal responses - might be appropriate. While sexual cajoling may not be a proper matter for criminalization, perhaps it *is* a matter for state intervention in terms of public education aimed at shifting social norms. Or, perhaps it is a matter for a less comprehensive public response: that is, a response not by the state acting on behalf of the political community as a whole, but a response by some group acting within a more limited understanding of the public sphere. For example, perhaps it is a matter for school discipline, suspension from an athletic team, condemnation by one's peer group, or public accountability on social media, etc.

This first part of my worry is illustrated in Duff's characterization of the wrong of betraying a friend by "selling salacious gossip [about my friend] to the popular press" (Duff 2018: 99). On Duff's view, one way to make sense of criminalizing such a wrong is "to begin from a totalitarian, rather than a liberal, conception of the polity and its res publica: a conception according to which the polity has a proper interest in every aspect of its citizens' lives" (ibid). For Duff, once law-makers begin "weighing the reasons for and against criminalizing the wrong I do to my friend, [they] already go wrong by seeing the relationship of which that wrong is part as any of [the polity's] business—as something to which it has reason to attend." According to Duff, this approach would be "totalitarian" because some things (here, relations between friends) are simply not the polity's business (ibid). In making this point, Duff quotes my work, from a chapter in which he and I engaged in a similar debate regarding "the criminal law's business." The quote from Duff (quoting me), goes as follows: "…[they]

8 See, for example, Kim Ferzan's illuminating discussion of the Aziz Ansari case (Ferzan 2019).

already go wrong by seeing the relationship of which that wrong is part as any of [the polity's] business—as something to which it has reason to attend, as part of 'a full philosophical account of the reasons that actually do bear on the justifiability' of its decisions." Duff (2018: 99), quoting (Dempsey 2011: 256).) Yet, the thesis of my chapter which Duff quotes was that there is conceptual space between the reasons to which law-makers have reason to attend (those reasons that are "the law-makers' business") and "a full philosophical account of the reasons that actually do bear on the justifiability" of their decisions. The former set of reasons are limited by political considerations (e.g. the res publica of the polity for which they are creating law). These kinds of reasons are the business of law-makers. On this, I take it that Duff and I agree. The latter set of reasons are limited only by their merit (whether they are genuine reasons – that is, value-ladened facts). These kinds of reasons are the business of philosophers, who should aim to take on board "a full philosophical account of the reasons that actually do bear on the justifiability" of law-maker's decisions.

While Duff's thin account of criminalization does not make any substantive claim regarding whether selling gossip about a friend is a private or public wrong, his discussion on this point is reminiscent of the view taken of the marital relationship, which shielded domestic violence from public accountability until recently. According to Duff, domestic violence is now properly regarded as public (despite being previously regarded as private), due to "the gradual recognition of [the institution of marriage] as 'public' [which] marked a development in our collective understanding of that institution" (2018: 165). On my view, this account gets matters in reverse order. It is not that a development in our collective understanding of marriage made domestic violence a public wrong. It was always a public wrong for a husband to beat his wife, we just did not do anything about it until recently.[9] That is to say, it has always been justifiable for domestic batterers to be called to account in ways that bear the hallmarks of public accountability outlined above: to call them to account *by a representative* of the political community; to do so *in the interests of* the political community (not just the victim); and to do so in a way that was *observable* by the entire political community. The problem was simply that, in practice, we either did not call batterers to account at all, or we left domestic abuse to more private forms of accountability.

The second part of my worry is that in order to answer Duff's true

9 I place the husband in the role of the abuser because that is overwhelmingly the form that domestic violence has taken historically, and husbands were granted legal immunity under the defense of chastisement. I do not, of course, deny that women/wives can, do, and have committed domestic violence. For a disaggregation of distinct forms of domestic violence, see (Dempsey 2009: ch. 6).

question – the normative question of what should count as a public matter – we need to start by asking what values can be realized by treating a wrong as a public matter.[10] That is, we must ask what reasons we have for calling for accountability in a more public, rather than merely private, way. This is a question we simply cannot begin to ask unless we first ask what, if anything, makes the conduct wrong. For example, if part of what makes some instances of domestic violence wrongful is the fact that they tend to promote and sustain patriarchal structural inequalities in society generally, this tells us something important about the reasons we might have for treating those wrongs as public rather than private. That is, it tells us something about why a comprehensive public response to those wrongs is more likely to be justifiable. Specifically, it tells us two things. First, if some instances of domestic violence tend to promote and sustain patriarchal structural inequalities in society generally, that tells us that the character of our political community is marred by patriarchal structural inequalities which establish the social context in which such wrongs are committed (Dempsey 2009: chs. 6-7). If our society is marred by this character defect, we have reason to cultivate a more valuable (non-patriarchal) character. Second, given that calling certain kinds of wrongdoing to account can reconstitute our character in valuable ways, criminalizing this kind of wrongdoing can realize such value (what I have called *constitutive value*), and thus lend weight to the justification of criminalization (ibid). My point here is simply that if we do not start by asking whether the conduct is wrongful, and if so, why it is wrongful, we cannot begin to see the reasons that might weigh in favor (or against) responding to it in a private or public manner. We need to know why it is wrong, so that we can know whether its wrongfulness is grounded (in part) in a character defect of our community. For, if that is the case, it lends strength to the case for a public response to the wrong - perhaps one rising to the level of criminalization.

4. THE VARIOUS PUBLIC REALMS: ALTERNATIVE SITES OF PUBLIC ACCOUNTABILITY (ASPAs)

This section identifies various public realms, and illustrates the relevance of two considerations in thinking about the justifiability of calling wrongdoers to account: (1) whether and how the wrongdoer will otherwise be called to account; and (2) how practices of accountability are constitutive of the character of the person or group that does (or fails to do) the calling.

10 On Duff's search for a normative theory of criminalization, see (Duff 2018: 3-6).

In one of the most instructive parts of *The Realm of Criminal Law*, Duff draws an analogy between the criminal law and self-regulating professional organizations, such as the medical profession, legal profession, academia, etc. (Duff 2018: 80-91). In his discussion, Duff draws attention to the fact that there are a vast range of public realms. He observes, almost sheepishly, that he is using the label "'public'…in an admittedly slightly unusual way," by conceptualizing accountability to one's professional organization (or even one's golf club) as a form of "public" accountability (Duff 2018: 83). I suppose Duff might be correct in thinking it slightly unusual to call these normative contexts "public" – but doing so is tremendously illuminating. Let's call these non-criminal, yet still public, normative contexts, "Alternative Sites of Public Accountability" or ASPAs for short.

There is much to be gained by keeping in mind the various ASPAs out there - and the ways that ASPAs can respond to something as a "public wrong," even when those public responses never come anywhere near the comprehensively public response of criminalization. For one thing, it reminds us that the case for criminalization of any given public wrong is strengthened, or weakened, depending on how well we do in holding perpetrators accountable in other public realms. Duff's discussion of queue-jumping illustrates this point well (Duff 2018: 280-282). Part of the case for not criminalizing queue-jumping, even if it is a public wrong, depends on whether queue-jumpers will be held to some other kind of public accountability for their wrongs (even if the "public" here merely consists of "fellow queue members" – and the ASPA at issue is merely the site of public queue itself). As Duff correctly observes, whenever "we can, without criminalizing, achieve the good for whose sake we would criminalize," the case in favor of criminalization is correspondingly diminished, because "criminalization would be unnecessary" (ibid:282). Conversely, when we fail to make use of ASPAs to hold each other accountable for our public wrongs, the case for using the criminal law gets stronger. The criminal law, as it were, then becomes the "last resort" for achieving the goods that can be realized through public accountability.

Consider the implications of this thought in the context of the #MeToo movement and related phenomena, such as universities holding students accountable for sexual misconduct, sports teams suspending players for sexual assault, or perpetrators being held accountable on social media. In all of these ways, political communities are creating new and multiple sites for holding wrongdoers publicly accountable for sexual wrongdoing. If we are concerned about over-criminalization, we should support these ASPAs - since, if they are effective in holding wrongdoers to public account, they take some of the pressure off the criminal law and thereby reduce the

strength of the case in favor of criminalization. Moreover, these ASPAs are valuable insofar as they allow space for a political community to debate and revise its understanding of sexual wrongdoing. Through ASPAs, members of the political community can stand against a broad range of sexual wrongdoing - including wrongs that have not been understood as matters of public concern until recently – such as unwanted sex, sexual cajoling, etc. [11] In doing so, ASPAs play a critical role in debating and shaping our public norms.

Some key implications of my way of thinking about public wrongs can fruitfully be compared to Duff's in the following hypothetical. Meet Dave: in addition to being a dentist and member of the American Dental Association (ADA), Dave is also a member of the National Organization of Men Against Sexism (NOMAS), The American Numismatic Society (ANS, dedicated to coin-collecting), and the Philadelphia Mummers (a group consisting of almost exclusively white men, who put on a parade in Philadelphia each New Year's Day). Dave's membership in both NOMAS and the Mummers is surprising, since the Mummer's are widely regarded as being rather sexist (Tynes 2016).[12] Still, these four organizations are four of the various ASPAs that might justifiably call Dave to account in the event that he commits a wrong.

As it happens, Dave often tells sexist jokes at work to his dental hygienist. On Duff's account, Dave's sexist jokes "come to our attention already marked as public" in the context of the dental profession, "since they will come to our attention as we think about matters that are already understood as belonging in [this] realm" (Duff 2018: 79). Specifically, we know that Dave's conduct falls within the realm of the ADA's business because the conduct occurs at work, in his dental practice. We should thus agree with the ADA when it claims to be justified in calling Dave to account for his sexist jokes under its Principles of Ethics and Code of Professional Conduct for "undermining professional relationships among team members" (ADA, 2018: 8). Whether or not Dave's conduct is wrongful, it is the business of the ADA, because it is happening in the context of a dental practice by one of its members. So, too, on Duff's account, Dave's sexist jokes "come to our attention already marked as public" in the context of Dave's membership in NOMAS, since sexist behavior, whether it is wrong or not, is "already understood as belonging in [NOMAS's] realm" (Duff: 79). Whether it's fighting sexism or being sexist, if conduct concerns sexism, it falls within the realm of NOMAS's business. We do not need to bother ourselves,

11 For discussion, see (Ferzan 2019); (West 2002).

12 Note that I am not claiming that the Mummers actually are sexist. I am merely drawing on their widely acknowledged reputation to get this hypothetical off the ground.

according to Duff, with asking whether Dave's conduct is wrong – we just need to know that it involves sexism and that he is a member of a group that makes sexism its business.

Conversely, on Duff's account, Dave's sexist jokes are simply not the business of ANS or the Mummers: neither group has any standing to call Dave to account for this conduct, because Dave's conduct does not concern his behavior *qua* coin-collector or *qua* Mummer. If Dave sold fraudulent coins to fellow collectors, then his conduct would "come to our attention already marked as public" as to the ANS. So, too, if Dave showed up late to parade practices, then his conduct would "come to our attention already marked as public" as to the Mummers. But when it comes to Dave tellingsexist jokes to his dental hygienist, neither the ANS nor the Mummers have any reason to call him to account, according to Duff.

On my account, we do not start with the assumption that conduct "comes to our attention already marked as public" in some context. Rather, in order to know if the conduct is a public wrong (in any sense), we first need to know whether the conduct is wrong, and why, so that we can consider which, if any, calling to account (public or private) might be justifiable. With Dave's workplace sexist jokes, it is easy enough to see why the conduct is wrong and why there would be a strong case justifying the ADA and NOMAS for calling him to account for it. In their distinctive ways, each can make salient different aspects of the wrong by calling Dave to account. The ADA can make salient the aspects of the wrong that relate to dental workplace professionalism, while NOMAS can make salient the aspects of the wrong that relate to sexism. In making these aspects of Dave's wrongdoing salient, each group realizes values central to their missions.

Duff would perhaps rather phrase this last point in terms of the *res publica* of the organizations (2018: 81), rather than their missions. In contrast, I would not want to trade on the positive connotation of res publica to motivate the assumption that an organization's realization of its mission is valuable.[13] Rather, on my account, it remains an open question. In the *Dave* hypothetical, I assume that the missions of the ADA and NOMAS are valuable: that is, I assume it is a good-making feature of the world that organizations like these should exist and further the goals

13 The ordinary meaning of *res publica* is associated with the "common *good*" (emphasis added) (OED 2010). Duff's use of the term is stripped of value – mostly. At some points, he characterizes the *res publica* of an organization as the "the *proper* concerns and aims, of the practice in question" (2010: 81). At other points, even on the same page, Duff makes use of a more analytic conception of the *res publica* of an organization - as "what, given its aims, falls within its scope" – irrespective of whether those aims are valuable or disvaluable (ibid).

grounded in their missions. If these were not valuable organizations, then furthering goals grounded in their missions would not realize any value. (Compare the Nazi party calling one of its members, Nick, to account for denying Aryan supremacy. In calling Nick to account for denying a tenet central to their mission, the Nazis do not realize value, despite the fact that calling Nick to account would further a goal grounded in their mission. This is so because the mission of the Nazi party is not itself valuable.[14])

All that said, however, there is so far very little that distinguishes my account from Duff's when it comes to calling Dave to account. If we grant the assumption that the ADA and NOMAS are valuable organizations, and that calling Dave to account realizes values grounded in their missions, then my account, too, has a story to tell as to why Dave's conduct falls within the ADA's and NOMAS's business. Structurally, its different than Duff's account, but it gets us to the same place so far.

Yet, when we ask whether the ANS or the Mummers might justifiably call Dave to account, my way of thinking about public wrongs raises considerations that are off-limits according to Duff. Imagine that the ANS and the Mummers have become aware of Dave's sexist workplace jokes and are considering whether to ask him to account for his conduct or risk being disinvited to the annual ANS meeting or barred from the annual Mummer's parade. On Duff's account, the ANS and Mummers are already out-of-bounds for even entertaining the possibility of such a response. According to my way of thinking about public wrongs, however, we would need more information before evaluating the justifiability of the ANS or the Mummers calling Dave to account. Again, we would start by asking whether Dave's conduct was even wrongful in the first place – for, if it was not wrongful, then it would not be justifiable for anyone to call him to account for it.[15] Second, we would need to know whether Dave has been held accountable, either publicly or privately, for his wrongful conduct – for, if he has not been held accountable in any other context, this fact would add to the case in favor of the ANS or Mummers calling Dave to account for his wrongdoing. (On its own, it would not suffice to justify either group calling him to account, but it would lend some strength to the case in favor of doing so.) Third, we would need to consider what values might be realized by the ANS and/or the Mummers calling Dave to account.

Recall that when Dave is called to account by the ADA and NOMAS,

14 To be clear, on my view, it is not justifiable to call someone to account for conduct that is not morally wrong, simply because it goes against a (purported, but not genuine) value that the group or polity has embraced. Thank you to an anonymous reviewer for requesting clarification on this point.

15 By "wrongful" here, I mean that the conduct is at least pro tanto wrong.

their calling to him account realizes values that are central to their missions. The same cannot be said when it comes to the ANS and the Mummers. For Duff, that's the end of the story. When he is called to account by the ANS and the Mummers, Dave can simply (and, according to Duff, correctly) reply: "I don't have to account for my conduct to you. It's none of your business."

On my view, things are a bit more complicated, depending on the values that might be realized by calling Dave to account. For example, if the Mummers calling Dave to account were part of a broader good-faith attempt to transform their sexist ways (that is, to reconstitute their sexist character), then *ceteris paribus* they would have a stronger case in favor of calling him to account than the ANS would have.[16] Indeed, it might even be justifiable for the Mummers to call Dave to account for his sexist jokes, despite the fact that their mission concerns performing in an annual parade, not fighting sexism. When it comes to the ANS, things seem rather less complicated. Their mission concerns coin-collecting, not dental practice professionalism and not sexism. Given the terms of the hypothetical so far, we have no reason to think that the ANS's calling Dave to account will realize any significant value. Thus, it seems unlikely that it would be justifiable for the ANS to call Dave to account. We can thus conclude, in agreement with Duff, that Dave's conduct is none of their business. Indeed, it seems that the ANS would not be justified in calling Dave to account even if we take on board a full philosophical account of the reasons that bear on the justifiability of their doing so.[17] My point is simply that we cannot reach that conclusion without first identifying whether and why Dave's conduct is wrong in the first place, and considering the values that the ANS might realize in calling Dave to account.[18]

16 Again, I'm not offering any factual claims about the Mummer's actual character.

17 Depending on the extent of Dave's misconduct, the ANS might be justified in disinviting Dave to the annual meeting – but only for preventative reasons, to avoid the possibility of his sexually harassing fellow attendees. They would not, however, be justified in calling him to account for his past wrongs, or imposing retributive punishment.

18 We can imagine a modified hypothetical where we include the fact that the ANS is concerned to diversify its membership and encourage more women to participate in their annual meeting – and stipulate that doing so would be valuable. In that case, the goal of creating a welcoming and respectful space at their annual meeting might lend some justification to their calling Dave to account, on pain of being disinvited to the meeting if he fails to explain himself, apologize if unjustified, etc. My point is simply that these facts matter, because they bear on whether the ANS would realize value in calling Dave to account, such that their doing so might be justifiable, under some variant of the hypothetical.

5. CONCLUSION

This paper has presented a way of thinking about public wrongs, with the goal of offering an alternative to the view articulated by Antony Duff, most recently in his book, *The Realm of Criminal Law* (2018). Of course, the view defended here owes much to Duff's work as it has developed over the decades. Most importantly, it is grounded in Duff's important insight that criminal law is best understood as a communicative enterprise through which the members of a political community call one another to account for wrongdoing. Moreover, it draws on the Duffian theme that a community's accountability practices, in part, make that community what it is. While we have slightly different ways of putting that last point, I take us to be in substantive agreement.

We part ways, however, when it comes to how to think about public wrongs. The view defended here is distinct from Duff's in three ways. First, it denies a sharp distinction between public and private accountability, preferring instead to view matters as a matter of degree. Second, it requires that we ask whether conduct is wrong, before asking whether it is a *public* wrong. Third, it views public accountability as taking place within various public realms, rather than conceiving of the public realm as primarily associated with the state. In explaining and defending this alternative way of thinking about public wrongs, this paper has presented a brief sketch of public and private accountability, and illustrated both the structural and practical implications of Duff's account of public wrongs, as compared to my account.

In the end, I suspect Duff will view this paper as yet another failed attempt to defend a modified version of "Moorean legal moralism" (Duff 2018: 93-96). Perhaps, if accused, I should plead guilty this time. For Duff's alternative, I fear, prevents us – as philosophers - from taking on board a full philosophical account of the reasons that actually do bear on the justifiability of calling someone to account for their wrongdoing. While perhaps legal officials ought not to take all of these considerations on board in their deliberations regarding what to criminalize - we, as philosophers, ought to keep them in view. The account of public wrongs offered here strikes me as one way to keep the full range of considerations in view, as we continue to work toward an illuminating account of justifiable criminalization.

BIBLIOGRAPHY

American Dental Association. 2018. *Principles of Ethics and Code of Professional Conduct*, Chicago: American Dental Association.

Dempsey, M. M., 2009: *Prosecuting Domestic Violence: A Philosophical Analysis*, Oxford: Oxford University Press.

—2011: "Public Wrongs and the 'Criminal Law's Business': When Victims Won't Share", in *Crime, Punishment, and Responsibility*, eds. R. Cruft, M.H. Kramer, and M.R. Reiff, 254-272, Oxford: Oxford University Press.

Duff, R.A., 1986: *Trials and Punishments*, Cambridge, Cambridge University Press.

—2001: *Punishment, Communication and Community,* Oxford, OUP.

—2007: *Answering for Crime: Responsibility and Liability in Criminal Law,* Oxford, Hart Publishing.

—2018: *The Realm of Criminal Law*, Oxford, Oxford University Press.

Ferzan, K. K., 2019: "Consent and Coercion", *Arizona State Law Journal* 50: 951-1007.

Oxford English Dictionary, 2010: (3rd ed.) online < https://www-oed-com> (accessed 18 May 2019).

Tynes, T., 2016: "This Famous Philadelphia Tradition Has An Extensive History Of Racism And Sexism", *Huffington Post*, <https://www.huffpost.com/entry/philadelphia-mummers-parade_n_568c091ce4b0b958f65d18bd> .

West, R., 2002: "The Harms of Consensual Sex", in ed. A. Soble, *The Philosophy of Sex: Contemporary Readings* (4th ed.), 317-322, New York: Rowman & Littlefield.

Wolfenden, J., 1957: *Report of the Committee on Homosexual Offences and Prostitution*, Cmnd. 247, London: H.M.S.O.

DOI: 10.31009/LEAP.2019.V7.10

Professional Ethics and Criminal Law*

KIMBERLEY BROWNLEE
University of Warwick

ABSTRACT

According to Antony Duff, certain realms of wrongdoing are properly the business of only the members of a particular scheme of practice – those members are the relevant public to whom other members are accountable. The polity is one such scheme of practice and the criminal law addresses the wrongs that are properly the polity's business. The criminal law is essentially a code of civic ethics for the members of a polity to hold one another to account. To elucidate these ideas, Duff draws an analogy between the criminal law and other codes of ethics, notably, codes of professional ethics, which also concern the conduct of a particular set of people acting within a particular system of practice. This article explores the merits of Duff's analogy between professional codes of ethics and the criminal law. First, the article details some problematic features of professional ethics codes, such as their disinterest in the moral merits of the profession they regulate and their tendency not only to be rigid and clunky, but also to shield members from full accountability. Second, the paper shows that many of the flaws in ethics codes are to be found unavoidably in the criminal law as well, which casts doubt on the claim that the criminal law is well-placed to identify and address wrongs that are genuinely morally wrong independent of and prior to their criminalization.

Keywords: Antony Duff, codes of ethics, criminal law, professional ethics, public wrongs, wrong

* I thank Simon Gansinger and two anonymous referees for valuable written feedback on this paper. For a helpful discussion, I thank both Antony Duff and the participants of the Symposium on Antony Duff's *Realm of Criminal Law* at Pompeu Fabra University, which I attended via Skype on 8 February 2019.

1. INTRODUCTION

In Chapter 2 of *The Realm of Criminal Law*, Antony Duff argues that *legal moralism* need not be moralistic in a pejorative sense. Indeed, we can develop a plausible version of legal moralism that doesn't seem to recommend a state-controlled witch-hunt whereby legislators seek out moral wrongdoing so that it can be criminalized and punished. To develop this plausible version of legal moralism, we must focus on *public wrongs*, but we must start with the public part of *public wrongs*, rather than the wrong part of *public wrongs*. That is to say, we must start with the *public realm* – the realm in which public wrongs are identified – and determine what is the business of the polity. As Duff puts it:

> "Rather than starting with the whole realm of moral wrongs as our canvas, and then asking which of them should be the criminal law's business, we must think about the criminal law's business, as a distinctive kind of legal institution; and to do that we must think about the polity's business (its *res publica*), since the criminal law's proper business must be to contribute, in some distinctive way, to the polity's business (79)." [1]

To flesh out these ideas of the *public realm*, the *polity's business*, and *public wrongs*, Duff draws an analogy between codes of professional ethics, which are concerned with misconduct within a particular system of practice, and the criminal law, which is concerned with wrongdoing that falls within the distinctive practice of civic life – the life of a polity. The criminal law can be seen (with some qualifications) as a code of civic ethics, Duff says.

My aim here is to examine the merits of Duff's analogy between codes of professional ethics and the criminal law. I highlight some problematic features of professional ethics codes, beginning with the fact that they often reflect insiders' lack of self-scrutiny about the moral merits of their profession along with the fact that codes are rigid and clunky, and tend to shield members from full accountability. Some of these problems are non-contingent: they arise from the very nature of such codes. Other problems depend on how the code is designed. I then show that several of the non-contingent faults of ethics codes are to be found, unavoidably, in the criminal law too. In doing so, I raise some general concerns about how Duff carves up the moral landscape and instrumentalizes moral reasoning.[2]

1 The page references are taken from the online version of Duff (2018).
2 For a related discussion, see Zaibert (2018) 118ff.

2. CODES IGNORE THE MORAL STATUS OF THE PROFESSION

2.1 Duff's Account of How a Profession Might Develop a Code of Ethics

Duff states that a code of professional ethics will define the kinds of wrongdoing which amount to professional misconduct and which the people bound by that code ought to avoid. Also, typically though not necessarily, a code will provide a process through which charges of misconduct can be adjudicated, and suitable sanctions imposed if those charges are proved (80).

Duff observes that, logically speaking, if not chronologically speaking, when the members of a profession are working out their governing structures and practices, their first item of business cannot be to design a code of ethics for themselves. To know what falls within the scope of a code of *professional* ethics, the drafters must first know what goals and activities fall within the *res publica* – the public realm – of their profession.

According to Duff, his distinction between the *public realm* and the *non-public* or *private realm* is normative: whether something is a public matter turns on whether it is something that *properly* concerns the relevant 'public', i.e. whether the community or profession in question has a *legitimate* interest in the matter. A doctor's medical malpractice is a public matter within her profession. Her cheating at golf, however, is a private matter as it relates to her profession. In other words, her cheating is not the business of the members of her profession. By contrast, her cheating at golf is a public matter for the members of her golf club, and her medical malpractice is not: it's a private matter in relation to her fellow golfers. A profession or community's *res publica* '…expresses its members' normative self-understanding of what they are (of what they are about) as doctors or as academics', etc. (161).

Once the members of a profession know what the business of their profession is, they can discuss the proper content and scope of their ethics code: 'they will not begin with the whole realm of moral wrongdoing, and of all possible moral wrongdoers, as their potential canvas.' This is because '…they will already know that they are concerned (only) with professional misconduct—that is, with wrongdoing committed within the realm of the profession and its practice…' and by the members of that profession (82).

2.2 The Moral Problems with Codes of Ethics

One problem with professional ethics codes is that members of a profession

are insiders who, almost by definition in virtue of their membership, view their profession as broadly morally acceptable. As such, they're not best-placed to reflect on whether its activities and goals are *genuinely* morally acceptable. Although they probably have far greater knowledge of, and expertise in, many aspects of their work than outsiders do, nevertheless they are also more likely to rationalize their activities and goals than outsiders are, especially when outsiders have nothing at stake.

When members of a profession move to the second step in Duff's logical order of events – namely, devising their code of ethics – they signal that they take their profession's goals and activities to be morally acceptable and amenable to a code of genuinely ethical conduct. But, many professions that take this second step are morally suspect, such as professions that administer capital punishment, impose solitary confinement in prison, use extreme interrogation techniques in investigations, and engage in profit-hungry high finance that can devastate the global economy.[3] A more extreme example is the SS in Nazi Germany, whose members were bound by an honor code implemented by an independent system of SS courts. According to Herlinde Pauer-Studer,

> "Himmler set the standards for the SS legal system by cultivating the idea of the SS as a specific elite bound by a code of 'honour and decency'. Members of the SS lost their status and could be excluded from the SS if they violated this code. Loyalty and honour were interpreted as unconditional obedience and 'ethical acceptance' of the orders of the authorities. The SS ethos rested on voluntary 'ethical fulfilment' of duties (Pauer-Studer, 2012, 375)." [4]

The codes of professional ethics governing such activities are morally suspect because the activities they inform are morally suspect and, in some cases, rotten to the core.

If a key activity within a profession is rotten, or if all the specific goals and activities that define that profession are rotten, then the so-called 'misconduct' that the ethics code identifies will include conduct that

3 A quick online search yielded the Goldman Sachs' Code of Conduct and Business Ethics: https://www.goldmansachs.com/investor-relations/corporate-governance/corporate-governance-documents/revise-code-of-conduct.pdf, as well as the American Correctional Association's Code of Ethics: http://www.aca.org/ACA_Prod_IMIS/ACA_Member/About_Us/Code_of_Ethics/ACA_Member/AboutUs/Code_of_Ethics.aspx?hkey=61577ed2-c0c3-4529-bc01-36a248f79eba, and the Code of Business Ethics and Conduct for the private military company Academi, formerly Blackwater: https://www.academi.com/pages/about-us/code-of-conduct.

4 I thank Simon Gansinger for highlighting this example.

people ought to engage in. The conduct is only 'wrong' according to the norms, expectations, and demands of the indefensible profession. It is not actually morally wrong. And certainly, the conduct is not wrong *prior* to its inclusion in the code. Following inclusion in the code, certain kinds of otherwise good conduct might become wrong – possibly – for complicated reasons. For instance, if a prison guard releases an innocent person from a brutal prison without authorization, she hasn't made his life any easier; he's now on the run. But, that is another matter.

This point is particularly relevant to Duff's analogy between ethics codes and the criminal law because Duff maintains, with some nuances, that the conduct that is to be included in the criminal code is wrong *prior* to its inclusion in the code. To maintain this view, Duff must hold that there is a disanalogy between the criminal law and many professions' codes of ethics on the grounds that, unlike morally suspect professions, the criminal law is clearly aiming at something good in his view, namely, determining how to live together; and, hence, the conduct it deems to be wrong actually is morally wrong *prior* to its inclusion in the criminal law.

The idea that the criminal law aims at something good (broadly speaking) is plausible enough. But, this won't expose a disanalogy between it and morally suspect professions because, at a *very* basic level, any profession that we recognize as a profession can be seen as aiming at something good, such as getting people the drugs they need, or appropriately managing the movement of money, or dealing with serious wrongdoing. So, the analogy between codes of ethics and the criminal law still holds, but not in a way that makes either come out smelling like roses. Both can be viewed against a backdrop of broadly acceptable aims, while nonetheless being morally objectionable in their specific goals and activities.

In many professions, members' lack of self-scrutiny – lack of *capacity* for clear-eyed self-scrutiny – about the morality of their *res publica* stems partly, but only partly, from the fact that their profession long predates its code of ethics. (Duff finesses this by speaking of the logical 'if not the chronological' order of drafters' activities. He also finesses this in his discussion of the criminal law by focusing on the efforts of Founding Parents who are enjoying a *constitutional moment* in which the problem of drafting a code *ex post* for a public that already exists doesn't arise.) It's unsurprising that insiders who have long absorbed the norms and values radiating from their profession's established community will operate on the assumption that their profession is morally acceptable. But, their conviction gives us no reason to accept their testimony about their profession's moral merits.

Another problem with professional codes of ethics is that they beef up a

profession's *appearance* of ethical credibility. As Duff himself notes rather unconcernedly, people 'who seek public respectability for their practice might seek to define themselves as a profession partly by creating a code of ethics' (80). According to Bob Brecher, such ethical window-dressing is deeply problematic:

> "It seems that the more immoral a society becomes, the greater its need to dress up its political and ideological priorities in ethical clothes…[E]thics so often – and so easily – serves as (attempted) justification of what is clearly morally wrong, whatever the reasons of those who make use of such 'ethical cover' (Brecher 2010, 351)." [5]

The 'ethical cover' given by a code can make morally objectionable professions that urgently need reconsideration less vulnerable to such review. Ethical cover can make external criticism of a profession, as well as self-scrutiny, harder to mount since members can point to the Code and say that they are 'properly' managing the moral pitfalls of their work.

As this suggests, the terms 'proper' and 'legitimate' – which Duff uses to specify the appropriate focus of the criminal law – are slippery characters. Codes of ethics reflect a formal rather than a substantive view of what *properly* concerns a profession's members. If we adopt a formal notion of *propriety*, then some types of activity become not the 'proper' concern of a certain professional simply because they fall within a different profession's jurisdiction. Prison guards shouldn't be deciding without authorization to release prisoners who they believe are innocent or have been confined too long. Parole boards, judges, and senior management teams should decide upon such matters. Similarly, prison guards shouldn't be executing people on death row the night before their scheduled execution. But, the impropriety of a prison guard deciding to release a person or executing a person on death row says nothing about the moral acceptability of either of these activities as such. The formal notion of *propriety* (as something pertaining to jurisdiction) leaves unanswered whether *anyone at all* should be imprisoning people or executing people.

Moreover, taking a formal rather than a substantive view of a profession's proper concerns invites an impoverished vision of *professionalism*, one that focuses on appearances rather than substance, and adherence to the codified rules rather than respect for genuine moral concerns. Codes often highlight norms of professionalism that pertain to members' treatment of each other: the norms that regulate profession members' romantic relations with each other, their relations with senior management, their

5 Related to this, when codes of ethics do more than window-dressing, they seem to be plasters on enterprises that are inherently immoral.

conduct in general, and so on. But, most of those norms are content-independent and transportable from one morally suspect enterprise to another.

Even when a profession is, on balance, morally defensible, its code can still prohibit conduct which members actually should engage in, and indeed should engage in for the sake of the *res publica* of their profession. For instance, the American Correctional Association Code of Ethics states that:

> "Members shall make public criticism of their colleagues or their agencies only when warranted, verifiable, and constructive." [6]

That trinity of requirements - warranted, verifiable, and constructive – is too demanding. It requires not just that a member's proposed public criticism of a colleague be warranted and verifiable, but also that it be *constructive*. And, presumably, criticism that is (perceived as) non-constructive or destructive could include criticism that could harm the profession's public image. As this example shows, codes of ethics can subject members to professional pressures to be unprofessional in a richer sense by directing them to act unethically. Even the codes of morally acceptable professions do this when they are clunky, inflexible, and general when morality requires specificity, subtlety, and context-sensitivity.

Codes can also help powerful people to ensure that the finger points away from them. Brecher observes that:

> "…to criticize a whistle-blower such as Margaret Haywood for breaching her ethical responsibilities to her (immediate) patients at once diverts attention from the political realities governing the situation in a National Health Service (NHS) hospital where elderly patients are routinely mistreated and focuses instead on the 'ethical' shortcomings of the person who blew the whistle (Brecher 2010, 352)."

Haywood's conduct aligned with a richer, more interesting and credible notion of *professionalism* that requires an understanding of the moral reasons for a profession to exist at all (if there are good moral reasons for it to exist), and honors those reasons when specific professional orders clash with them. Slavishly following the norms of a profession is not the best way to honor the responsibilities entrusted to members. Indeed, being slavishly deferential to the rules can dull the mind, the judgment, and the energy to challenge injustices when necessary, making it all the harder for members

6 American Correctional Association Code of Ethics: http://www.aca.org/ACA_Prod_IMIS/ACA_Member/About_Us/Code_of_Ethics/ACA_Member/AboutUs/Code_of_Ethics.aspx?hkey=61577ed2-c0c3-4529-bc01-36a248f79eba

to honor their real moral responsibilities.[7]

These comments on both the self-effacing nature of professions (i.e. members' inability to engage in clear-eyed self-scrutiny about their goals and activities) and the implications of that self-effacingness for how codes regulate conduct, are significant because Duff uses the boundaries of a given public realm to determine who has and who lacks the standing to object to what members of that public do. If members cannot reliably track when their fellows commit genuine moral wrongs, then they are unable to hold each other properly to account. And yet, in Duff's view, outsiders lack the standing to hold them to account, with the consequence that often they may escape real accountability. Let me elaborate.

3. DUFF'S CONTEXT-SPECIFIC PUBLIC / PRIVATE DISTINCTION IGNORES SOME PARTIES

Duff notes that his distinction between the 'public' realm and the 'private' realm is not only normative (in a formal sense), but also context-specific. As noted above:

> "…a doctor's treatment of her patient counts as 'public' in the context of her medical life, but as 'private' in the context of her membership of a golf club. It is something in which her fellow doctors properly take an interest, and something for which (if it involves mistreatment) she may be called to account by her fellow doctors or by a professional disciplinary committee: but fellow members of her golf club have no such interest in it, and no standing to call her to account…Her unsporting behaviour on the golf course, by contrast, is a public matter in the context of the golf club, but private in the context of her medical life… (83)."

This distinction seems plausible at first glance, but proves to be problematic.

7 In their paper on distortions of normativity, Pauer-Studer and Velleman (2011, 351-2) recount the case of Johann Paul Kremer, a physician, who served at Auschwitz during the time when its gas chambers were in operation: 'Kremer's environment was one in which aspects of mass-murder were given medical labels such as "triage", "quarantine", and "therapy". Kremer fell in with this professional mind-set, viewing camp inmates as patients, as anatomical specimens, as subjects of scientific research. Because he never killed anyone with his own hands, he could think of himself as upholding the standards of his profession. And he could be complacent about his professional rectitude because of having proven himself a "fanatic of truth" in opposition to the racial science of the day. With his attention drawn to the "medical" features of his activities, and to the medical ethic that they seemed to call for, he was applying rules of normative salience that did not reveal the true moral weight of what he was doing.' I thank Simon Gansinger for highlighting this example.

Suppose the person who spots the doctor cheating at golf is a golf fan and an audience member at the tournament, not a player or referee. Is that audience member a member of the relevant community? Suppose that the viewer who spots the doctor cheating is a walker out with his son, who pauses to watch the game. Suppose the walker's young son spots the cheating too, and looks to his father to see what he will do as a model for how to behave. Is the father, or the son, a member of the relevant community?

Duff would embrace the idea, I presume, that the doctor's conduct both in golf and in medicine is not the father's business, as he's not part of the relevant public. But, this is a worrying idea since, drawing on Brecher again, 'professional codes of ethics [can] serve to protect professionals from the [wider] public rather than vice versa.' Indeed, Brecher objects more strongly to codes of ethics than I have here. He maintains that they are insidious (2010, 354). First, in his view, the very notion of *professional ethics* makes a false distinction between our moral responsibilities as people and our putative moral responsibilities as professionals. Second, codes instrumentalize moral thinking. Their function is both to hold members accountable and to protect members from moral accountability. Many parties learn to play the system to appear ethically compliant while allowing profit or power to govern their decision-making.

I am not as skeptical as Brecher is about codes of ethics since, if well-designed, they can help us at least a bit to navigate treacherous moral waters that we have not encountered before, that we cannot have clear moral intuitions about, and that we have no good analogues for in our ordinary lives. But, if codes of ethics are as morally objectionable as Brecher argues, or even as I have argued, what does this mean for Duff's analogy between them and the criminal law?

Duff would no doubt acknowledge that professional codes of ethics have many the problems I have identified. If many of these problems are indeed unavoidable and, hence, make professional codes 'insidious' as Brecher puts it, then we must ask whether codes offer a good model for how to think about the criminal law. Duff is well aware that similar problems beset many criminal legal systems. The question is whether such problems are an inescapable feature of criminal codes as such and, if they are, whether that makes the criminal law morally suspect too. I will return to this theme below.

4. CAN THERE BE GOOD NAZI FATHERS OR RESPECTABLE WIFE-ABUSING ETHICISTS?

Another issue worth highlighting pertains to *who* is doing the professional (or public) wrong. Duff says:

> "I should make clear what I am and am not arguing: in particular, I am not arguing that fellow members of a person's profession should just ignore the serious moral wrongs that he commits if they do not bear on his professional conduct. Suppose that I come to know about some egregious moral wrong committed by an academic colleague—perhaps a wrong that is also criminal: he has left his partner in a cruelly damaging way, or committed a fraud against the polity (evading his taxes) or a corporation. Suppose too that the wrong has no direct impact on or implication for his academic conduct: he still behaves properly towards students and colleagues; he is still honest in his teaching and research. Suppose too that I and my university rightly do not share the extensive definition of academic misconduct that includes some version of 'gross moral turpitude'. In that case, I have no reason to respond to, to take any notice of, his conduct *as an academic colleague*: no reason to call him to account for it in our professional dealings, or to table it for discussion at a departmental meeting, or to report it to the university authorities as a disciplinary matter. It is not a matter of academic conduct, and therefore cannot be a case of academic misconduct: which is to say that it is not my business as his academic colleague (86)."

There are two sentences in this passage that I wish to flag:

> "Suppose…that the [academic's] wrong has no direct impact on or implication for his academic conduct: he still behaves properly towards students and colleagues; he is still honest in his teaching and research."

And:

> "Suppose too that I and my university rightly do not share the extensive definition of academic misconduct that includes some version of 'gross moral turpitude'. In that case, I have no reason to respond to, to take any notice of, his conduct *as an academic colleague*."

In terms of morality, can we divorce our spheres of activity so sharply? Could a Nazi prison guard actually be a good and loving father to his children despite his day job? Could an academic actually be a good, honest, respectable, and responsible colleague and teacher even if he's cruel to his wife or corrupt in his activities as a citizen? This is an empirical question, and we won't settle it here. It is at least possible, and sometimes likely, that

a person's behavior in one sphere will bleed into his behavior in other spheres, that his private attitudes about women or the law will creep into his attitudes about his professional duties, colleagues, and students through his manner, moods, opinions, and conduct.

Moreover, even if his behaviors in different spheres remain distinct and wholly detached from each other, his personal behaviors could become known to colleagues and, more importantly, to students who operate within an institutional setting that encourages them to revere their professors. The exposure of his non-professional wrongs cannot but influence students' impressions of him and his scholarship as well as, possibly, their sense of how they should conduct themselves. If non-professional wrongs do bleed into his professional life, then, contrary to Duff's second statement highlighted above, colleagues do have a reason, in principle, to take notice of his non-professional conduct *as academic colleagues*. Although a professional tribunal is not the right body to hold him to account for non-professional wrongs that fall outside its jurisdiction, colleagues have reasons to attend to the repercussions of those wrongs for his standing and behavior within the profession.

A further, interesting thought to consider is whether the relevance of an academic's conduct outside of his profession depends partly on the kind of academic he is. Duff says, referring to a stylized version of himself as an academic at the University of Stirling:

> "…if I mistreat my children, this is not a violation of my academic duties, nor something for which my university has the right to hold me to account…(155)." [8]

My question is: Does it matter if it's a moral philosopher, rather than a chemist, who is cruel to his partner, guilty of corruption, or mistreating his children, especially if that philosopher defends positions that would not countenance those modes of conduct? (As a discipline, Philosophy has real-world examples of leading philosophers who defend objectivist, egalitarian views of ethics and justice, but whose alleged behavior toward women is reprehensible.) Or, does it matter if the chemist has actively championed women within academia and yet, outside of academia, is cruel to his wife? Similarly, does it matter if it's a pediatrician rather than a gerontologist who is a neglectful or vicious parent to her children?

The political equivalent might be Donald Trump, whose alleged treatment of women in personal relationships would seem to be despicable

8 Duff notes in a footnote that, 'There is, of course, plenty of room for argument about what does, or does not, fall within the res publica of a profession…but on any account some aspects of its members' lives fall outside it.' (155).

and, more importantly for present purposes, *criminal*. We might say that those behaviors are not the business of the Congress who might hold him to account, but those behaviors have bled into his political activities – in the form of hush money at least – and they inform people's opinions of the qualities and behaviors necessary to conduct oneself appropriately in the executive office that he holds.

In response to my above questions, I think it does matter *who* does the wrong. Contrary to Duff's view, it seems that, in at least the cases of the moral philosopher, the chemist who champions women in academia, the pediatrician, and indeed President Trump, their specific interpersonal wrongs are matters of professional misconduct, as well as non-professional wrongdoing, and hence it is their colleagues' business as *colleagues* as well as *tout court*.[9]

With these observations about codes and members on the table, let's turn to Duff's vision of the criminal law as a code of civic ethics.

5. A CODE OF CIVIC ETHICS

5.1 The Business of the Polity

In Duff's view, the same logical order that applies to the development of a profession's code of ethics applies to the development of the criminal law. Logically speaking, the first task is to specify what the business of the polity is and, with it, a conception of the *public realm*. Then,

> "[the 'public' wrongs] '…that are, in principle, the business of the criminal law…will not need to be distinguished, within the whole realm of wrongs, from 'private' wrongs: they will come to our attention already marked as public wrongs, since they will come to our attention as we think about matters that are already understood as belonging in the public realm (79)."

Duff leaves questions about the polity's business unexplored in Chapter 2 where he examines legal moralism, professions, and their codes. In Chapter 4, he observes that we can list the formal elements which a specification of a polity's civil order needs to include:

9 These observations are distinct from, but related to Duff's observation that '… professions [such] as medicine, policing, banking, and other financial services (not to mention the law), are not self-contained enterprises that directly affect only their professional members: they deal with 'the public'…' (89) Also to be fair, Duff's point is that the dividing line between the wrongs that professions or the criminal law should concern themselves with is elusive.

- the central institutions of governance and law,
- the substantive ends and values towards which the polity's life is to be oriented, and
- the rights or freedoms that are to be respected both by the state (as the polity's institutional order) and by its members (158).

Further clauses and amendments then give more substantive detail to these ambitions and how they will be pursued. Duff notes that '…civil order (as I am using the idea) depends on law and the rule of law, it extends beyond the realm of legal institutions; it encompasses the informal, extra-legal practices and expectations by which our social lives as citizens are structured.' (159)

Specifying the polity's business more precisely is not straightforward since questions about that business can take at least two forms: What is the central purpose, aim, or goal of the polity? or Which matters, issues, or enterprises properly concern the polity? The two questions are related of course, and Duff references both, albeit in disjunctive terms: 'A political community, like any kind of community, must be defined (it must define itself) in terms of some shared *goals* or *activities* that constitute its business.' (155, emphasis added). Nonetheless, the two questions give different images of what we should seek to specify.

If we focus on which matters, issues, and enterprises properly concern the polity, then a host of possible answers present themselves, including security, resource allocation, regulation of membership, use of coercive force, punishment, and determination of the boundaries of jurisdiction. But, these answers are all formal answers. As such, they highlight again the slipperiness of the notion of *propriety* and what *properly* interests the polity, and they fail to guide us in our efforts to determine what to do. Being told that punishment and security are 'proper' concerns of the polity tells us nothing about how to approach them.

If, by contrast, we focus on what is the central purpose or goal of the polity, then we move closer to something substantive, but face a host of candidate goals, some of which are incompatible with each other and some of which are impoverished. One comparatively rich answer, which the Secretary General of the OECD got members to endorse, is that the end of government is the wellbeing of the people. This is impressive ambition. A similar, though underspecified answer comes in Article 1 of the Polish constitution, which Duff quotes: the 'Republic of Poland shall be the common good of all its citizens' (154). As Duff notes, the content of

'common good' could be fleshed out in terms of a wide variety of values (not all of which equate to wellbeing).

An alternative vision of the polity presents it in night-watchman terms, which, to my mind, offers an impoverished vision of a *res publica*, in the same way that a set of doctors would have an impoverished vision of their *res publica* if they specified their treatment and care ambitions in the minimalistic terms of merely keeping people alive.

Similarly, just as a set of doctors would have a morally suspect vision of their *res publica* if they specified it in terms of treating one or two ailments that predominantly affected people like themselves, so too a polity would have a morally suspect vision if it took the interests of one set of people as the 'good' it should secure, valuing other people as three-fifths of a man, for example.

However, even when doctors have an impoverished or suspect vision of their *res publica*, a particular valuable goal still drives them, namely, protecting health. Which is more than can be said of some professions, such as executioners or interrogators who use extreme techniques like torture. Although, as noted above, at a very basic level, these enterprises can be viewed against a backdrop of morally plausible aims (namely, the aim of dealing with serious wrongdoing), nevertheless the specific goals that drive their work are reprehensible. The question is: Is the criminal law like doctors' codes, or occasionally more like executioners' codes? As noted above, Duff maintains that polities and the criminal law are clearly aiming at something good, namely, determining how we are to live together. But, this may claim too much on behalf of the criminal law. When a polity's vision for itself is morally suspect, it may be driven less by the good aim of determining how to live together than by the suspect aim of ensuring that those on top remain there.

Since not all polities have a morally acceptable self-understanding, just as not all professions have a morally acceptable self-understanding, we should not take a polity's declared *res publica* as an unassailable reference point for what *properly* concerns it. And, since political insiders, like professional insiders, are poorly placed to engage in clear-eyed self-scrutiny, we should be hesitant to accept their declarations about the nature, function, and value of their code of civic ethics. Just as we need an independent defense for establishing or retaining any given profession, so too we need an independent defense for establishing or retaining a given public realm and the code of civic ethics created to regulate it.

Duff stresses at the outset of Chapter 2 that, although different polities will have different outlooks and visions for themselves, his interest is not in

what a given polity positively declares to be morally wrong, but in what is *actually morally wrong*. We want to allow then that, although there is scope for cultural variation in the specification of the business of the polity, nevertheless many polities fundamentally get it wrong, and are built around ideas that should not be the driving purposes of polities. What they declare to be 'wrong' is *not* morally wrong either *tout court* or in relation to the specific interests of that 'public'.

5.2 Founding Parents and Other Problems

Duff says that Founding Parents '…cannot take the existence of criminal law for granted; they must justify its creation and maintenance…we will need a justification, in terms of the polity's appropriate aims. That justification will need to be grounded in a political theory of the polity,…' (97) The term *appropriate* here is like the terms *proper* and *legitimate* noted above. It's a slippery character. In the case of the law, this slipperiness is partly a matter of final jurisdiction. The state presents itself as the final authority on the boundaries of jurisdiction. Hence, the appropriate aims of the polity are (ostensibly) largely what the polity says they are. And, Duff acknowledges that, unlike drafters of professional codes, a set of Founding Parents could, in principle, decide that their criminal code will see every kind of moral wrong as a candidate for criminalization. But, he says that it's highly implausible that they should do so; such an approach is inconsistent with core liberal values, and it relies on a deeply retributivist outlook. This answer offers some reassurance – that there are broad substantive limits to what reasonable Founding Parents' vision will be for their polity – but nonetheless these insiders, who may well lack the capacity for clear-eyed self-scrutiny, seem to be given the final say, in Duff's view, on their polity's appropriate aims.

A further problem concerning Founding Parents, as well as all drafters of Codes, which I didn't highlight above, pertains to who those drafters are. The language of *Founding Parents* is as jarring as it is illuminating. One problem with Duff's description of the drafters as 'Founding Parents' is that it occludes the fact that the people drafting the laws in almost all polities have not been Parents, but Fathers and, indeed, white Fathers. And, in cases like the US, the thinking of those Fathers has had considerable influence over judicial reasoning and legislative activity. (Presumably, it would be, or should be, problematic for the medical profession if only male doctors were permitted to participate in devising its code of ethics, regardless of what they devised.) This leaves us with the question: Can legitimate and morally acceptable polities be produced through suspect

constitutional processes?

Two further problems with criminal codes that are akin to the problems of professional codes are, first, that much of the criminal law is written *ex post*, long after a polity has entrenched itself as a 'legitimate' entity. When this is the case, the criminal law reflects the polity's self-effacing attitude toward its *res publica*; the law is built on the assumption that the polity's goals and activities are indeed legitimate, when it fact they might not be. Second, even in a polity with a morally respectable set of goals and activities, the criminal law is clunky, inflexible, and general. Both codes and the criminal law, to the extent that it resembles them, limit and distort moral consideration.

6. CONCLUSION

This paper has highlighted problems in Duff's use of an analogy with professional codes of ethics to flesh out an image of the criminal law as a code of civic ethics which, like professional codes, regulates the conduct of members of a particular 'public' realm. First, members – as insiders – are biased toward their enterprise and, hence, are poorly placed to judge its merits or devise a code to regulate conduct. Second, the enterprise that members seek to regulate with a Code may be morally abhorrent. Indeed, the most morally suspect professions and states seem to dress up their regulation of their activities in the most morally-purifying language. Third, both empirically and morally, people cannot isolate their different spheres of activity within different 'public' realms. When they do wrong in one sphere, that wrong is often the business of people whom Duff would deem to be outsiders. Fourth, the business of the polity cannot be specified adequately in formal terms alone; substantive commitments have to be made, and insiders – Founding Parents or code drafters - should not have the final say on what those commitments should be. Finally, it matters who the code drafters are. In the case of the criminal law, we can reasonably wonder whether it's morally suspect because it has been drafted largely by one segment of society, privileged white men.

In closing, let me note two points of disanalogy between professional codes and the criminal law. These points won't undercut my main conclusion that we have reason to be wary of both codes and any criminal law like them as accurate trackers of moral wrongdoing and accountability regardless of the *res publicae* they regulate. But, these points of disanalogy do highlight the limits of appealing to professional codes to inform our thinking about the criminal law.

First, one minor point of disanalogy that Duff notes is that: unlike the criminal law, 'A professional code might cover not only matters of ethical (mis)conduct, but also matters of competence… (i.e. fitness to practise medicine; responsibility to remain informed of new findings, keep skills up to date, maintain level of cognitive and physical competence, etc.)' (80).

Second, another disanalogy is that most, if not all, professional codes include a sentence at the outset along the lines that, upon appointment, all members are required to confirm their commitment to this code of conduct.[10] By contrast, although voluntarist accounts of political obligation have their defenders, and although immigrants often do have to declare their allegiance, nevertheless most members of a polity do not join it voluntarily and, typically, their polity cannot revoke their membership.

Codes of professional ethics are a seemingly useful analogue for the criminal law because they are codified. They have the same 'look' as the formal part of the criminal law. But, their use in an analogy comes at a high price and may heighten, rather than lessen, critics' skepticism about Duff's account of public wrongs and our standing to hold each other to account.

BIBLIOGRAPHY

Brecher, B. 2010. 'The Politics of Professional Ethics'. *Journal of Evaluation in Clinical Practice*, 16(2), 351–355.

Duff, A. 2018. *The Realm of Criminal Law*. Oxford: Oxford University Press.

Pauer-Studer, H. 2012. 'Law and Morality under Evil Conditions: The SS Judge Konrad Morgen'. *Jurisprudence* 3(2), 367–390.

Pauer-Studer, H., and J. D. Velleman. 2011. 'Distortions of Normativity', *Ethical Theory and Moral Practice* 14, 329–356.

Zaibert, L. 2018. *Rethinking Punishment*. Cambridge: Cambridge University Press.

10 The General Medical Council's Code of Conduct opens with the statement: 'For us to command the confidence of all of our key interests, it is necessary that Council as the governing body should adopt and comply with appropriate standards of conduct. Upon appointment, all Council members are required to confirm their commitment to the Members' Code of Conduct.' See: https://www.gmc-uk.org/about/how-we-work/governance/council/code-of-conduct

DOI: 10.31009/LEAP.2019.V7.11

Democratic dialogue, multiculturalism and "public wrongs" *

ROBERTO GARGARELLA
CONICET

ABSTRACT

In this work, I challenge some of the ideas presented by Antony Duff's in his book *The Realm of Criminal Law* and, more particularly, his approach to the idea of "public wrongs." I claim that his views on the subject unjustifiably put into question some shared ideas about what it means to live in a democratic and multicultural society. More particularly, I maintain that, in multicultural societies, we define the basic elements of our identities in dialogue with different and "significant others," which –I submit- makes it difficult to identify public wrongs that are –in his words- "clearly inconsistent with… any remotely plausible conception of civil order". In my opinion, many of the problems that I find in his views originate in Duff's implausible understanding of democracy: the "voice of the community" –I claim- can and should only be expressed by the community itself, through an ongoing and unending collective conversation.

Keywords: democracy, "public wrongs", mululticulturalism, liberalism, communitarianism

> The Ulama was a popular ballgame with ritual associations played since 1400 BC by the pre-Columbian people of Ancient Mesoamerica. Among the Mayans, the game was known as Pok a Tok, while the Aztec called it Tlachtli. The game, which is still played in a few places by the indigenous population, constituted a central part of the indigenous American culture. The ballcourts were the size of a modern day football field. The game was played on a rectangular court with slanted side-walls against which the ball could bounce. Usually, those walls were decorated with the images serpents and jaguars and also with images of human sacrifice. The rules of the game are unknown, but – according to the images and sculptures that were found in the courts- it was like the racquetball, where the players struck a ball, made of rubber, with their

* I want to thank Fernando Bracaccini for his generous reading of a previous draft of this paper.

> hips. In some late developments of the Ulama, which seemingly took place during the Classic era, the game was combined with religious human sacrifices. In these occasions, one of the teams was offered as a sacrifice after the end of the game. Remarkably, it seems that the winning squad or its captain –rather than the defeated team or some of its individual members- were the ones to be decapitated after the game.[1]

1. INTRODUCTION AND ILLUSTRATIONS

In what follows, I want to critically examine some aspects of Antony Duff's book The Realm of Critical Law (Duff 2019).[2] In particular, I want to call the attention about some problems related to the notion of "public wrongs", which plays a central role in this and some other recent works of his (Duff 2009, 141; Marshall and Duff 1998). In this paper, I shall mainly discuss four of those problems: i) the difficulty of maintaining a notion of "public wrongs" such as the one that Duff wants to maintain, in the context of (radically) plural societies; ii) the confusions created by Duff's oscillations between liberal and communitarian positions; iii) the institutional difficulties implied by Duff's views on the matter; iv) the tensions that may emerge between "the voice of the community" and democracy.

As a general, introductory point, I would claim that Duff defines public wrongs in ways that –I submit- conflict with the demands of our multicultural societies, and also with some of our most basic assumptions about democracy. By the end of the paper, I will briefly sketch an alternative approach –related to the notion of deliberative democracy- which I think could better accommodate those serious demands.

Before presenting some of my criticisms to Duff's views, let me quickly offer a few examples related to the challenges that multiculturalism poses to our traditional Western or liberal approaches to the law in general, and to the Criminal Law in particular.

In recent years, the Colombian Constitutional Court recognized as legitimate certain forms of indigenous justice that many still consider as examples of "torture." For instance, in a 1996 (*Tutela* 349/1996, p. 10), and following a claim brought by an Embera-Chamí Indian, the Colombian Court defended the institution of the *cepo*, which was a a kind of corporal

1 The content and details of these events are still unconfirmed, but the story, which has been told once and again, belongs to our shared knowledge -and beliefs- about the life of aboriginal groups before colonization. In the end, the story simply offers an extreme illustration of cultural diversity and its implications.

2 From here on, I shall only quote the pages of the book when I refer to *The Realm of Criminal Law.*

punishment common among certain indigenous communities, which was imported from Spanish colonial law (Van Cott 2000, 218). Similarly, on 15 October 1997 (T-523/1997, in the "*Jambalá* decision"), and in the name of protecting the community's customary law, the Court upheld a form of indigenous punishment that included whipping, and which Amnesty Law considered at that time an expression of torture (ibid., 221).[3]

Other "hard cases" may be those involving minors living in indigenous communities. For example, in the province of Salta, Argentina, there was recently a case related to a Wichí girl victim of alleged sexual abuse by her stepfather (the case is from 2005).[4] In that opportunity, members of the Wichí community maintained that the alleged abuse represented, in fact, a common practice that belonged to the group's traditional culture. Another difficult and paradigmatic case was that of the U'wa twins in Colombia, where an aboriginal couple refused to raise the twins out of the traditional belief that giving birth to twins could create misfortunes for the community. In Santa Cruz del Quiché, Guatemala, a local indigenous leader "punished" with a spanking a 12-year-old child, accused of theft, which generated complains from the Human Rights Ombudsman, who accused the community's leader of "child abuse"[5] (Ramírez 2017, 16). All these cases offer us extreme examples, difficult to accept or tolerate for most of us, individuals who have grown up in social contexts defined by the Western culture and a liberal philosophy.

2. RADICAL PLURALISM AND THE DIFFICULTY OF DEFINING "PUBLIC WRONGS". DIALOGIC IDENTITIES

The examples of radical pluralism that I have presented in the previous section call our attention about the existence of a plurality of world-views, which may include substantially different understandings of how to live a good life, but also radically different approaches to the notions of "right",

3 According to ex-Magistrate Ciro Angarita, the Court's decision reflected a division within the Court, between those who "absolutely reject the possibility that indigenous 'usos y costumbres' can be considered sources of law… [and] another, which accepts, on the contrary, that respect for this alternative source of law – to the extent that it is not contrary to the Constitution and the law – constitutes an expression of the ethnic and cultural diversity of the Colombian Nation and, as such, has a firm but conditional pretext in our [normative] system" (ibid.). See also Bonilla (2003).

4 The case was the object of a judicial process in "Ruiz, José Fabián. Abuso sexual con acceso carnal calificado". Expte 3399/5.

5 See the newspaper "Prensa Libre", 13th October 2017.

"wrong," and the like.[6] The mere existence of these radical differences has interesting implications for a theory of Criminal Law like the one that Antony Duff has been building up in recent years. In the first place, I shall maintain, those examples cast doubts about the author's approach to the concepts of *public wrongs, mala in se* and *mala prohibita*, in particular.[7]

If we take into consideration Duff's strongest definition of the concept of "public wrong", the challenge posed by our examples seems more serious. For instance, in different parts of his work, and thinking mostly about extreme cases like those of murder and rape, Duff defined *public wrongs* as faults that "are clearly inconsistent with, manifest violations of, any remotely plausible conception of civil order –any conception, that is, of how the members of a polity can live together as citizens" (300; and also Duff 2009, 141; Marshall and Duff 1998). Similarly, Duff defined *mala in se* as "crimes consisting in conduct that is (uncontroversially) wrongful independently of its criminalization" (20; also Duff 2001, 23-4; Duff 2013, 181). When we define public wrongs (and like concepts, such as *mala in se*), in this strong form, we leave little room for radical pluralism. To put it differently, in this way we show having little respect to alternative cultures.[8]

It should be clear, in fact, that different cultures may value different things for different reasons, at different times of their own history: why is it that a culture different from ours could not treat "our most cherished values" in a completely different way than we do, and still be a decent community, worth of respect?[9] Why could not they understand the right to private property, or even the right to life or the right to sexual integrity in ways that radically –and reasonably- challenged our own approaches to

6 The idea of radical pluralism is fundamentally related to John Rawls' notion about "the fact of reasonable pluralism" (Rawls 1991). However, with this alternative concept I also want to encompass world-views that in principle may sound unreasonable –while (I would suggest) in the end they should not be considered as such. The same Rawls makes an effort to consider such rather extreme views, for example, in Rawls 2002. Thus, for instance, in his references to "hierarchical" and "nonliberal" peoples (ibid., Part II).

7 Duff –I shall suggest- defines these concepts differently in different parts of the book, but none of his definitions seem to deal properly with radical pluralism.

8 As Duncan Ivison put it: "Do 'we' share, as a social, political, and legal community, a form of life within which diverse and overlapping sub-communities can find their place? Or are 'we' instead merely a diverse collection of incompletely articulated communities, between whom communication –when it occurs- is erratic, superficial and at worst, hostile?" (Ivison 1999, 89-90).

9 Note that I am making reference, most of all, to different communities sharing the same geographical space, this is to say, being part of the same multicultural society (Duff recognizes, of course, that different –even repugnant- civil orders may exist. See, for instance, Duff 2016)

those matters?[10] In the context of democratic and multicultural societies, I believe, we should be able to accept the "fact of disagreement," and learn to live with it (Waldron 1999, Bracaccini 2018).

Of course, to state this –to recognize that other cultures may have a completely different approach to our most cherished values- does not imply endorsing cultural relativism. What I am trying to say, instead, is that we have reasons to be more attentive to and respectful of other cultures, particularly in the face of established practices of cultural imperialism and social imposition. More significantly, to accept the fact that other cultures may repudiate or radically challenge our most cherished values, does not necessarily imply saying something in favor of those alternative values or cultural practices: those shared practices, shared values and shared understanding may result valuable -or not at all- from a critical perspective. We may have reasons, or not, to accept, support or enforce those alternative shared understandings. But these are things to be discussed apart, with independence of the point I am making here. The point is: we need to recognize that different cultures may resist, repudiate or dismiss, rather than endorse or share, what we consider to be our "shared values".

Moreover, one may consistently recognize the importance of cultural differences and be at the same critical in relation to all established cultural practices –even, or more particularly, concerning long-standing shared practices. Usually –one could claim- those shared practices tend to be the result of cultural constructions, and as such the product of a complex and problematical history –a history that may combine deep convictions and shared commitments, together with profound injustices, unjustified inequalities, exploitation and domination. The history of private property may be a good example of this combination of noble reasons and obscure impositions. In sum, we could both challenge Duff's approach to the concept of public wrongs (public wrongs as wrongs that are "clearly inconsistent with…any remotely plausible conception of civil order") and be critical about cultural relativism.

In other passages of the book, Duff seems to define the concept of public wrongs in a slightly different way. He thus refers to public wrongs as

10 Let me suggest this: perhaps, at the deepest level, all of us, as human beings, share certain attitudes of disgust or profound discomfort with regard to certain human conducts. For instance, some could claim that is "intrinsic" to our human nature the tendency to react against unfair behavior. This may be true, but then our reaction against murder, for example, would be derivative from that "intrinsic" rejection towards unfairness. As a consequence, too, "murder" would not be "intrinsically" wrong. What would be wrong, instead, would be the act of unfairly depriving someone of his life. In addition, the fact that we had these reactions or feelings of disgust would not give us reasons, per se, against those actions capable of triggering such feelings on us.

conducts that are "intrinsically, rather than consequentially, inconsistent with civil order [because] they violate it in themselves, since they are utterly inconsistent with the terms on which citizens could live together" (300). This different definition seems also odd, particularly in the work of an author who has consistently show to have clear communitarian sensibilities (more about this below). Given those sensibilities, how could he then refer to conducts that are "intrinsically…inconsistent with the terms on which citizens could live together"? How to maintain such an approach, in the face of societies that are in a continuous process of cultural development? What would be "utterly inconsistent…", etc., in societies that are in a permanent process of learning and changing? Think, for example, about the Me Too movement, its meaning and implications: if the *me Too* movement taught us something, this was that what seemed to be "natural" a few years ago –say, a certain distribution of work and social roles between men and women- may result "outrageous" nowadays. The point is: we "construct" the notions of "appropriateness"; "outrageous", "immoral"; etc., in an ongoing process of "dialogue" with others. Within this alternative picture, there seems to be little room for "public wrongs" understood as wrongs that are "intrinsically…inconsistent with civil order". As Nicola Lacey put it (in a review of Antony Duff's most recent work), there is "no space outside culture: no 'view from nowhere' from which to appropriate criteria for the relevance of evidence, or reasonableness, or indeed the definition of criminal wrongs can be determined" (Lacey 2011, 307).

The communitarian author Charles Taylor presented a similar point a few years ago, with singular clarity. For him, "the crucial feature of human life is its fundamentally *dialogical* character" (Taylor 1992, 32). Following George Herbert Mead and his idea of "significant others" he adds that "the genesis of the human mind is…not monological, not something each person accomplishes on his or her own, but dialogical" (ibid.). For Taylor, "we define our identity always in dialogue with sometimes in struggle against, the things our significant others want to see in us" (ibid., 32-33). Taking into account these perspectives, Duff's claim about public wrongs sounds problematic. In what concerns our shared values, there is nothing that stands "out there," nothing to be "discovered" or "declared" as a pre-legal or pre-existing public wrong. Rather -I would suggest, following Taylor- we craft those shared values in a continuous interaction with our "significant others". As a result of these interactions, we may finally decide, through more or less formal procedures, to encourage, discourage, and even punish, certain kinds of behaviors. But all of this would then be the product of a collective, human creation, rather than something to be discovered in the "intrinsic nature" of things, or "declared", as pre-existing or pre-legal wrongs (more about this below).

3. THE DIFFICULTY OF BEING LIBERAL AND COMMUNITATARIAN AT THE SAME TIME

Let me now consider a different definition of "public wrongs", which can also be found in Antony Duff's recent publications. The alternative definition that I am now considering seems to be more sensitive to context, and therefore more dependent on considerations of "time" and "place". For instance, in certain occasions, Duff defines public wrongs as wrongs that are of concern to citizens as participants in the political community, where citizens are "engaged in a distinctive enterprise of living together, structured by a set of values that help to define that enterprise" (Duff 2010, 5; Edwards & Simister 2017, 110). For this view, what counts as the polity's business "depends on an account of the defining aims and values of such a polity", where those "aims and values" –it is here assumed- can "change over time" (Duff 2007, 142; Edwards & Simister 2017, 110). Similarly, in certain parts of his new book, Duff defines public wrongs as wrongs "directed against or impacting on some aspect of the polity's collective life…which concerns all citizens as citizens, and to which they must therefore respond collectively" (299). What all these views have in common, I believe, is a certain communitarian flavor: public wrongs are not seen here as permanent or unchangeable features of our life in common.

Now, and taking into account this new approach to the notion of public wrongs: Why would we have reasons to provide especial protections to those "aspect[s] of the polity's collective life…which concern[] all citizens as citizens"? Or, more specifically: What reasons would we have to define as public wrongs all those conducts that seriously challenged "significant aspects of the polity´s collective life"? Consider this example: knowing of the profound love we have for our community in Catalunya, someone may choose to defy our national feelings, make fun of our ancient communal traditions, and even insult us for belonging to this "land of nationalists". The vast majority of us could take such conducts as directed against us, and contrary to what define us as citizens of this community. Someone could seriously claim, in Catalunya, that those insults "deny or negate the conditions under which we can live as citizens who share a civic life" (300). In fact, if we define our civic identity in terms of a shared language and same shared traditions, then to launch an attack precisely against those shared values (those values that define us as Catalans), would imply to attack the same social bonds that keep us together and define who we are. In sum, someone could claim that the challenge posed by those anti-nationalists aggressions seriously and negatively impacts "on some aspect of the polity´s collective life," in ways that concern all us as citizens of this

country. But the question is: do we really have reasons to consider such wrongs (those insults, that mockery) as public wrongs? Do we really have reasons to take those anti-nationalist outbreaks so seriously (this is to say, in ways that could even require the use of our coercive powers)? Could we really say, in those circumstances, and following Duff, that those offenders are required to "answer to [their fellow citizens] for wrongs that are their business", because they have violated the values that help to define what the polity is? (Duff 2007, 142; Edwards & Simister 2017, 110).

Probably, Duff would object to that example. And he would (now) do so in the name (not of shared, localized, communal values, but rather) of more universal values. This answer, however, would be unattractive, and its unattractiveness would be originated in Duff's unjustified oscillation between the universal and the communal; the general and the particular; the critical view about the structures in which we live, and the attachment to the communities where we live. Typically, Duff claims: "we must begin with individuals in community: with individuals who already recognize themselves as living in community with others" (191) and also that we see our "good as partly bound up with the collective good of my fellow citizens" (190), *but also*, and at the same time, that "we should look critically at the structures and practices of our polities: they are not given, as the unchangeable framework of our civic life" (ibid.). These quotes, which appear almost in the same page, represent –as I shall illustrate- just one example of Duff's persistent attempt to reaffirm and resist liberal and communitarian values.

Duff, it seems to me, wants to vindicate both things at the same time (say, universal and local values), which may sound attractive in principle, but results untenable in the end: one cannot consistently defend both things as part of the same theoretical framework. Unfortunately, the fact is that Duff endorses a peculiar political philosophy, which at the same time wants to be republican, liberal and communitarian (186, 189). To reclaim, at the same time, all these different philosophical commitments, sounds strange within the context of contemporary political philosophy. In the end, what explains the existence of these diverse philosophical approaches is the fact that they offer different, and sometimes even opposite answers, to similar questions. Then, one may choose to declare his adhesion to different political philosophies at the same time but, in the face of the difficult questions –in those situations where having one or another political philosophy really matters- one needs to choose, and declare where he stands. At that point, it results simply unreasonable to continue advocating for opposing political philosophies at the same time.

What are we supposed to do, for instance, when the so-called "shared values" of our communities appear not to be shared by one –say, the dissident, the anarchist or the anti-conformist? Are we supposed to enforce the values of our community, so as to ensure that the basic cultural structure of our society is not undermined, or should we rather protect the individual rights of the dissident, no matter the impact of his challenge on the bonds that held us together? What should we do in the face of conducts that profoundly offend our shared understandings? What are we supposed to do with the anti-nationalist that mocks the symbols that –we assume- define us as "Catalans" –the symbols that structure our common identity?[11]

In sum, it is in the face of these difficult cases that we are required to define whether we are primarily liberals or communitarians. It is precisely in these cases when we cannot just claim to be, at the same time, liberals, communitarians, republicans and deliberative democrats. Unfortunately or not, we are required to take sides.

4. THE INSTITUTIONAL IMPLICATIONS OF "DECLARING PUBLIC WRONGS": A PATCHWORK OF VIEWS

Having reached this point, I shall now focus on another aspect of Duff's approach to the notion of "public wrongs", which I also find problematic. I am referring to the question about how to reach public decisions about "public wrongs," through the institutional system.

In a crucial section of his book, which appears under the title "The Law's Voice", Professor Duff refers to the way in which public wrongs are "declared" as such. He states: "The criminal law…,should be seen as a communicative enterprise: in its substantive dimension, it declares certain

11 Note that the reasoning behind this example –where we consider the anti-nationalist attack as an attack against "us"- does not significantly differ from the one that Marshall and Duff offer, for example, in order to justify the treatment of "rape" as a shared wrong. They claim that women may consider rape "as a collective, not merely an individual, wrong (as an attack on them), insofar as they associate and identify themselves with the individual victim. For they define themselves as a group, in terms of a certain shared identity, shared values, mutual concern –and shared dangers which threaten them: an attack on a member of the group is thus an attack on the group…the attack in this individual victim is itself also an attack on us –on her as a member of the group and on us fellow members" (Marshall & Duff 1998, 19-20). The question is then: Why should the attack against one women be considered an "attack on the group" (given our "shared identity, shared values, mutual concern," etc.), but not so the attack against the one who sings our national anthem? Finally, I want to note that, for reasons of space, I am not making reference, here, about another significant in Duff's view of the subject, which concerns the cases in which victims (i.e., victims of domestic violence) refuse to share their wrongs with the community. In this respect, see, for example, Dempsey (2011).

kinds of conduct to constitute public wrongs—to be wrongs that merit a formal public response; in its procedural dimension, it provides the process through which those accused of committing such wrongs are called to formal public account" (109; Duff 2001).

The fact that certain wrongs are "declared" to be public wrongs is particularly important for our approach to the matter, because it implies denying that those wrongs are, in one or another way, created, or collectively defined as such. Duff claims, for instance, that "the law that defines such *mala in se* as criminal does not create their wrongfulness: rather, its point is to declare that these (pre-existing, pre-legal) wrongs are, when committed within the polity's jurisdiction, wrongs for which their perpetrators are liable to be called to answer by the polity through its criminal courts" (123). Moreover, for him, "[t]o say [that substantive criminal law[declares these norms is to deny, implicitly, that this is a matter of norm-*creation*; although it involves a process of developing precise determinations of norms that are extra-legally somewhat vague or indeterminate" (207).[12]

We have already challenged Duff's claims about intrinsic wrongs, and suggested instead that people –and communities- constitute themselves through dialogic, creative processes. Now, we are going to express our reservations concerning Duff's idea about public wrongs as wrongs to be "declared," rather than created or collectively defined (collectively defined, for instance, through an ongoing and unending process of public conversations).

There are numerous things to be said concerning Duff's view about "declaring public wrongs", but here I just want to concentrate on one particular aspect of the matter. The problem I am thinking about relates to the *institutional mediations* required by these "declarations". The problem originates here: according to Duff's approach, some of the most crucial

12 Notably, in this book, more than in any other of his works, Duff acknowledges the "fact of disagreement", and the importance of public deliberation in the area of Criminal Law. In the conclusions of his work he for instance refers that he tried to argue for a "thin, formal principle [of criminalization, related to public wrongs], which leaves most of the substantive work to be done through what should be a process of public deliberation about the construction of the polity's civil order, about what kinds of conduct should be counted as public wrongs within that order, and about appropriate ways of responding to such wrongs" (332-3). These kinds of statements represent an interesting and novel development in his theory, given the way in which they recognize both the existence of disagreements and the importance of deliberation. However, they also show the persistent ambiguities we find in his views, between universalism and localism; liberalism and communitarism; legal moralism and subjectivism; etc. Duff needs to still clarify, for instance, how those views fit with the ideas of mala in se; or the pre-legal, pre-existing, non-created and declared public wrongs: What, then, is going to be discussed in those deliberations? The substance of the Criminal Law? Its content? Its details?.

decisions about our life together would be adopted "for the people," "in the name of the people," but finally "by a few people," and without communicating or discussing with the rest of them.[13] The fact that someone is going to "declare" what wrongs are to be considered "public wrongs", and do it in our name, but without talking to us, creates serious difficulties of "mediation", and serious risks of "distortions": who is going to do decide such important matters for us? Who and for what reasons could be authorized to speak in our name? How is this person or group going to know what matters to us, without talking to us first? Note, again, that we are dealing with very risky material here: we do not want to commit mistakes in this regard (say, regarding what to define as public wrongs), because our life and liberty may result thus compromised.

Now, those "declared" public wrongs have to go through a complex institutional process before they can become an integral part of our Criminal Law. A legislature may include those public wrongs as part of our Criminal Code; an administrative agency –the police- may detain someone in the name of a Criminal Law that criminalizes those public wrongs; a judge may be required to decide a conflict related to the application of the Criminal Law; etc. All those institutional mediations complicate things further, in a serious way.[14]

In fact, for the Criminal Law to be able to "declare" certain pre-existing wrongs as public wrongs, a complex institutional system needs to be in place. When the "declarative" process becomes in this way institutionalized– say, through the intervention of legislative, administrative and judicial agencies- some serious risks immediately appear. For instance, through that process, the voice of the law may come to distort, rather than express in any meaningful way, the voice of the community. I mean: in those conditions it is highly probable that the voice of the law (say, the voice of the law declaring certain wrongs as public wrongs) will have little connection with what the community itself would have said, had it been consulted about those matters. Predictably, at each stage of the process, interest groups, pressure groups, private interests, bureaucratic interests, etc., will come into play, adding new "distortions" or "noise" into our

13 That the decisions at stake are crucial for our life in common should be clear from the start: we are talking about issues related to our most cherished values, and also about questions that may require the use the coercive state apparatus, in its most extreme forms. This is why we need to be especially careful regarding how we define and act upon those public wrongs.

14 Note, however, that Duff refers to a criminal law system that first, "declare[s] and define[s] wrongs", and second "provide a process through which those accused of committing such wrongs can be called to formal account" (214). The "process" seems to be provided for the second task, and not for the first one.

decision-making (or, for this matter, "declarative") process.

In sum, after the intervention of so many different public agents with very different democratic credentials, different legitimacy, and different interests, elected in different times and through different processes, what we most probably obtain is a *patchwork of views* rather than decisions that express the community's voice. As Carlos Nino put it:

> "there is no guarantee whatsoever that the result of this awkward mix of different decisions, which can ultimately respond to a combination of findings from different debates, carried out by different groups of people at different times, have some resemblance to the majority consensus that obtained after an open and free debate" (Nino 1991, 578).

Duff's theory, which aims to be able to make sense for actual societies, would need to address these risks more seriously. In the end, and given the institutional framework that we have in place, in what sense could we expect that the Criminal Law became "our" Criminal Law? Under those institutional conditions, what reason could we have for maintaining that the voice of the law expresses the voice of the people? How could someone reasonably claim that the Criminal Law speaks with the community's language?

5. THE TENSIONS BETWEEN "THE VOICE OF THE COMMUNITY" AND DEMOCRACY. A MILLEAN APPROACH

Let me now explore a little bit more Duff's statements about the law as an expression of "the community's voice". What Duff says in this respect is enormously relevant for those of us interested in establishing stronger connections between the Criminal Law and democracy.

In one of the many paragraphs he wrote in this respect, Duff stated: "when the substantive criminal law defines certain kinds of wrong as public wrongs, it must speak, or claim to speak, in the voice of a linguistic and normative community whose law it is, and whose values it thus declares; and it must speak to an audience who can be expected to understand those declarations as grounded in what could be reasons for action—reasons for them to act. In defining theft, for instance, as a crime, the law defines it as a wrong—something that its addressees have normally conclusive reason not to do, for which they will be liable to be called to public account if they do it; but they can have such reasons only if they can

understand the (supposed) wrongness of theft, and the law that declares its wrongness" (110). This view, I want to emphasize, has always been present and has always played a central role in Duff's approach to the subject.[15]

Now, I find numerous problems in this view, which begin with Duff's suggestion about a law that "must speak or claim to speak" the voice of the community. Obviously, it is one thing to "speak" the voice of the community, and another very different one to "claim to speak" its voice. Anyone can claim to speak for the community, but only the same community –I shall maintain- can speak for itself, with its own voice.

The idea that only the same community can speak for itself relates to the particular conception of democracy that I endorse, namely a deliberative conception of democracy, and its Millean foundations. My point of departure is, in fact, John Stuart Mill's claim, according to which each person is to be considered the best judge of his own interests. For Mill, in fact, each person has something personal and non-transferable to say about him or herself and how the external world affects his or her own interests. In his famous book *On Liberty*, and trying to provide support to his proposed "harm principle", Mill introduced his view on the matter. He claimed:

> "That the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise, or even right. These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him, or visiting him with any evil in case he do otherwise.

15 In other crucial writing on the same topic, Duff maintained: "my understanding of what is said to me, and of the possibility of saying it for myself, depends not merely on the content of what is said, but on the voice in which it is said –on who speaks to me, and in what accident; indeed, the content of what is said, what It means to the hearer, cannot be entirely divorced from the voice in which it is said…we must… wonder whether it is a voice which all citizens could hear as their –or as a voice which could be theirs" (Duff 1998, 204). In another piece of his work, he claimed: "If there are individuals or groups within the society who are (in effect, even if not by design) persistently and systematically excluded from participation in its political life and in its material goods, who are normatively excluded in that their treatment at the hands of the society's governing laws and institutions does not display any genuine regard for them as sharing in the community's values and who are linguistically excluded in that the voice of the law (through which the community speaks to its members in the language of their shared values) sounds to them as an alien voice that is not and could not be theirs, then the claim that they are, as citizens, bound by the laws and answerable to the community becomes a hollow one" (Duff 2001, 195-6).

> To justify that, the conduct from which it is desired to deter him must be calculated to produce evil to someone else" (Mill 2003, 18).

In support of his view, Mill resorted to at least two additional reasons, which we could call the *motivational argument* and (following Cass Sunstein) the *epistemic argument* (Thaler and Sunstein 2009, Sunstein 2014). The motivational argument says that each individual "is the person most interested in his own well-being". The interest that each person has in his or her own interest, contrasts with the interest that others may have in one's well-being. For Mill, the interest that any other person could have in one's well-being was "trifling, compared with that which he himself has"; while the interest that society could have in one's situation was "fractional, and altogether indirect." Individual self-government would thus be valuable not only in itself, as an expression of basic individual freedom and choice, but also because its connection with each person's well-being.

Together with the motivational argument, Mill made reference to the epistemic argument, which is still more interesting for our present purposes. This argument was based on an empirical assumption according to which, concerning his own sentiments and circumstances, "the most ordinary man or woman…has means of knowledge immeasurably surpassing those that can be possessed by anyone else" (see also Sunstein 2014, introduction). Let me quote –again, at some length- Mill's view on the topic, which is very rich:

> "neither one person, nor any number of persons, is warranted in saying to another human creature of ripe years, that he shall not do with his life for his own benefit what he chooses to do with it. He is the person most interested in his own well-being: the interest which any other person, except in cases of strong personal attachment, can have in it, is trifling, compared with that which he himself has; the interest which society has in him individually (except as to his conduct to others) is fractional, and altogether indirect: while, with respect to his own feelings and circumstances, the most ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by any one else. The interference of society to overrule his judgment and purposes in what only regards himself, must be grounded on general presumptions; which may be altogether wrong, and even if right, are as likely as not to be misapplied to individual cases, by persons no better acquainted with the circumstances of such cases than those are who look at them merely from without...he

> himself is the final judge. All errors which he is likely to commit against advice and warning, are far outweighed by the evil of allowing others to constrain him to what they deem his good" (ibid., 64, emphasis added).

To summarize: for Mill, all attempts by society to decide in the name of a particular individual, thus overruling his or her own will, would be based on "general presumptions, which may be altogether wrong" or "misapplied to individual cases". This is why –Mill assumed- each person had to be taken as "the final judge" in what regards his or her own interests.

Based on these Millean assumptions, and in agreement with other theorists of democracy, I would suggest that no one, but the same community, should be considered authorized to speak in the name of the community's interests (see, for example, Nino 1991). To state it differently: in the same way that each individual should be considered to be "the final judge" of his or her own interests, every community should be taken to be "sovereign" in what concerns its own affairs.

Now, in order to know what the community actually says about its own interests, what it thinks, what it considers to be good or bad, what it takes to be right or wrong –I submit- we have only one reasonable possibility: we need to launch a public, collective conversation, where each person participates as an equal together with all the others. Every voice has then to be listened, every viewpoint has to be considered, every interest has to be taken into account, so as to make it possible that the "voice of the law" expresses the "voice of community". If we did not allow this process to take place, the impartial character of our decision-making process would become compromised –the possibility of properly balancing the demands of all would result seriously affected.[16] Think, for instance, about what Duncan Ivison reported, concerning the situation of aboriginal people in Australia, Canada and the United States. He stated: "It is striking that, historically…many Aboriginal people have come to see the criminal law as simply a means to impose an alien and hostile conception of community over them, justified usually in terms of being for their own good. The fact that indigenous people in Australia, Canada, and the USA are amongst the most arrested and jailed people *in the world*, lends (depressingly) ample support to the acuity of such a perception" (Ivison 1999, 94).[17]

16 In previous work, I distinguished between this "deliberative" approach and both elitist and populist understandings of the Criminal Law. See, for example, Gargarella 2017.

17 And he adds: "How can we speak of the offender violating our common norms and values. Whose norms? Which values?" (ibid., 92).

In sum, either we allow someone to speak in the name of the community -with all the risks of "distortions" here involved- or we design an institutional -deliberative- process aimed at listening and balancing the voice of "all those potentially affected", which would maximize our chances of deciding in ways that are respectful to the demands of all.[18]

6. CONCLUSIONS

In this work, I have challenged some of the ideas presented by Antony Duff's in his book *The Realm of Criminal Law* and, more particularly, I questioned his approach to the idea of "public wrongs." I claimed that his views on the subject unjustifiably put into question some shared ideas about what it means to live in a democratic and multicultural society. More particularly, I maintained that, in multicultural societies, we define the basic elements of our identities in dialogue with different and "significant others," which –I submit- makes it difficult to identify public wrongs that are –in his words- "clearly inconsistent with… any remotely plausible conception of civil order". Our views on the subject –I claimed- tend to be different in different times and places, which does not necessarily mean to say something in favor of these changing cultural values. In addition, I suggested that Duff's approach to this and similar subjects failed as a consequence of his unjustified theoretical oscillations between a liberal and a communitarian political philosophy. Moreover, I objected to Duff's opinions about how public wrongs are "declared", by making reference to the problems that unavoidably emerge when we situate the task of "declaring public wrongs" in the context of an institutionalized democratic system. Finally, I maintained that many of the problems that I find in his views originate in Duff's implausible understanding of democracy: the "voice of the community" –I claimed- can and should only be expressed by the community itself, through an ongoing and unending collective conversation.

18 One of the referees who evaluated this article asked me to clarify how my approach could endorse, at the same time, something like the "harm principle" and the claim that the community should always have the final word concerning how to deal with its own public affairs: it may well happen that the community's deliberative decision may work "against the requirements of the harm principle." I understand the point, but I don't see it as really problematic, as a result of reasons that I cannot fully address within the scope of this article. Summarily speaking, however, I would suggest the following: i) first, I see the "harm principle" as directly related to the principle of "equal moral dignity", which I take to constitute the fundamental presupposition of a deliberative democracy: in a society where that principle is severely violated, democracy (any version of it) could not work; ii) assuming a Jeffersonian view, I would suggest that societies would tend not destroy themselves, if they had fair, proper chances to deliberate about their own future.

their allegiance, nevertheless most members of a polity do not join it voluntarily and, typically, their polity cannot revoke their membership.

Codes of professional ethics are a seemingly useful analogue for the criminal law because they are codified. They have the same 'look' as the formal part of the criminal law. But, their use in an analogy comes at a high price and may heighten, rather than lessen, critics' skepticism about Duff's account of public wrongs and our standing to hold each other to account.

BIBLIOGRAPHY

Bonilla, D., 2003: "Los derechos fundamentales y la diferencia cultural. Análisis del caso colombiano," *Seminario Latinoamericano (SELA)*, Yale University. Manuscript in file with the author.

Bracaccini, F., 2018: "A Law that Speaks in the Voice of the Community? The Problem of Disagreement in the Theory of Criminal Law," unpublished manuscript, in file with the author.

Dempsey, M., 2011: "Public Wrongs and the 'Criminal Law Business': When Victims Won't Share", in R. Cruft, M. Kramer & M. Reiff, *Crime, Punishment, and Responsibility. The Jurisprudence of Antony Duff*, Oxford: Oxford University Press.

Duff, A., 1998: "Law, Language and Community: Some Preconditions of Criminal Liability," 18 *Oxford Journal of Legal Studies* 189-206.

Duff, A., 2001: *Punishment, Communication, and Community*, Oxford: Oxford University Press.

Duff, A., 2009: *Answering for Crime*, Oxford: Hart Publishing.

Duff, A., 2010: "Towards a theory of Criminal Law?" 84 *Aristotelian Society*, Supp. vol. 1.

Duff, A., 2013: "Relational Reasons and the Criminal Law" 2 *Oxford Studies in Legal Philosophy* 175-208.

Duff, A., 2016: "Legal Moralism and Public Wrongs", in K. Ferzan & S. Morse, eds., *Legal Moral and Metaphysical Truths. The Philosophy of Michael S. Moore*, Oxford: Oxford University Press.

Duff, A., 2019: *The Realm of Criminal Law*, Oxford: Oxford University Press.

Edwards, J. & Simester, A. (2017), "What's Public About Crime?", *Oxford Journal of Legal Studies*, vo. 37, n. 1, 105-133.

Gargarella, R., 2017: *Castigar al prójimo*, Buenos Aires: Siglo XXI.

Ivison, D., 1998: "Whose Values? Which Norms? Justifying punishment in intercultural contexts. In M Matravers (Eds.), *Punishment and Political Theory*, (pp. 88-107). Oxford: Hart Publishing

Lacey, N., 2011: "Community, Culture, and Criminalization," in R. Cruft, M. Kramer & M. Reiff, Crime, *Punishment, and Responsibility. The Jurisprudence of Antony Duff*, Oxford: Oxford University Press.

Marshall, S. & Duff, A., 1998: "Criminalization and Shared Wrongs", *Canadian Journal of Law and Jurisprudence* 11, 7-22.

Mill, J. [1859] 2003: *On Liberty*, London: Dover Publications Inc.

Nino, C., 1991: *The Ethics of Human Rights*, Oxford: Oxford University Press.

Ramírez, S., 2017: "La justicia indígena y la justicia ordinaria frente a los conflictos civiles. Camino para su articulación," unpublished report, in file with the author.
Rawls, J., 1991: *Political Liberalism*, New York: Columbia University Press.
Rawls, J., 2002: *The Law of Peoples*, Cambridge: Harvard University Press.
Sunstein, C., 2014: *Why Nudge? The politics of libertarian paternalism*, Conn.: Yale University Press.
Taylor, C., 1992: "The Politics of Recognition", in A. Gutmann (Ed.), *Multiculturalism and the Politics of Recognition*, Princeton: Princeton University Press.
Thaler, R. & Sunstein, C., 2009: *Nudge: Improving decisions about health, wealth and happines*s, Penguin LCC.
Van Cott, D., 2000: "A Political Analysis of Legal Pluralism in Bolivia and Colombia," *J. Lat. Am. Stud*. 32, 207-234.
Waldron, J., 1999: *Law and Disagreement*, Oxford: Oxford University Press.

DOI: 10.31009/LEAP.2019.V7.12

One Masterprinciple of Criminalization – Or Several Principles?

TATJANA HÖRNLE
Humboldt-Universität zu Berlin

ABSTRACT

The paper challenges a fundamental premise that underlies R A Duff's new book "The Realm of the Criminal Law": the premise that criminalization theory must lead us to one master principle. Instead of striving for one unified theory, it seems preferable to be pluralistic from the outset and to develop separate theories of criminalization. The principles that are necessary to evaluate (potential) prohibitions in criminal laws should be developed separately for different groups of conduct: first, conduct that is incompatible with important collective interests; secondly, violent attacks against other persons; third, other conduct that violates another person's right to non-intervention.

Keywords: criminalization theory; master principles; Rechtsgut; rights; collective interests; pluralism.

1. INTRODUCTION

The purpose of this paper is to provide a short comment on one feature in R A Duff's new book "The Realm of the Criminal Law"[1]: the assumption that criminalization theory must develop *one* master principle that could be applied to a vast range of conduct. Against this commitment to a unified theory of criminalization, I propose that we should proceed from different starting points. The task of scrutinizing criminal laws and legislative proposals, or developing legislative proposals, becomes easier if one distinguishes from the outset between conduct that endangers collective interests and conduct that violates the rights of individual persons. In the latter case, it is also useful to distinguish between rights violations in the

1 I would like to thank José Luis Marti for organizing the symposium on this book in Barcelona, February 2019, and for providing us with a wonderful occasion to exchange views on criminalization theory. For the publication, I have mainly adhered to the form of the oral comment, with only a few footnotes added.

core area of criminal law (such as violent attacks against life and physical integrity) and other rights violations (such as theft, fraud and other disregard for property rights). I hope to show that a non-unified approach has the advantage of giving more structure to thinking about criminalization compared to the more general and more abstract principles that might be applied to all kinds of (potentially) criminal behavior.

2. THIN AND THICK THEORIES OF CRIMINALIZATION

What can we expect from theories of criminalization? We might hope to identify *one (and only one)* master principle for decisions on whether a certain kind of behavior should be criminalized. To be called a master principle, the principle would need to be exhaustive and exclusive (Duff 2018: 234). And, if it were to guide decision-making in an effective way, we would need a *thick* theory of criminalization. Duff calls principles thick if they have a rich descriptive content that allows them to be applied without adding further normative judgments.[2] Criminalization theory that focuses on an exhaustive, exclusive and thick master principle would identify an overarching goal or value that could be applied to *all* public discussions about criminalizing conduct and all legislative proposals. It would provide one recipe for a wide range of political proposals. One general idea would be broken down into conclusions and, ultimatelly, lead to norms that prohibit (or demand) certain actions in criminal laws or judicial definitions of crimes. Criminal law theorists can be expected to like the idea that their efforts could steer criminalization this way, as an exercise in deductive reasoning (members of parliament and other political voices might take a different view). For legal scholars in particular, it seems natural to work with deduction because this is a familiar methodological approach, at least in the Continental-European tradition.

Duff is skeptical about the possibility of a thick concept that allows for a substantive theory of criminalization. His ambition is more modest: to propose a thin theory of criminalization. The thin master principle that he develops argues that:

> "A. We have reason to criminalize a type of conduct if, and only if, it constitutes a public wrong.
>
> B. A type of conduct constitutes a public wrong if, and only if, it violates the polity's civil order." (Duff 2018: 275, 277).

2 See for the difference between thick and thin principles Duff (2018): 253 ff.

I agree with his skeptical stance towards the idea that a master principle could be exhaustive, exclusive *and* thick at the same time. The spectrum of behavior that we have serious reasons to consider for criminalization is broad. The range of justifications for criminal norms cannot be reduced to one substantive, thick master value that guides decision-making in a clear-cut, uncontroversial way. The more substantive content a principle has, the smaller its scope of application will be. For some criminal prohibitions, protection of human dignity can be cited as the supporting rationale.[3] However, for most crimes this would be a rather far-fetched idea, for instance, with regard to property offenses and tax crimes. Even if one sets relatively narrow notions such as dignity aside and turns to broader concepts such as autonomy and self-determination, they do not yield a "one idea fits all", substantive master principle. Of course, quite a few criminal prohibitions protect autonomy and self-determination in a straightforward way (for instance, all crimes that involve coercion, such as robbery and the traditional concept of rape). But even in the area of crimes against individuals, there are reasons to prohibit conduct that has no impact on a human being's ability to lead an autonomous life (for instance, injuring an infant should be a crime even if wounds heal without any lasting impact on her future life), and for many offenses against collective interests, it is not the most straightforward explanation that they protect the autonomy of individual human beings. Duff convincingly identifies a negative relationship between thickness and inclusiveness: the more substance a master principle has, the more it will be under-inclusive. Thinner principles cover more cases, but they do not give decision-makers clear guidance (Duff 2018: 255). Duff still opts for *one* overarching concept: the idea of public wrongs while admitting that "the thinness of a principle radically undercuts its substantive utility" (Duff 2018: 262). However, it seems more promising to abandon the idea that there is *one* theory of criminalization, which starts off with one, albeit thin master principle. My own, more radical conclusion is: It does not make sense to strive for a unified theory of criminalization. The alternative is to be *pluralistic from the outset*: we need several separate theories of criminalization.

Before I elaborate the thesis that theories (plural) should be preferred to one singular theory of criminalization, I should mention one more approach of the latter kind. As I am a German legal scholar, readers might expect to read something about *Rechtsgüter* (in a literal translation: legal goods). This is one version of a master principle, which is still popular in the German literature, but it has been increasingly debated in recent years

3 Hörnle and Kremnitzer (2011); Hörnle (2012).

more and more debated.[4] Duff mentions the Rechtsgutslehre in his book and arrives at a rather polite conclusion by describing it as "relatively thin" (Duff 2018: 254). This criticism can be phrased more strongly: It is not only a thin concept, but also so thin as to be an empty concept. The notion of a "good" does not give any guidance at all – every state of affairs could be labeled this way.

3. THE DISADVANTAGES OF THIN MASTER PRINCIPLES AND THE NEED FOR A PLURALISTIC APPROACH

Within academic discourse, the highly abstract, thin nature of master principles might not be considered a serious disadvantage. However, if scholars strive to have some impact on legislative activities, they should also strive for more substantive content. The more arguments turn from the highly abstract to more specific scenarios, the more they have a chance to be heard. Academic writings about criminalization theory should have a purpose beyond our internal discussions: ideally, they should help to make political decisions somewhat more rational. Legislators are not trained in systematic analysis. Their decisions are mostly based on gut reactions, typically emotional responses to incidents that were reported in the media. These emotional responses may reflect moral judgments, but moral judgments that are holistic and intuitive. The essential point is to provide structures for thinking about criminalization. This is not a pointless enterprise. At least in the German system, one can suppose some willingness to engage in reflections during the legislative process. There are formal procedures for parliamentary hearings and the need to provide written legislative material. Both the staff within the ministries who write drafts for new laws and the members of parliamentary committees need arguments. Admittedly, criminal law theory will hardly succeed in convincing politicians to revise opinions that are anchored in deeply felt intuitive judgments. But sometimes there is genuine puzzlement, and these are the occasions when it is possible to have an impact on legislation by providing arguments.

Criminal law theorists tend to prefer one theory of criminalization, a theory that is homogenous and well-rounded. Several different approaches might seem less pleasing, too pragmatic and incoherent from an aesthetic point of view. However, it is more important to differentiate and to decrease the degree of abstraction. My main argument is that it is possible to get *somewhat thicker* theories. The thesis I propose is: we need *several distinct*

4 See Dubber (2005); for competing opinions Engländer (2015) and Kudlich (2015).

theories of criminalization: one for crimes against individuals and another for crimes against collective interests (and, as I will explain below, another subdivision is recommendable for crimes against individuals). My main assumption is that for each of these groups, the logic of thinking about criminalization is different. Rather than glossing over differences and grouping all kinds of conduct together under one thin master principle such as Duff's public wrong principle, (potential) prohibitions should be analyzed *ab initio* separately for different groups of protected interests. If a choice has to be made between one unified, but very thin theory of criminalization and several, somewhat thicker theories, we should sacrifice our academic preference for the neater unified theory.

If one looks at crimes against persons and crimes against collective interests separately, some of the analytical distinctions that Duff introduces (Duff 2018: Chapter 6) can be used to explain different underlying logics, for instance, the distinction between responsive and preventive. Duff uses it to give the harm principle two different meanings:

> "We have reason to prohibit X if conduct X is harmful (responsive)
>
> We have reason to prohibit X if this will prevent harm (preventive)" (Duff 2018: 238).

The responsive-preventive distinction can be applied to the two categories of crimes against individuals and crimes against collectively shared interests. The logic of laws that protect collective interests is straightforwardly preventive. Citizens share an interest in maintaining institutions (and other shared resources) if their existence and well functioning is important for (different) individual ways of living. The point of protecting institutions and resources is, as a rule, that they are needed in a forward-looking perspective.[5] Prohibitions against, for instance, perjury and corruption serve to prevent damage to the effectiveness of institutions (lack of trust in courts and administrative agencies). Even if the content of a shared institution could be described in a backward-looking manner (for instance, foundational narratives about historical events), the purpose of a criminal prohibition (in the example against false narratives) would be future-oriented, based on the assumption that the institution is essential for our living together. I will argue that the reasoning behind criminal prohibitions must include responsive features if the behavior impacts individual persons rather than collective interests.

5 Exceptions regarding natural resources, mainly living species, are conceivable. Arguments for criminalization can be made in a responsive, non-preventive way if the issues are rights of animals (and perhaps plants) rather than their usefulness for humans' purposes – but these debates cannot be taken up here.

4. CRIMES AGAINST PERSONS AS VIOLATIONS OF RIGHTS

4.1 Why Rights to Non-Intervention Matter

A master principle for crimes against persons could be phrased in the following way: We have reason to criminalize conduct if it violates the rights of another person. Duff mentions criminalization theories that emphasize the notion of rights, but he raises several objections. He argues that this approach does not work well for *mala prohibita* or regulatory offenses, Duff 2018: 263. This is a convincing point, which has led me to question the search for one overarching master principle. If one accepts, however, a plurality of criminalization theories, the rights-based approach becomes interesting again for crimes against individuals.

Duff points out that a rights-based theory is (also) just a thin theory. Discusion is still needed as to what kind of interests should be acknowledged as individuals' rights that ought to be protected by the criminal law (Duff 2018: 263). Invoking the notion of rights does not settle all questions. One crucial question concerns the origins of rights. One possibility is to rely on catalogues of pre-legal, moral rights.[6] I prefer a different approach. If the task is to explain and justify practices of *state punishment*, reference to rights should reconstruct basic relations between citizens in this very social role as citizens (and not in the role as moral beings). It is not the vertical relationship between polity and offender that should be the primary focus point of analysis, but the horizontal relationship between citizens. Crimes against persons are not foremost conduct that violates the civil order. They are this, too, but this is not the most salient description. The main feature is interpersonal: one citizen disregarding the right of another citizen, that is, the right to non-intervention. German constitutional theory uses the term *Abwehrrechte* (defensive rights). Constitutional lawyers are usually interested in the right to non-intervention that citizens have against the state (see, for instance, Poscher 2003). However, for the purpose of criminalization theory, the rights to non-intervention that citizens have *against each other* take center stage. The basic relations that underpin state punishment can be depicted graphically as a triangle.[7] The three corners of the triangle represent the state, the offender, and the victim, the three sides relational structures that are characterized by rights. The basis of the triangle is formed by the victim's right to

6 See for such an approach Wellman (2017): 6.

7 The state, at the apex of the triangle, stands in rights-relations to both offenders and victims *qua* citizens. The rights of offenders against the state become important in criminal proceedings and sentence enforcement. For our purpose, the crucial right is the one at the bottom of the triangle: the right to non-intervention. See for a more comprehensive treatment Hörnle (2019).

non-intervention that the offender disregards. Again, this is not a theory about moral agents and moral duties. The rights to non-intervention that persons as citizens have against each other are fewer and much less demanding than their mutual moral duties and moral rights.

4.2 The Core Area

Are rights to non-intervention *pro tanto* rights or categorical rights (see for this dichotomy Duff 2018: 249-253)? *Pro tanto* means that there might be, all things considered, reasons for refraining from criminalizing conduct, even if it violates a right to non-intervention. Categorical rights are sufficient reasons to criminalize the conduct in question, without the need to weigh them against countervailing considerations. At this point, I draw another distinction that leads to three theories of criminalization, arguing that we should distinguish core rights to non-intervention from the wider scope of other rights to non-intervention. One might object that it is not possible to demarcate a "core or the criminal law". Vincent Chiao has challenged the notion of a core, mainly arguing that this paints a distorted image of the criminal law because the vast majority of crimes does not fall into this category (Chiao 2018: 150-159). This is certainly true with regard to the number of offense descriptions and the number of crimes committed. Conceptually, however, a distinction between "core" and "other offenses against persons" makes sense.

One can approach the core of criminal law by identifying individuals' most important rights (rights to life, bodily integrity, sexual autonomy, freedom of movement, freedom from coercion). I propose to define the core of criminal law as violent attacks that violate these most important rights to non-intervention. This definition can be supported by pointing to historical developments, historical developments that were more than just contingent events but crucial in the formation of states and public law. The need to maintain peaceful relationships by outlawing interpersonal aggression was a major force for the emergence of criminal law as public law.[8] Thomas Hobbes, in the "Leviathan", emphasized the commonwealth's protective function against interpersonal violence and transgressions that are to be expected in the state of nature. There is no need to point out that modern states have to fulfill other functions, too, but protective duties remain foundational duties for the state.

If one agrees that the protection of citizens against violent attacks of others is central to justifying why states are legitimate institutions, this supports

8 See for the so-called Landfriedensbewegung Wadle (2001).

the following conclusion: criminal law for this core area must be *both* responsive *and* categorical. The argument that a responsive logic applies needs more explanation. It might seem counterintuitive at first. Arguing that the state has a primary duty to protect citizens against serious interpersonal violence points towards a preventive justification. Duff contends that criminal law is "essentially preventive" (Duff 2018: 261). I agree with the acknowledgment of criminal law's preventive features, but would argue that there is a second level. Imagine that empirical research shows that a criminal prohibition in the core area, for instance rape, does not have measurable preventive effects. This is not a realistic assumption, and we can hardly explore how human beings would respond in a world *without* criminal law in the core area. Existing legal systems do criminalize this kind of conduct. And even if social scientists were to find a gap in one country and design a comparative study, they would need to control a lot of other factors that might influence the frequency of the relevant conduct. However, imagine for a moment a study that concludes "individuals' desistence from conduct X depends on a lot of factors, but not on the existence of a criminal prohibition". Should such a finding lead legislators to abolish the prohibition? Seen from a preventive-instrumental perspective, this would be a logical consequence. One should, however, take into account that criminal laws have a double function: they are norms of conduct, but also the prerequisite for official censure after the crime has occurred. The second feature is important, too. For my argument, the crucial point is to step from states' duties to maintain basic standards of security to state's duties to censure and sanction offenders who disregard the fundamental "no physical violence"-right. The right to be protected against violent attacks and the right to have them censured and punished are separate, but intertwined rights. Having a right to non-intervention against one's fellow citizens would be pointless if its violation goes unnoticed. Rights only are taken seriously if violations are registered and if there is a negative response. For these reasons, criminalization based on the notion of individuals' rights to non-intervention would need to be responsive, too. Because of the outstanding importance of the rights at stake, for the core area, there are also good reasons to assume that states are under a categorical duty to respond with serious censure and sanctions, that is, with criminal punishment. Duff argues "even if there are some egregious wrongs that we *must* criminalize, this holds only if we assume that we have a criminal law" (Duff 2018: 277). I disagree with this conclusion and argue for a *categorical* duty to create criminal law in the core area (not necessarily with all the features of existing criminal justice systems, but with serious, that is not only verbal censure). Other, non-censuring responses are not sufficient, even if they are equally effective in terms of prevention.

4.3 Other Rights Violations

The foregoing thoughts about the core area of criminal law were based on the perspective of criminal law theory. They are hardly important for real political debates about criminalization: We can assume that legal systems, however diverse they are, include prohibitions against violent attacks on the most important rights to non-intervention, such as the right to life or the right to sexual autonomy. Contemporary debates take place in the two other fields. How far should the criminal law safeguard collective interests (for example, how might the environment best be protected or what constitutes corruption)? Do individuals have a right to non-intervention against other citizens beyond the universally acknowledged core area of violent attacks? With regard to the second question, one can demand or criticize criminal prohibitions from two angles. First, as already mentioned, the question could be whether a certain interest is important enough to justify a right to non-intervention. This leaves rooms for debate, but it is possible to develop sub-principles that give some structure to the debates. The relevance of the interest in question for quality of life should be a central criterion. For legal purposes, statements about quality of life and rankings of interests must apply standardized measures, because general norms cannot rely on idiosyncratic preferences.[9] For judgments about rights to non-intervention beyond the core area, it is also conceivable to additionally apply the collective perspective in the mode of a filter. For our shared, collective interest, one can argue that rights to non-intervention should not be recognized too generously. Particularly in heterogeneous societies with diverging group interests, there might be reasons not to accommodate all claims for defensive rights but instead to promote resilience in some areas – but this opens a wide discussion that cannot be covered here.

Beyond the core area, a right to non-intervention is only a *pro tanto* reason to criminalize behavior. It does not amount to a categorical duty. Countervailing reasons need to be taken into account (the same holds for crimes against collective interests). For instance, recognizing property rights does not necessarily mean that the criminal law must be employed to protect *all* private property without the option to differentiate according to the degree of harm, the costs of law enforcement and other factors. Within the contemporary debate about the proper scope of sexual offenses, there is by now agreement on the right to sexual autonomy, a right to non-intervention that is wider than the traditional prohibition of rape involving physical violence (see, for instance, Archard 1998). However, the discourse

9 See for arguments referring to quality of life von Hirsch and Jareborg (1991).

about new sexual offenses (beyond the older concept of rape as a violent act) might leave *some* room to consider, for instance, questions regarding proof and other possible countervailing reasons.

With regard to countervailing reasons, one point deserves more attention: citizens' obligations to protect their own interests. At this point, my proposal deviates from Duff's analysis. He emphasizes that in contrast to tort law, criminal verdicts should not consider splitting responsibility between offenders and victims (Duff 2018: 294). The opposite view contends that decisions about criminalization should take *both* offenders' duties to refrain from rights violations *and* victims' obligations to protect their rights into account. The rationale stems from the notion of subsidiarity: we should only make moderate use of the intrusive (and costly) criminal justice system and thus value other paths to protect individual's interests, including options to protect one's own interests. In German criminal law doctrine, the heading for these discussions about the subsidiary role of the criminal law is "*Viktimodogmatik*" (see Hillenkamp 1981; Hillenkamp 2017). There should be limits to victims' obligations, that is, they should only be a reason not to criminalize conduct if self-protection is easy, efficient and does not require sacrificing legitimate liberty interests (see for more details Hörnle 2009), and if the conduct does not fall in the core of criminal law. As mentioned above, in the case of a violent attack against the most important rights, I would assume a categorical duty of the state to respond to rights violations, but this should not be extended to all kinds of rights violations.

This idea is unique to legal theory. From a moral perspective, one would focus on the agent's conduct and come to different results. Morally seen, it enhances blameworthiness if offenders exploit others who were oblivious or not very competent in protecting their own interests. The notion of victims' obligations stems exclusively from the victim's role as a citizen. The German Federal Constitutional Court uses the notion of "*Menschenbild des Rechts*" to emphasize citizens' autonomy and responsibility.[10] This is a normative, ascriptive concept, not a social or psychological description of human beings, who will, in real life, only partly live up to the normative expectations of the role described with the term citizen (Hörnle 2015). The German Constitutional Court points to our constitution (the *Grundgesetz*), but the underlying assumptions make sense beyond German positive law. They are fitting for all modern states with legal systems that emphasize citizens' autonomy – the obligation to care for one's own interests in a responsible way can be described as the flipside of autonomy rights.

10 See for instance Entscheidungen des Bundesverfassungsgerichts (BVerfGE), Volume 41, 29, 58; Volume 108, 282, 300.

This idea of victims' obligations obviously needs more elaboration, with regard to the question of how much self-protection can be required and what unacceptable restrictions of liberty would be. The point I want to make here is a limited one. First, these questions tend not to be covered sufficiently in the field of criminal law doctrine and deserve more attention. Secondly, for the area of criminalization theory, such considerations could contribute to creating a denser net of guiding principles beyond a very thin master principle. For illustration, let me mention two examples from recent debates about criminalization. The first example concerns the reform of sexual offenses and the choice between an "only yes means yes"-model and a "no means no"-model. The difference between the two models shows up in cases where neither approval nor rejection was expressed: they would be punishable under the first model, but not under the second. The second example relates to a recent German scandal where a young offender had assembled and posted vast amounts of personal and private information about politicians in the Internet that he collected from a wide range of websites (for instance, photos, chats with family members, mobile phone numbers, etc.). Should such disregard for sexual autonomy and privacy of others lead to criminal punishment? If the question is how much citizens ought to protect their own interests, for both examples one can point to victims' obligations that are easy to fulfill and do not unduly interfere with their liberty. It does not seem unreasonable to demand an expression of disapproval when confronted with another person's sexual advances or to demand the use of a password to protect personal data. Under this premise, criminal norms can be drafted in a way that excludes constellations where victims of a rights violation themselves have neglected to protect themselves in easy ways. Such details of the victim-offender-interaction are not grasped easily with Duff's approach that focuses on public wrongs and the vertical relationship between the polity and the offender.

5. CONCLUDING REMARKS

The main point in my comment is that it would be preferable not to search for one master principle to structure criminalization. If we distinguish three groups of objectionable conduct, one could formulate three master principles:

> Principle 1 (preventive, *pro tanto*): The state has reason to criminalize conduct if it is incompatible with important collective interests that cannot be adequately protected by other means, provided that there are no stronger countervailing reasons.

Principle 2 (responsive, *pro tanto*): The state has reason to criminalize conduct if it violates another person's right to non-intervention (however, there may be countervailing reasons).

Principle 3 (responsive, categorical): If conduct consists in a violent attack that disregards important rights of others, it must be criminalized.

Admittedly, with regard to principles 1 and 2, arguments in support of important collective interests and rights to non-intervention are open to debate. The same holds for weighing them against countervailing reasons. Nevertheless, it seems feasible to develop a framework of principles, including victims' obligations along these lines. Developing sub-principles for each of the three areas could (compared to the more general notion of "the polity's civil order") give somewhat more structure to criminalization decisions.

BIBLIOGRAPHY

Archard, D., 1998: *Sexual Consent, Boulder*, Colorado: Westview Press.

Chiao, V., 2018: *Criminal Law in the Age of the Administrative State*, Oxford: Oxford University Press.

Dubber, M., 2005: "Theories of Crime and Punishment in German Criminal Law", *American Journal of Comparative Law* 53: 679-707.

Duff, A., 2018: *The Realm of Criminal Law*, Oxford: Oxford University Press.

Engländer, A., 2015: "Revitalisierung der materiellen Rechtsgutslehre durch das Verfassungsrecht?", *Zeitschrift für die Gesamte Strafrechtswissenschaft* 127: 616-634.

Hillenkamp, T., 1981: *Vorsatztat und Opferverhalten, Göttingen: Schwartz Verlag.*

Hillenkamp, T., 2017: "Was macht eigentlich die Viktimodogmatik?", *Zeitschrift für die Gesamte Strafrechtswissenschaft* 129: 596-628.

Hörnle, T., 2009: "Die Obliegenheit, sich selbst zu schützen, und ihre Bedeutung für das Strafrecht", *Goltdammer's Archiv für Strafrecht* 156: 626-635.

Hörnle, T., 2012: "Criminalizing Behaviour to Protect Human Dignity", *Criminal Law and Philosophy* 6: 307-325.

Hörnle, T., 2015: "Das Menschenbild des Rechts", in *Anthropologie und Ethik*, eds. J. Heilinger and J. Nida-Rümelin, 97-111, Berlin: de Gruyter.

Hörnle, T., 2019: "The Role of Victims' Rights in Punishment Theory", in *Penal Censure. Engagements Within and Beyond Desert Theory*, eds. A. du Bois-Pedain and A. Bottoms, 207-226, Oxford: Hart.

Hörnle, T. and M. Kremnitzer, 2011: "*Human Dignity as a Protected Interest in Criminal Law*", Israel Law Review 44: 143-167.

Kudlicleh, H., 2015: "Die Relevanz der Rechtsgutstheorie im modernen Verfassungsstaat", *Zeitschrift für die Gesamte Strafrechtswissenschaft* 127: 635-653.

Poscher, R., 2003: *Grundrechte als Abwehrrechte. Reflexive Regelung rechtlich geordneter Freiheit*, Tübingen: Mohr Siebeck.

Von Hirsch, A., and N. Jareborg, 1991: "Gauging Criminal Harm: A Living-Standard Analysis", *Oxford Journal of Legal Studies* 11: 1-38.
Wadle, E., 2001: *Landfrieden, Strafe, Recht*, Berlin: Duncker & Humblot.
Wellman, C., 2017: *Rights Forfeiture and Punishment*, Oxford: Oxford University Press.

DOI: 10.31009/LEAP.2019.V7.13

Should Public Blame replace the Criminal Trial? A comment on R A Duff, *The Realm of Criminal Law*[*]

GUSTAVO A. BEADE
University of Palermo

ABSTRACT

What should we do in cases in which public blame takes an important role? Should we rely on it and forget about Institutional alternatives? Here I present one of Duff's new developments from *The Realm of Criminal Law* that will certainly generate new debates: the right to be prosecuted. I begin with an example followed by an argument that I believe will meet with Duff's approval. My goal in the paper will be to expand the right to be prosecuted, as presented in *The Realm of Criminal Law* and argue that the proper solution will be to defend the exercise of the right to be prosecuted and respond to the accusations.

Keywords: public blame, criminal trial, communicative punishment, right to be prosecuted.

1. INTRODUCTION

During the last ten years I have been inspired by Antony Duff's work and developed many of my own central ideas in reference to this corpus. However there has always been something that concerned me about Duff's ideas on blame and the communicative function of punishment.[1]

I have doubts about Duff's response to particular cases, e.g.: should we have punished X for doing Y? or do we have moral standing to blame X for

* For helpful comments and criticisms I am grateful to Fernando Bracaccini, Roberto Gargarella, Tatjana Hörnle, Santiago Mollis, Ezequiel Monti, Gideon Yaffe, two anonymous reviewers and of course, Antony Duff. I would also like to thank José Luis Marti for organizing the symposium on this book in Barcelona in February 2019. Finally I would like to thank the University of Palermo for the financial support during the writing of the paper.

1 For instance, there are identifiable differences about blame in particular between what he said in Duff (1986) and in Duff (2010).

doing Y? At the same time, Duff's framework has often been characterized in different ways: many legal theorists read him as being close to abolitionism, although Duff is not an abolitionist. Others criticize him for being neither a pure liberal nor a pure republican. Over the past ten years Duff many times criticized abolitionism and developed a particular liberal republican approach. Nevertheless, confusion still surrounds his work.

Here I present one of Duff's new developments from *The Realm of Criminal Law* that will certainly generate new debates: the right to be prosecuted. I begin with an example followed by an argument that I believe will meet with Duff's approval. Ultimately, my goal will be to expand the right to be prosecuted, as presented in *The Realm of Criminal Law*.

2. INCITEMENT TO RAPE

I begin with a particular example which I will call *Mr. Lamb's case*: Mr. Lamb is a musician and well recognized public figure in his country. During a promotional event for his forthcoming record, he surprisingly makes claims such as, "It is a legal atrocity that if a sixteen years girl wants to have sex with you, you can't because it is forbidden," or even more notoriously, "Some women need to be raped to have sex because they are hysterical and feel guilty for not being able to have sex freely," and something like, "What are these so-called 'Rights of Women'? I do not believe in the laws of men, but only in those of nature." Obviously, such statements will generate a public outcry.

Suppose that many citizens begin to protest against him, discussions erupt in social networks, and the media rejects his claims. Public blame and censure against him is unanimous and everyone agrees that Mr. Lamb is wrong. Some of his fans refuse to continue to listen to his music and not a single other musician will defend him in public. After the conference, radio and TV stations decide not to play his music and concert promoters throughout the country cancel his shows. A few days later, Mr. Lamb presents a written statement apologizing for his conduct and explains that he was misunderstood. However, after the negative response from the public he decides to cancel his forthcoming tour to promote the new album and to retire – temporarily – from the music business.

I assume that many legal theorists and philosophers would say that we have censured and blamed Mr. Lamb enough.[2] Many would probably say that we should be satisfied because we live in a polity that rejects these kinds of conduct and forget about the issue, suspending recourse to any

2 See for instance Husak (2010: 433)

sort of legal procedure. I think that blame plays a central role in our relations with other members of our communities and consider that sometimes public blame or critical attitudes e.g., public humiliation, 'lynching' in the mass media, is probably the only kind of social response in a political community.[3] Blame is an expression of the community's disapproval and clearly distinguishes who is the subject of reactive attitudes and who is not: therefore, blame stigmatizes but also acquits. Furthermore, constitutes the foundation on which a certain type of retributive-communicative punishment is based.[4] However, I am unconvinced about relying heavily on public blame. I believe that in this particular case we need to do something more than expressing anger and disapproval.[5] If Mr. Lamb committed an offense (probably an incitement offense), we should prosecute him and put him on trial.

In what follows I will defend the idea that we should prosecute him and call him to account. I tentatively expand Duff's idea of the conditional right to be prosecuted and argue against the idea that public blame generates enough censure for Mr. Lamb.

3. CATEGORICAL AND CONDITIONAL RIGHT TO BE PROSECUTED

According to Duff, to call someone to account for an alleged wrong, however condemnatory that calling to account might be, involves showing that person a certain kind of respect, even concern, as a fellow citizen who, along with everyone else, is bound and protected by the values to which that calling to account appeals (Duff, 2018: 210). To say that we owe it to the wrongdoer to call him to account is to say that he has a right to be thus called to account and, therefore, if the calling to account takes the form of a criminal prosecution, a right to be prosecuted. However, Duff distinguishes between a *conditional right and a categorical right* to be prosecuted.

The *conditional right to be prosecuted* is the right that if we are going to respond to the wrong in a way that enacts or displays our condemnation of it, and that bears adversely on the perpetrator, we should prosecute him rather than subject him to other kinds of an adverse response (Duff, 2018: 210). When he refers to these other responses, Duff is thinking about

3 About the (possible) relationship between public blame and criminal law see Hörnle (2018).

4 I developed this idea in Beade (2019).

5 I would like to thank an anonymous reviewer for encourage me to develop this point.

immediate ostracism or exclusion, or about punishment without a trial, which he describes as lynching. A *categorical right to be prosecuted* would be a right to be prosecuted even if the alternative was simply to have one's wrong ignored. On Duff's view the right to be prosecuted can be *categorical* rather than a merely *conditiona*l right. If to prosecute someone is to address her as a responsible member of the political community, as a participant in this form of civic life, it is an aspect of what we owe to each other as fellow citizens: the 'equal concern and respect' that we owe each other, the recognition of each other as fellow citizens, is displayed in, *inter alia*, our responses to the wrongs that we commit. To fail to respond to another's wrong, when that other is a fellow citizen and the wrong a (non-trivial) civic wrong, is to imply that neither she nor her wrong is worthy of our attention or notice; but that is not how we should address our fellow citizens (Duff 2018: 210). However, a categorical right to be prosecuted or punished is not an absolute right: there might be good reasons not to prosecute, or punish, culpable wrongdoers (for instance, having to do with 'the public interest;' in which case, non-prosecution or non punishment does not violate their rights (Duff, 2018: 211 n 95)).

According to Duff, what we owe to victims and perpetrators is not (simply) to condemn the wrong or the wrongdoer. Condemnation does indeed express a concern for the victim, and for the values which that wrong violated and it at least implies a recognition of the wrongdoer as a responsible agent: but it does not by itself address him as a responsible fellow member of the normative community, because it does not require or invite him to respond to the charge of wrongdoing - to deny the charge, to defend himself, or to accept his guilt. A polity that is quick to exclude wrongdoers might not invest much in a process that calls wrongdoers to answer for their deeds - beyond what was needed to make reasonably sure that they were guilty. But a polity that takes an inclusionary attitude to its members and that still treats them as citizens even when they commit serious wrongs, will be committed to calling them to account for their civic wrongdoing (Duff 2018: 211).

Let me return to Mr. Lamb´s case: he has been censored and his behavior was widely criticized in the media. His statements were rejected by a large part of the polity who believe that there are matters about which we agree: that we - as members of the polity - do not trivialize sexual assaults and the suffering of the victims and that we respect the laws that govern us and punish those crimes. We assume that there are things that we are not willing to accept passively. In addition, the actions taken by some private companies to cancel his shows and stop playing his music are a sign that

these are public matters. In short, it seems that Mr. Lamb received what he deserves and it seems that to call him to account, to prosecute, and to punish him will not add anything but suffering.

I criticize many of the decisions that were taken as a result of Mr. Lamb's claims. I do not think that censoring his music on the media, cancelling his concerts, nor hurting him financially are fair decisions. In fact, they are out of proportion and undeserved because they affect people other than Mr. Lamb. These actions affect his musicians and his staff as well. They lost their jobs and their future seems to be compromised by these decisions. It does not matter how much anger we feel, nor how much we repudiate Mr. Lamb on social networks. The central point is that beyond individual and unconnected responses such as public blame we need a response from someone who represents us as an authorized member of the polity. Someone that can speak in the name of the polity. Moreover, we owe Mr. Lamb an adequate treatment for his conduct. This is where the right to be prosecuted plays a significant role. In Mr. Lambs' case, the right to be prosecuted will be conditional and categorical. Public blame initially results in ostracism and exclusion for Mr. Lamb, but then this kind of blame will be probably ignored by public officials.

The categorical and conditional right to be prosecuted seem to be important categories when we think about prosecution and trial. However, at first sight these look like underdeveloped categories. Taking into account Mr. Lamb's case there are two important questions that need to be answered; For example, why to prosecute or not Mr. Lamb will be the first decision that a polity needs to take when facing similar situations? Or why should Mr. Lamb agree to be prosecuted? In what follows, I will try to answer these two questions.

4. WHY PUBLIC BLAME IS NOT ENOUGH

Although, members of the polity individually expressed themselves and private corporations took decisions that censured Mr. Lamb's claims, the polity as a whole has not said anything about Mr. Lamb. However, how can the polity express itself? Evidently, there are several ways whereby a polity can express itself. I cannot present them all here, but prosecutors should represent the polity as a whole in cases like this. If a prosecutor reports someone or starts a judicial inquiry, she is speaking in the name of the polity. The prosecutor as a member and a representative of the polity will may be able to discuss and refute each of Mr. Lamb's claims and remind Mr. Lamb that the "laws of men" are his laws, and that in this polity we do

not tolerate such assertions. Furthermore, a prosecutor will point out that nature (or the laws of the nature) does not force us to treat women as if they were objects.

However, someone might ask what would be the point of calling him to account and threatening him with a criminal trial (and claim that as his right!) when we have already censured and blamed him. We might be accused of prosecuting Mr. Lamb as a consequence of social pressure or because we are advocates of penal populism. Furthermore, someone could argue that in virtue of the rage that the case has provoked it seems to be undeniable that Mr. Lamb's trial will be seriously contaminated and this could constitute another type of retaliation. Nevertheless, my argument is that, apart from public blame and private responses against Mr. Lamb, if we are advocates of a communicative theory of punishment the polity as a whole must do something. The aim of calling him to account and putting him on trial is to give him an opportunity to explain what he did (and why he did it), realize that he made a mistake and, if he wishes, to repent for what he did and (if possible) to repair the harm he caused.[6] The fact that Mr. Lamb´s statements where rejected and discussed in the polity is not a reason to suspend or left aside the reasons that we have to prosecute him. If Mr. Lamb suffered disproportionate public blame we would not have strong reasons to prevent Mr. Lamb from being prosecuted but to guarantee a proportionate trial.[7]

The offender's repentance and apology after thinking about the censure of the polity in which he lives allows him to recognize that what he did was wrong. Note that the idea of imposing suffering in order to facilitate an apology has met with some criticism.[8] Some people might have doubts about the general idea: why would an offender apologize after committing a crime? Why should Mr. Lamb repent for what he did? My answer would be the following: Mr. Lamb is part of this polity and he probably wants to continue to belong to it and actively participate in it. Repentance allows for this possibility. Under these circumstances the polity is obliged to react positively to the apologies or the repentance of an offender.[9] Let me elaborate further on this idea.

6 According to Michelle Madden Dempsey this is an act that kicks *off a moral dialogue*. See Dempsey (2015).

7 I thank an annonymous reviewer for prompting me to reflect on this issue.

8 See for instance Tadros (2011: 101).

9 To call someone to account also generates certain obligations for the polity. One of them is to be willing to listen to the offender's reasons and to re-integrate him quickly after being punished. Unfortunately I cannot develop this point further here. However I think that is related to the problem of the so-called collateral restrictions. See Hoskins (2016).

In one, perhaps, overrated idea of the criminal trial, Mr. Lamb can give reasons that explain what he did. Moreover, we - the polity - need to listen to him. It is important, when guaranteeing his right to be prosecuted, to understand why he is prosecuted and why what he said constitutes an offense. In J.M. Coetzee´s *Disgrace* a university professor, David Lurie, is accused of having a relationship with a student named Melanie Isaacs and must appear before a special university committee. Lurie is informed of the charges against him, which are based on a statement by Melanie Isaacs and a report that although Melanie did not attend his classes, nor take exams, Lurie gave her a passing grade for the course. Wishing to finish the investigation quickly, Lurie pleads guilty. However, when the committee asks Lurie which of Isaacs' claims he accepts, he responds that he accepts all of Isaacs' statement without giving any reasons. He says that she has no motives to lie. The committee insists that Lurie read Isaacs' statement. He refuses to do so. This situation creates a problem for the committee members who have reservations about ending the matter without delving deeper into what happened. They tell Lurie that the community has the right to know which of her claims he accepts and that it is also necessary for him to know why he is being censured. Lurie stands by his refusal to read her statement. Finally, he is suspended and expelled from the University.

In *Disgrace*, the boundaries of the polity collide with the right of someone to refuse to explain what happened. The polity cannot force citizens to accept or discuss the details of an accusation. A defendant can make use of a specific right to explain what has happened but we cannot put him under a legal obligation to do it. We cannot claim that a polity may demand an explanation, nor therefore maintain that the wrongdoer is under an obligation to explain himself.

Nevertheless, repentance is an act which is important in re-establishing the relationship between the offender and the community. If I am part of a certain community and I make a mistake, I would not like to be excluded from the community. I am part of the community, I belong to it, and I have an interest in continuing to be a member. It is my interest in being part of a community (or a polity) which gives me reasons to recognize my errors and to repent for what I have done. For these reasons I will be able to accept the right to be prosecuted. It is clear that the community also has obligations that derive from the wrongdoer's repentance. Among these, I think, are the obligations to accept repentance and to welcome the person who intends to return to be part of the community. Needless to say that respect for the rule of law will be part of the right to be prosecuted and of the trial. Another prior obligation therefore will be to secure access to the right to be

prosecuted in order to avoid unwanted acts of public lynching.[10]

Duff claims that the trial is a process through which citizens are made to answer to charges by their fellows citizens of civic wrongdoing and to answer for that wrongdoing if the charges are proven: it thus provides an appropriate response to what the polity defines as crimes. The trial constitutes the calling to account that civic wrongdoing requires - a calling that takes the wrongdoing seriously and addresses its agent as a responsible citizen. That trial focuses on that wrongdoing as a public rather than a private matter: the polity calls the alleged perpetrator to account for a wrong that concerns all citizens. However it respects the bounds of private life: what is at issue is whether the defendant is guilty of the particular wrong specified in the indictment; only matters bearing directly on that issue are to figure in the trial (Duff 2018: 214).

Advocates of restorative justice might suggest that there will be better opportunities for a communicative process in a different kind of informal process. However, I am unpersuaded by their arguments. Let me say something about this alternative.

5. WHY NOT THE RESTORATIVE JUSTICE ALTERNATIVE?

Calling to account, as a way of treating each other as responsible agents, might not seem to require a formal criminal process. Duff suggests that the advocates of restorative justice often also make responsibility salient in their accounts of how we should collectively respond to (what we see as) crimes, but they advocate informal processes of mediation rather than a formal process like the criminal trial.[11] What is important, they argue, is to develop less formal structures and procedures through which responsibility can be discussed, negotiated, accepted, and discharged - through which people can take responsibility for their actions and agree on how to discharge that responsibility through reparative actions. They differ, however, over just what those structures and procedures should be, and sometimes they fail to appreciate the significance of the distinctive kind of calling to account that the criminal law can provide. However, according to Duff, a process that is to do justice to victims and to what they have

10 To blame someone means pointing out that that agent did something wrong; blame is conceptually linked to wrongdoing. If blame implies the idea that we are treating the offender as a moral agent, as a member of our moral community we can expect that the she could offer a justification or and excuse that explains what she did. In this sense blame could be perceive as a form of dialogue. As I understand it public lynching is far from being considered as a dialogue but to an extreme way of accusation.

11 See e.g., Walgrave (2008) and O´Mahony and Doak (2017).

suffered must not merely aim at calling the perpetrator of harm to account or at bringing him to recognize his responsibility for that harm, but must also call a wrongdoer to account for the wrong that he has done.

Duff argues that the advocates of restorative justice often prefer an informal negotiation between the parties (who are most directly involved in the case) rather than a formal determination by a court. What matters, they argue, is that people should discuss their concerns, explain their actions and reactions, freely and openly. They should be able to come to accept their own responsibilities and to recognize that the responsibility for a harm might not be properly allocated to just one person. We might do better to try to resolve our problems through such informal processes rather than by appealing to the formal (and coercive) apparatus of the criminal law. But, according to Duff, such informal processes will not always be adequate. Duff proposes that some kinds of wrong should be treated as public matters which concern us all as citizens. If these wrongs violate our collective values, that gives us reason to look for a formal, public, collective response (Duff 2018: 212).

Duff claims that we should treat as typically public wrongs those acts that require categorical recognition and condemnation rather than the kinds of nuanced negotiation that a 'conflict' oriented process is likely to involve. It is difficult to think about institutional alternatives in which Mr. Lamb responds without being forced to do so and having the chance of offering a justification, an explanation, expressing regret, offering compensation, or taking part in a kind of dialogue. Mr. Lamb's offense is one that has no particular victim but the polity. It directly violates our values and, as Duff advocates, it is a kind of wrong that should be treated as a public matter. When we commit wrongs that count as 'public,' we must be willing to answer for them publically (Duff 2018: 213). If we propose an alternative to this we will be favoring the privatization of the conflict, one in which the polity will not take part. Moreover, if our starting point was public blame, we will need to think of a public process like a trial. In short, if we are to call wrongdoers to account, we must establish that they are wrongdoers first. And if we are to do that in a way that treats them as responsible citizens, we should do it through a process that invites them to answer the accusation of wrongdoing and which protects them against unwarranted judgments of guilt (Duff 2018: 214).

6. PROSECUTING INCITEMENT IN A LIBERAL REPUBLIC?

There is one issue, which is not central to the argument, but which I need to tackle. Mr. Lamb will be accused of committing an offence of incitement

(or encouragement), i.e. to persuade or encourage others to commit a crime.[12] My argument is that we have a particular offense of incitement and we need to decide if we are going to prosecute Mr. Lamb or not. However, in normative terms, the question is: should we prosecute citizens for incitement offenses in a liberal republic?

Probably influenced by liberalism I used to have reservations about this kind of inchoate offenses. I used to think that to criminalize this kind of conduct overestimates the influence certain people have over others and underestimates the critical faculties of the recipients of those messages. If we accept that we live in a democratic polity composed of autonomous citizens, i.e., agents that can make their own decisions, the criminalization of this type of behavior appears to be insulting. Autonomous citizens are able to make their own decisions and have the capability to avoid committing wrongs.[13] Many of the prosecutions for incitement are related to the use of drugs or the promotion of certain substances. However, I am skeptical about the influence that defenders of certain drugs could have on autonomous citizens. Such defenses may possibly influence people under a certain age but they are unlikely to influence adults. We have, as adult citizens, different visions and moral judgments about the good and the bad. Our own decisions make us responsible, rather than any "suggestions" by others. Thus, it is difficult to understand why we should accept incitement offenses in a liberal republic. However, let me try to present an answer based on Duff's account.

According to Duff, the criminal law's primary purpose is the provision of an appropriate response to those kinds of moral wrong that constitutes public wrongs. Taking this into account, Duff argues that one function of the substantive criminal law is simple declarative: the polity has reason to declare (in some suitable formal way) the norms that its citizen have determined should guide their civic conduct - its public norms. Duff's modest claim is that an institution can have good reasons to make a formal declaration of the norms that are to guide its members' conduct beyond the desire to ensure compliance with those norms by threatening sanctions against those who violate them. These are norms whose breach will render perpetrators liable to a formal public response (Duff 2018: 206-207). However, even without the prospect of punishment, such formal declarations of what constitutes public wrongs can have some preventive effects: they can remind us of how we should or should not conduct ourselves and they can warn us that if we fall short in our conduct we will face the prospect of being prosecuted. Furthermore, even without the

12 More details in Ashworth & Horder (2013: 476 ff).

13 See Husak (1992).

prospect of punishment following a conviction, the prospect of prosecution can have a dissuasive effect, since prosecution itself is a burdensome business and conviction is, for most people, unwelcome and unpleasant.

There are differences between drug cases and Mr. Lamb's case. In particular, Mr. Lamb's case looks like an "Incitement to Hatred" offense. [14]Duff argues that to hate a fellow citizen is thus already a lack of civic virtue, i.e. a lack of concern and respect for others. However, if I act simply as an individual who hates another individual, nothing of civic importance might hang on my action. If my conduct has a collective aspect and is directed against members of an already vulnerable group, its character changes: if my hatred is directed against a group, treating them as contemptible or hateful, as disqualifying members from proper citizenship, then such conduct becomes genuinely threatening (Duff 2018: 200-201). In this case, Mr. Lamb, fails to recognize others (women in particular) as equals and as fellow citizens and this makes a big difference in comparison to drug cases.

7. FINAL REMARKS

Remain calm, my abolitionist friends: there is no risk of imprisonment for Mr. Lamb. He will exercise his right to be prosecuted and will be able respond to the accusations. To call him to account will allow him to limit disproportionate public blame and give him the chance to defend himself against the polity's representative: the prosecutor. Our common law, addressing us in our collective voice, can do much more for our civic ideals and this is part of Duff's proposal in *The Realm of Criminal Law*.

BIBLIOGRAPHY

Ashworth, A. & Horder, J., 2013: *Principles of Criminal Law*, Oxford: Oxford University Press

Beade, G.A. 2019: "Who Can Blame Whom? Moral Standing to Blame and Punish Deprived Citizens," *Criminal Law and Philosophy* 13:271

Dempsey, M.M. 2015: "Prosecuting Violence Against Women: Toward a "Merits-Based" Approach to Evidential Sufficiency," *Revista Jurídica de la Universidad de Palermo* 14:241

Duff, R.A., 1986, *Trials and Punishment*, Cambridge: Cambridge University Press

—2010, "Blame, Moral Standing and the Legitimacy of the Criminal Trial," 23(2) *Ratio*, 123–140

14 See for instance the German Criminal Code (StGB § 130 (1)) which defines "Incitement to Hatred" as 'assault the human dignity of others by insulting, maliciously maligning or defaming segments of the population'. See also Hörnle (2012).

Duff, R.A., 2018: *The Realm of Criminal Law, Oxford*: Oxford University Press
Husak, D. 1992: *Drugs and Rights, Cambridge*: Cambridge University Press
—"Already Punished Enough" in *The Philosophy of Criminal Law, Selected Essays Oxford*: Oxford University Press
Hörnle, T. 2012: "Criminalizing Behaviour to Protect Human Dignity," *Criminal Law and Philosophy* 6: 307
—2018: "#MeToo - Implications for Criminal Law?" *Bergen Journal of Criminal Law and Criminal Justice*, Volume 6, Issue 2, p. 115-135
Hoskins, Z. 2016: "Collateral Restrictions" in Flanders Ch. and Hoskins Z., *The New Philosophy of Criminal Law*, London: Rowman & Littlefield
O´Mahony, D. and Doak: J. 2017, *Reimagining Restorative Justice*, Oxford: Hart Publishing
Tadros, V. 2011: *The Ends of Harm: The Moral Foundations of Criminal Law*, Oxford: Oxford University Press
Walgrave, L., 2008: *Restorative Justice, Self-Interest and Responsible Citizenship*, Cullompton: Willan Publishing

DOI: 10.31009/LEAP.2019.V7.014

Criminal Law, Civil Order and Public Wrongs*

R. A. DUFF
University of Stirling

ABSTRACT

This is a response to five critiques of my 2018 book *The Realm of Criminal Law*, by Michelle Dempsey, Kimberley Brownlee, Roberto Gargarella, Tatjana Hörnle, and Gustavo Beade. Topics covered include the idea of a "public wrong"; the usefulness of the analogy I draw between criminal law and codes of professional ethics; how criminal law can function in polities characterized by deep cultural and normative differences; what scope there is for thick principles of criminalization; and whether offenders have a "right to be prosecuted".

Keywords: pCriminal law, public wrongs, professional ethics, multiculturalism, principles of criminalization, criminal trials, a right to be prosecuted.

1. INTRODUCTION

I would like to thank the other contributors to this symposium for their critically constructive responses to my book (Duff 2018): I cannot do justice to all their comments here, but will tackle (what I take to be) their most significant criticisms. They certainly expose some of the ways in which I failed to make my claims and arguments clear enough (to myself as well as to others), and they bring out some of the ways in which the book's central themes need to be further developed; but I hope to show that they do not threaten the central ideas that I tried to explain and defend. In what follows I first give a brief account of the book's aims and main themes (s. 2), and then respond to each of the commentators: to Michelle Dempsey, who

* I am very grateful to José Luis Martí, Roberto Gargarella, and Gustavo Beade for arranging and editing this symposium, and for organizing the workshop at Universitat Pompeu Fabra at which drafts of the papers were discussed; and to the participants in that most enjoyable workshop.

offers an alternative account of the central idea of public wrongs (s. 3); to Kimberley Brownlee, who criticizes the use that I make of an analogy between criminal law and codes of professional ethics (s. 4); to Roberto Gargarella, who argues that my account of public wrongs cannot be sustained in the context of radically plural societies (s. 5); to Tatjana Hörnle, who argues that we should replace my single, thin, principle of criminalization by a more pluralist account grounded in three distinct principles (s. 6); and to Gustavo Beade, who develops the suggestion that public wrongdoers have a "right to be prosecuted" (s. 7).

2. THE REALM OF CRIMINAL LAW

The book emerged from a four-year research project on *Criminalization*, on which I worked with four colleagues: Lindsay Farmer, Sandra Marshall, Massimo Renzo, and Victor Tadros. The purpose of the project was, as we rashly described it, "to develop a normative theory of criminalization: an account of the principles and values that should guide decisions about what to criminalize",[1] and that was also my goal when I started work on the book. In one way, the book achieved that aim, since it offers a "public wrongs" theory of criminalization, whose central principle is:

> A. We have reason to criminalize a type of conduct if, and only if, it constitutes a public wrong.
>
> B. A type of conduct constitutes a public wrong if, and only if, it violates the polity's civil order.[2]

In two ways, however, this will disappoint those looking for a developed, substantive theory of criminalization. First, the principle is extremely thin, since it leaves most of the substantive work, in determining just what kinds of conduct we have good reason to criminalize, still to be done. For until we have given content to the ideas of "civil order" and "public wrongs", which is itself a major normative undertaking, we cannot begin to work out this theory's implications; and even when we have given those ideas substantive content, we will have worked out only what we have reason to criminalize, not what we should criminalize, all things considered. Second, most of the book is taken up, not with exploring the implications of this

1 See Duff et al (2014: 1), which describes the project, and the other publications that it produced—and explains why it did not produce such a theory. I should record here my enormous debt to my colleagues on the project, as well as to the UK Arts and Humanities Research Council, which funded the project.

2 This principle is formulated in Duff (2018: 232, 275, 277), and explained in chs 6-7 of the book.

principle, but with more preliminary discussions of the way in which, the grounds from which, we can arrive at this master principle of criminalization. I can best explain the aims of the book by responding to these two grounds for disappointment.

If we are to decide what kinds of conduct we have reason to criminalize, we must ask what counts as a reason to criminalize; to decide that, we must ask what the purpose of criminalization could be, which is to ask what the purpose of criminal law could be. To answer that question, we must examine the role that criminal law can play within the institutional framework of a political community: for criminal law is, and must therefore be theorized as, part of the political structure of a polity.[3] If we are to tackle that task, however, we must first have some idea of what criminal law is—what distinguishes it, as a particular kind of institutional practice, from other institutions? Accordingly, the book begins with an account of criminal law as a practice that defines a set of wrongs whose perpetrators (actual or alleged) are to be called to formal public account through a criminal process, and to be formally censured, and liable to punishment, if proved guilty. This is not intended to be a neutral analysis of "the concept of criminal law": rather, it offers a rational reconstruction of criminal law as an institution with a distinctive character, of which we can usefully ask what role (if any) it should play in a political community. Theorists of criminal law often focus on criminal punishment as its central characteristic, or as the end point towards which the substantive criminal law and the criminal process are directed: by contrast, I suggest that whilst criminal punishment is indeed a salient, and normatively problematic, feature of our systems of criminal law, a plausible justificatory rational reconstruction of criminal law need not, and should not, make punishment thus central; we can identify purposes and values in the criminal law that do not depend on punishment.

The claim that criminal law is essentially concerned with wrongs obviously implicates some version of legal moralism. Chapter 2 therefore discusses different species of legal moralism. It provides a defense of "negative legal moralism"—the claim that we may not criminalize conduct that is not morally wrongful prior to its criminalization; it distinguishes the version of "positive legal moralism" that I defend, according to which only "public" wrongs can give us reason to criminalize, from more familiar, more ambitious, versions, according to which we have reason to criminalize

3 My focus in the book is almost entirely on the domestic criminal law of nation states (though I say a little about international criminal law in ch. 3.1-2): whatever the future of the nation state might be, this is the context in which criminal law is most fully developed.

every type of morally wrongful conduct;[4] and it provides an initial explanation of the idea of a "public" wrong by discussing the codes of ethics that professional bodies draw up—codes that identify the kinds of wrong that count as "public" for the profession in the sense that they fall within the profession's distinctive practice. There is, I argue, a useful analogy between the criminal law, as the code of a political community, and a code of professional ethics: from which it follows that an account of public wrongs in relation to criminal law must depend on an account of the public realm of the polity. Those drawing up a code of medical ethics must begin with a conception of what medicine is as a distinctive practice—what kinds of activity medicine involves, what aims and values should guide those activities: only then can they work out what kinds of conduct fall within its scope, and then what kinds of conduct must count as wrongs that concern the medical profession—as wrongs that count in that context as "public" wrongs. So too, if we are to draft a criminal code, and decide what kinds of conduct it ought to cover, we must begin with an account of the distinctive kind of practice whose code this is to be, of the aims and values by which that practice is (supposed to be) structured, and of the kinds of conduct that fall within its scope. That practice is the practice of civic life—the practice of living together as members of a political community: the "public" realm, the realm within which criminal law operates and within which we can identify the kinds of public wrong with which criminal law is concerned, is the realm of civic life. If we are to work towards a theory of criminalization, we must therefore begin with an account of the distinctive practice of civic life: only then can we work out what kinds of conduct could count, from the perspective of the polity and its criminal law, as "public" wrongs.

Chapter 3 embarks on this task by looking first at the identity of those who participate in the practice of civic life—those whose law the criminal law is supposed to be. It argues that citizens, as members of the polity whose law it is, are the criminal law's primary addressees: it is to them that the law paradigmatically speaks, in terms of values that are meant to be their values; it is to their fellow citizens that they must answer for the public wrongs they commit; and in democratic polities it is they who "own" the criminal law—it is their law. Questions must then arise: about the relationship between the domestic criminal laws of nation states, as the laws of their citizens, and international and transnational criminal law (questions discussed only briefly in this book); and about the status of those who are not citizens of the polity but are still supposedly bound by its

4 Michael Moore is the best known contemporary proponent of this more ambitious kind of legal moralism: see e.g. Moore 1997: chs 1, 16, 18.

law, or of those who dissent from the polity's values as reflected in its criminal law (questions that I discuss in more detail, to show how an account that, like mine, makes citizenship central can still do justice to non-citizens, who are to be treated as guests, and to dissenters).

Chapters 4-5 then provide an account of the kind of practice in which citizens are engaged: the "public realm" that constitutes their life together as citizens, in which criminal law operates as a type of public law.[5] The concept of "civil order" is central to an understanding of this public realm (see Farmer 2016): a polity's civil order consists in the normative ordering of its civic life —of its existence as a polity. That normative ordering is structured by the set of goals and values through which the polity constitutes itself as a political community: it can be partly defined in a written constitution, as a formal statement of the fundamental goals towards which the polity's collective activities are oriented, and the values by which those activities are to be governed; but it is also implicit in the polity's institutions of government and its citizens' shared understandings of their civic life. Such a conception of civil order, as the ordering of the polity's public realm, depends on a normative distinction between "public" and "private" realms: a polity's civil order constitutes its public realm, in relation to which other aspects of its citizens' lives are "private"; we must therefore attend to different distinctions between "public" and "private" that can be drawn. We must also attend to the preconditions of civil order: what kinds of agreement, what kinds of shared understanding, among the citizens are necessary if civil order is to be possible, and what can those who share such understandings say to those who do not? But given a viable conception of a polity's civil order, and thus of its public realm, we can make better sense of the idea of a public wrong, as a kind of wrong that falls within that public realm, and that violates that civil order: such wrongs are in principle the business of the polity's criminal law, as wrongs to which a formal, public response is appropriate (or even necessary).

Chapter 5 puts more substantial flesh on the relatively formal account of civil order provided in Chapter 4, by sketching the central aspects of the civil order of a particular kind of polity—a liberal republic of free and equal citizens: it draws on the republican tradition of political thought (and on a liberal kind of communitarianism, or communitarian kind of liberalism) to portray the kind of polity that we can plausibly aspire to create. We can then see the role that criminal law could play in such a polity as helping to sustain, but also as partly constituting, its civil order. For once we

5 The slogan that criminal law is public law has become prominent recently, as part of the reaction against what its proponents see as unduly moralised conceptions of criminal law: see e.g. Thorburn 2010, Chiao 2019.

understand the idea of a public wrong, as a wrong that violates an aspect of civil order, we can see criminal law as an appropriate way of marking, and responding to, such wrongs. Its central role is to declare some of the central norms of that civil order, as norms that define what kinds of conduct citizens are entitled to expect from each other (and from the polity); and then to provide the process through which those who are accused of violating these norms can be called to formal, public account by their fellow citizens: in doing so, it gives institutional form to the polity's commitment to its defining values, and to its appropriately respectful concern for all its citizens—including both victims and offenders. (To say that the criminal law's central role is to declare such norms and to provide for those who violate them to be called to public account is to deny that punishment is the primary purpose of criminal law: but given the oppressive salience of punishment in our existing systems of law, I also say something about the role of punishment in a liberal republic's criminal law.)

Given this account of the role of criminal law in the civic life of a democratic republic, I turn, in Chapter 6, to criminalization: in particular, to the question of whether a normative theory of criminalization should offer a "master principle" by which our deliberations about criminalization could be guided. Different kinds of master principle are distinguished; some familiar master principles (notably the harm principles) are critically discussed. One central conclusion is that any plausible master principle will be very thin: it will provide, that is, not substantive criteria for criminalization, but formal criteria that leave most of the substantive normative work in deciding what kinds of conduct are to be criminalized still to be done. This is true of the master principle that emerges in the book: we have good reason to criminalize a type of conduct if and only it constitutes a public wrong;[6] and it constitutes a public wrong if and only if it violates the polity's civil order. This principle is very thin, in that it leaves the normative questions of how the polity's civil order is to be constituted, and of what kinds of conduct should be taken to constitute violations of that order, still to be decided (through the process of public deliberation by which a democratic republic forms its criminal law). But we should not expect to be able to identify a master principle that is both thick and plausible; and the thin principle that I propose is still fruitful in identifying the kinds of consideration that are relevant to debates about criminalization, and thus in showing how deliberations about criminalization ought to be structured—what kinds of claim and argument must be made and

6 The most a master principle can plausibly offer is an account of what gives us good reason to criminalize, not of what we should criminalize all things considered.

defended if we are to be justified in criminalizing a type of conduct.

Chapter 7 seeks to make good on that claim for a "public wrongs" master principle, but only after an essential qualification. Criminalization is just one among a range of possible responses to public wrongs, distinguished from other responses by its concentration on the wrongfulness of the criminalized conduct and on calling perpetrators to public, censorial account. So the question whether we have good reason to criminalize a type of conduct is the question whether we have good reason to criminalize it *rather than* to respond in some other way; and a master principle of criminalization must help us to make that kind of "rather than" decision. To illustrate the work that the "public wrongs" principle can do, I briefly discuss examples of three kinds of criminal offence. First, I discuss "*mala in se*", consisting in conduct that is held to be wrongful prior to its legal regulation. These are in one way the most straightforward kinds of crime to understand, but are in two respects somewhat complicated: for the criminalization process is a process not simply of trying to capture some pre-existing moral wrong, but of constructing a civic conception of the wrong as a public wrong; and, partly for this reason, the distinction between "*mala in se*" and "*mala prohibita*" (the latter being understood to consist in conduct that might be wrongful only because it is legally prohibited), whilst still important, is less clear cut than many think it is (or should be). Second, I discuss different species of "*malum prohibitum*", involving different ways in which violations of pre-criminal legal regulations can constitute criminalizable wrongs: the distinction between regulation and criminalization is important here, as are the different ways in which regulatory violations can be wrongful (even when the conduct does not cause the kind of mischief at which the law is aimed). Third, I discuss examples of "pre-emptive" offences, which criminalize conduct that does not itself involve the mischief at which the law is ultimately aimed, but is in some way preparatory to, or increases the risk of, that mischief: one way in which our criminal laws have been expanding, especially in relation to terrorism, has been in the creation of wider kinds of pre-emptive offence; it is important to see how far such offences can be justified.

These discussions of different kinds of criminal offence are intended to be illustrative, rather than exhaustive: they illustrate the kinds of argument and deliberation that will be needed if we take a "public wrongs" principle, of the kind I have argued for, as our guide to criminalization. Those arguments and deliberations will have different contents in different polities, reflecting as they will different conceptions of a polity's civil order, and I have not tried to argue in this book for one particular such conception:

which is not to say, as we will see, that I advocate a radical relativism according to which a polity's conception of its own civil order, and the criminal laws that it creates to sustain that order, are beyond criticism or argument; but such criticisms require an engagement with substantive political theory that goes beyond the scope of this book. I have indicated the kind of (republican, liberally communitarian) political theory that I favor, and I have sketched some of its implications for civil order and for the criminal law; but I have not tried to justify, or indeed to explicate in detail, that theory, or to defend it against others—that would be a task for another book (and this book was already long enough, if not too long).

I turn now to engage with my commentators and to respond, if not to all their suggestions and criticisms, at least to the most significant of them.

3. PUBLIC WRONGS AND PUBLIC REALMS

I begin with Michelle Dempsey's discussion of how we should understand the idea of the (or a) public realm, which is central to my account of crimes as public wrongs (Dempsey 2020). There is much about which Dempsey and I agree: in particular, that criminal law should be understood as a communicative practice in which wrongdoers are held to account, and which serves not merely to sustain, but also partly to constitute the civil order, the character, of the political community (Dempsey 2020: 158-9; Duff 2018: ch. 5.4-6). Our main disagreement is about how to understand the idea of a "public" wrong in this context.

On Dempsey's account, what makes wrongs more, or less, public is the way in which the perpetrator is called to account: a wrong is more "public" to the extent that the wrongdoer is or should be called to account more publicly; and a calling to account is more public insofar as it done, observably, by representatives of a larger group. Thus at one end of the "private"-"public" spectrum we have "comprehensively private" callings to account done by the direct victim on her own behalf, in private; and at the other end we find "comprehensively public" callings to account in public by representatives of "the entire political community" (Dempsey 2020: 165). This is certainly a possible account of the idea of a "public wrong", and Dempsey offers a useful analysis of what it involves; but it is not the (admittedly somewhat stipulative) conception of the "public" with which I work.

For Dempsey publicness is a univocal, scalar matter of how "public" a calling to account is; for me, it is a practice-relative matter that does not permit such a scalar analysis. A wrong, on my account, is a "public" wrong

if it is one that properly concerns "the public"; and "the public" consists in the participants in the relevant practice or community. Thus in the context of a political community, a wrong counts as "public" if, and only if, it properly concerns the members of that community, in virtue of its bearing on the practice of living together as a political community in which they are engaged; in the context of a golf club, by contrast, a wrong is "public" if and only if it properly concerns the members of the club, in virtue of the practice—playing golf—in which they are engaged together; in the context of an academic institution, a wrong is "public" if and only if it properly concerns the participants in that institution, as bearing on the scholarly, educational practice in which they are engaged. Two implications follow from this; both concern matters on which Dempsey and I disagree.

First, a wrong might be public in the context of one practice, but not in the context of other practices in which the wrongdoer also participates. If I cheat in a golf match, that is a public matter as far as my fellow golfers (fellow members of the club, or golfers anywhere, if we see golf as a practice in which all golfers participate) are concerned. It is their business: they have the standing to call me to account for it; I cannot rebut their challenge by insisting that it is a private matter. However, it is not a public matter in the context of the university philosophy department to which I belong: my academic colleagues, qua colleagues, lack the standing to call me to account for it, since it is no part of, has no bearing on, the practice in which we are mutually engaged; if they challenge me, I can properly tell them that it is none of their business—that I do not answer to *them* for my conduct on the golf course.[7] Similarly, my lackadaisical discharge of my academic duties, my failure to mark my students' essays on time, are public matters in the context of my department and my university, but private matters in the context of my golf club: I do not answer for them to my fellow golfers.

Second, to determine whether a wrong should count as a public wrong in the context of a particular practice, we need an idea of what that practice is: what are its distinctive goals and activities; what values structure those activities? Indeed, we can often identify a wrong as a wrong, or as a particular kind of wrong, only by attending to the practice within which it is a wrong: we understand what counts as cheating at golf, and why it matters, by grasping the character of golf as a distinctive kind of practice. That is part of what I meant by arguing that we must start with "the

7 More precisely, I do not answer to them *qua* colleagues: my cheating is not something that can properly be brought up at a departmental meeting, or made the subject of a disciplinary charge. But if a colleague is also a friend, or a member of the same golf club, she can *in that context* challenge me about my cheating.

public"—with an account of the "public realm", the distinctive aims and activities, of the practice within which we are operating: for only then can we identify the wrongs that should count as public within that practice. But "Start with the public" becomes a more important slogan when we imagine ourselves not as agents acting within the framework of an existing, fully formed practice, but as lawmakers—for instance as people charged with the task of formulating the golf club's rules of conduct, or a university's code of ethics. If that is our task, it would be misguided to start with a list of all the wrongs that can be committed, and then ask which of them should figure in the code. We must begin with a conception of the practice's "public realm"—of what kinds of conduct fall within the practice; only then can we identify kinds of *mis*conduct that should concern participants in the practice. Thus the slogan "Start with the public" can also be expressed by saying that wrongdoing, misconduct, presupposes conduct, which presupposes a public realm. A code of ethics, whether for a golf club or a university, or any other social practice, deals with misconduct—not misconduct in general, but misconduct in activities that belong to the practice ("professional misconduct" is misconduct in one's professional activities): to identify the kinds of misconduct with which our code should deal we must therefore first identify the kinds of conduct or activity that belong to that practice. That is why, if we imagine ourselves as tasked not to draw up a code of ethics for an existing practice, but to construct a constitution for a new or newly organized practice, we will not, and cannot, begin with a code of ethics: we must begin by formulating a conception of the practice, of its aims and guiding values, and of the kinds of activity that fall within it; only then can we turn to determining what should count as censurable misconduct in that context. We begin, that is, with a conception of the practice's *res publica*—its public realm; we are then in a position to identify kinds of wrongdoing that fall within that realm and therefore count as "public" in that context.

As with golf clubs and universities, and other practices and professions, so too (I argue) with political communities and their criminal laws (hence the discussion of codes of ethics in ch. 2). The members of a political community are engaged together in a distinctive practice, of living together as citizens; their criminal law will deal with kinds of wrongdoing that count as "public" in the context of that practice in virtue of their impact on it—on the polity's "civil order"; to identify those kinds of wrong, we must therefore first have some conception of that civil order—of the polity's constitution as a distinctive practice. As analytical theorists or as sociological inquirers, we can understand a polity's criminal law by seeing it in relation to that polity's conception of its civil order, and recognize that different polities are structured by different conceptions of civil order

(they have different constitutions; hence the discussion of constitutions in ch. 4). As normative theorists, or as active participants in the life of our own polity, we will argue about how this or any polity *should* conceive its civil order; only on that basis will we be able to argue sensibly about the proper scope of its criminal law—about what it should count as public wrongs, and about how it should respond to them (hence the discussion of liberal republicanism in ch. 5).

We can now see how, although Dempsey and I agree on the importance of "Alternative Sites of Public Accountability (ASPAs)" (Dempsey 2020: 167), we disagree about how they should be understood. On her view, they are different ways of achieving the same end: that wrongdoers are called to account. For any given (kind of) wrong, we can ask which ASPA is most suitable for calling the wrongdoer to account, and we might seek a sensible division of labour between different sites. Thus a golf cheat might be most suitably called to account not by the criminal law, or by his employer, but by fellow members of his golf club (or perhaps by the sport's central governing body): for the club is most likely to be best placed to call him to account; and in doing so it can also help to constitute itself as a valuable institution, which seeks to realise the values of fair play and honesty in its practices. It might be so obvious that given the nature of the wrong and the available ASPAs, the golfer should be called to account by his club, that legislators formulating the polity's criminal law would not even consider cheating at golf (or at games generally) as a candidate for criminalization (unless it involved some financial gain). It remains true, however, on Dempsey's account, that there is reason in principle to criminalize cheating at golf, even though there are obviously better reasons not to do so: for this would help to realise the good of wrongdoers being called to account for their wrongs. This might not be a reason to which legislators should attend in their deliberations: but it figures in "a full philosophical account of the reasons that … bear on the justifiability" of their decisions about the scope of the criminal law (Dempsey 2020: 171); and it must then surely figure in the thinking of self-reflective legislators who wonder what kinds of reason should figure in their first-order deliberations.

A crucial point, which highlights the difference between Dempsey and me, is that on her account, "in deciding whether any given calling to account is justified" (whether it would, for instance, be right for a legislature to have golf cheats called to account by the criminal law), "we should consider whether and how the wrongdoer will otherwise be called to account"; and one reason in favour of criminalization would be that this "is the only way the wrongdoer will be held to account" (Dempsey 2020: 162). In designing our institutions, we must attend to *all* those values that might

be realised through the institution's activities; and one value is that wrongdoers are called to account for their wrongdoing. On Dempsey's view, "if a person can realise a value through her action, then normally she will have a reason to act so as to realise that value":[8] since criminalizing cheating at games could help realise the value of wrongdoers being held to account, a legislature has reason to do so—although no doubt it has much better reasons not to do so, and not even to think of doing so.

I have two main responses to this line of thought. First, the relevant value here is not that of wrongdoers being held to account, but that of wrongdoers being held to account *by those to whom they are answerable*; and they are answerable to, and only to, fellow members of the normative practice in which the wrong is done. If a passing stranger seeks to call a golf cheat to account for her cheating, he might succeed in calling her to a kind of account: but he does not achieve the proper end that would be achieved if she was held to account by her fellow members, to whom she is properly answerable for her golfing (mis)conduct[9] We might miss this point if we think only of more serious kinds of moral wrong, since we might think that those who commit such wrongs are answerable to "the moral community", which includes all moral agents; but we live in other communities than the moral community, with different and more limited scopes.

Second, whether we are acting within some existing practice, or trying to construct one, it is wrong to suggest that we have reason (even in principle) to try so to design our practice as to realise each of all the many possible values—as if those founding a golf club have reason, in principle, to take the promotion of aesthetic appreciation as an aim, and to formulate the club's constitution accordingly (as the normal correspondence thesis implies). It matters that what they are doing has some value, and they might be accused of normative myopia if they attach undue importance to the value of what they are doing: but even if we look for "a full philosophical account of the reasons that actually do bear on the justifiability" of the practice that they construct, that account would not plausibly make reference to all the goods that they could help to realise in constructing the practice.

To which Dempsey might with justice reply that, even if there is some

8 Dempsey 2009: 83 (on "the normal correspondence thesis"); see also Gardner 2007: 62-3, on the "agent-neutralist" view of reasons. This is one part of an ongoing argument about the nature of practical reason between me and theorists such as Dempsey and Gardner: see Dempsey 2011, Gardner 2011, Duff 2013.

9 Analogously, if a vigilante mob lock up an unconvicted offender for the length of time for which he would have been imprisoned if convicted, they might bring about roughly the same material impact that his formal punishment would have brought about; but they do not achieve the end that would justify that punishment—that he be held to punitive account by those to whom he is answerable, his fellow citizens.

analogy between professions (or golf clubs) and polities, it breaks down just when I need it to do serious work. It might be ridiculous for those creating a golf club, or a profession's constitution, to attend to all the goods that they could in principle realize: but it is surely far from ridiculous, indeed it would surely be entirely proper, for the founders of a polity to think in such terms; and, in particular, if they take seriously the importance of wrongdoers being called to account, they should think about how they can help to realise that good. The first point noted above still remains: that if what matters is that wrongdoers be called to account to and by those to whom they are answerable, it is not clear that a polity could help to realise that end by criminalizing every kind of wrong; for that would presume that wrongdoers of all kinds are answerable, at least in principle, to their fellow citizens—but that is a central point at issue between me and Dempsey. However, why should a polity not take the collective view that, first, every kind of moral wrongdoing in principle falls within its purview, so that those who commit any kind of wrong are in principle answerable to the polity; but, second, that there are many wrongs with which it should not deal directly, since they are better delegated to one of the more specific practices that are found within the polity? Thus, for instance, wrongs committed in the course of medical practice are in principle, like all wrongs, criminalizable as the polity's business; but many of them can be better dealt with through the medical profession's own disciplinary procedures.[10] Now a polity *could* coherently take this view of the matter, just as it could take Moore's view of the proper aim of criminal law, as being to inflict retribution on culpable wrongdoers;[11] and Dempsey is ready to "plead guilty" to offering "a modified version of "Moorean legal moralism" (2020: 172). I think that is indeed what she offers—although the modification is crucial, since it replaces the infliction of retributive suffering by the more plausible goal of calling wrongdoers to account; and I do not offer anything resembling a refutation of either Moorean or Dempseyan legal moralism. But what I do argue against Moore, and would now argue against Dempsey, is that to advocate such legal moralism is to take a particular view of the proper concerns of a polity: that it has an in principle proper interest in every aspect of its citizens' lives, since it has such an in principle interest in all the wrongs that they might commit. Both Moore and Dempsey would then set tight limits on the kinds of wrong in which a liberal polity should actually take an interest; but my claim is that a truly liberal polity would not need to set such limits, because its

10 And I agree that this kind of delegation is sometimes possib profession's own disciplinary procedures le, and desirable: see Duff 2018: 89-91.

11 For Moore's view see e.g. Moore 1997: ch. 1; for detailed critical discussion see Duff 2018: ch. 2.

initial conception of itself —of the practice of living together as members of the polity, of the aspects of its members' lives and activities that fall within that practice—would not include every kind of wrong. To include every kind of wrong, it would need to include every kind of conduct (for misconduct presupposes conduct): but a central defining feature of liberal as against totalitarian polities is that they precisely reject such a view.

At times I wonder whether the difference between Dempsey and me is substantial: for we agree that a wide range of wrongs (those that a liberal polity would not criminalize) are not in the end the polity's business, and seem to differ only about the stage of deliberation at which "it's not the polity's business" comes into the picture. However, first, that difference matters, if our aim is to understand the proper structure of arguments about criminalization. Second, her account of the idea of the "public" seems to me not to do justice to what we might call the practice-relativity of the public, or to allow a sufficiently basic role to the thought that some wrongs are, from the polity's point of view, categorically "private".

4. PROFESSIONAL ETHICS AND CRIMINAL LAW

My account of public wrongs and of the criminal law's role in a political community depends in part on an analogy between criminal law and professional codes of ethics (Duff 2018: ch. 2.7-9). This analogy is the focus of Kimberley Brownlee's paper (2020). Brownlee points out various "problematic features" of existing professional codes—various pathologies to which they are vulnerable. One is that members of a profession are likely to accept without question the moral worth of the practice in which they are engaged, and base their code of ethics on that assumption, whereas a more critical scrutiny might show the practice to be, if not "rotten to the core", morally suspect in some of the aims it pursues or the means by which it pursues them: but if the practice is morally suspect, its code of ethics will fail to identify the kinds of conduct in which its members really ought, or ought not, to engage.[12] Another problem is that codes of ethics can provide a spurious respectability, an "appearance of ethical credibility" to a profession, so protecting it from critical scrutiny. They can also require members to behave in ways that are, seen from a larger perspective, unethical, and encourage (if not coerce) an uncritical obedience to what

12 Brownlee comments on those who "engage in profit-hungry high finance" (2020: 177): I once heard a talk on professional ethics in the financial sector, during which the lecturer told us that one of the "dilemmas" that people working in the sector faced was "How far should you exploit the customer's ignorance?"; if that figures in your conception of your professional activity as a "dilemma", something is badly wrong.

the code requires. Brownlee also points out that the distinction (a crucial one on my account) between what falls within the realm of "professional conduct" and thus legitimately concerns other participants in the practice, and what falls outside that realm and thus counts as "private" as far as the profession is concerned, is not a clear-cut or uncontroversial one. Similar problems, she argues, arise if we see criminal law as analogous to a code of professional ethics—indeed, they arise in sharper form. Most obviously, we must ask how members of a polity are to arrive at an acceptable conception of their *res publica*—of the goals and values that are to structure their collective life; we must recognise the danger that those goals and values will not be ones that are worth pursuing; and we must ask who gets to determine those goals and values, and the laws that are to reflect and protect them.[13]

I agree with almost everything Brownlee says about the kinds of problem that can infect, and undermine the moral authority of, both professional codes of ethics and the criminal law, and I should have made clear that in my discussion of such codes I was not assuming that the kinds of code we actually have are unproblematic: that those whom they claim to bind ought simply to accept them; that the processes by which they are drawn up are unproblematically legitimate; that they reflect conceptions of the worth of the relevant profession that are not to be questioned. I do assume that professions can be of significant value—that such professions as medicine, law, education can (if properly constituted and governed) contribute to human well-being; and that there is good reason for any profession's constitution to include a code of ethics, dealing with the kinds of conduct in which its members should (not) engage in the course of their professional activities, and procedures through which they can be called to account for professional misconduct. I also assume that, however much room there might be for disagreement around the borders, it will be possible, and ethically necessary, to draw a distinction between the "public" and the "private" in the context both of professions and of the criminal law. Now Brownlee seems to accept that codes of ethics, and thus presumably the professions whose codes they are, can be valuable, which is also to accept that they need not "instrumentalize moral reasoning"—that they need not pervert moral thought by making it an instrument of the

13 Brownlee finds my talk of the "Founding Parents" who are, I imagine, to draw up a polity's constitution and its criminal law "jarring", because "it occludes the fact that the people drafting the laws in almost all polities have not been Parents, but Fathers and, indeed, white Fathers" (Brownlee 2020: 188; see Duff 2018: 92-101). My talk of "Parents", rather than of the "Founding Fathers" who loom so large in the rhetoric of the US Constitution, was precisely intended to jar: to remind readers that if we are engaged in normative theory, rather than in positivist description, we must not take for granted the identity of the legislators.

profession's morally disreputable ends, or serve to protect its members from moral scrutiny (see Brownlee 2020: 182). So is her worry simply that I pay inadequate attention to the ways in which both professions and polities, and their codes, can go wrong: that my discussions display an unwarranted complacency about the kinds of code, and the kinds of criminal law, that we actually have? I think there is more to it than that—and that "more" highlights an important aspect of my account.

One problem, in Brownlee's eyes, seems to be that my account, both of professional codes and of criminal law (or of a polity's civil order, from which its criminal law will flow), is set in overly "formal" terms that do not allow us to distinguish the substantively acceptable from the unacceptable—those that are created, by suitable procedures, to serve worthwhile ends by morally appropriate means, from those that are not: I must recognize (as an "extreme example") that the SS had a code of ethics (Brownlee 2020: 177); and the "formal" character of professional codes too often serves to conceal, rather than to illuminate, the substantive moral values by which the profession's members ought to guide their conduct (2020: 179-80). There are two distinct worries about formalism here.

The first concerns the character of the codes and the role they play in professionals' lives: if they are expressed as sets of formal rules, which the members are simply to obey without further (self-)critical thought, they can indeed have the kind of morally blinkering effects that Brownlee highlights; similarly, if citizens see their criminal law simply as a set of rules that they are to obey, this is likely to blind them (or to help them blind themselves) to the more nuanced moral considerations that should guide their actions. However, that is an objection not to professional codes of ethics as such, but to some of the ways in which they can be formulated and treated; and, in the case of criminal law, I argued that the criminal law of a decent polity will not be a set of prohibitions or rules that citizens are simply to obey. The citizens will maintain a critical attitude towards their criminal law, as a fallible attempt to express the polity's shared values; they will be ready, as good citizens, to act if necessary in ways that it defines as criminal; and even when it can claim their allegiance rather than their critical opposition, they will apply it with a critically discretionary understanding—they will not simply follow it blindly. There are questions, in particular in the context of the kinds of regulation whose breach constitutes a *malum prohibitum*, about when we should subject our conduct to relatively strict, formal rules—for instance to regulations specifying the maximum speed at which we may drive, rather than to the injunction not to drive dangerously (see at nn 30-32 below): but neither codes nor criminal laws *need* be "formal" in ways that distort our practical reasoning.

The second worry concerns the "formal" character of my account of both professional codes and the criminal law: that it makes room for, or allows in, codes that serve the ends of morally repugnant "professions" (like the SS), or that fail to capture the moral considerations that should guide the profession's members; that it makes room for profoundly undemocratic systems of criminal law that function to protect the interests of the powerful rather than to serve the common good. This is certainly true—up to a point. My initial aim was to give a rather formal account of criminal law (as involving the definition of a set of public wrongs and the provision of a process through which those accused of such wrongs would be called to public account), and of the idea of civil order (and of a polity's constitution) in which the criminal law is to be grounded; analogously, my aim was to provide a rather formal account of professional codes of ethics, which did not try to say anything substantive about the aims a profession should pursue, about which professions are worthwhile, or about the appropriate content of their codes of ethics. The account is not *purely* formal: not *anything* could count as a profession or professional code, or as a conception of civil order, or as a system of criminal law. But it counts as professions both the good and the bad, so long as they can be seen to be oriented towards some intelligible conception of a good that is worth pursuing; it counts as codes of ethics codes that can be seen to govern members' conduct in the light of that good;[14] it counts as polities with conceptions of their civil order utterly undemocratic societies that are oriented towards substantively unacceptable goals, or that take substantively unacceptable views of what counts as "public" or as "private", and the content of whose criminal codes is therefore substantively unacceptable.[15] For a current example, if the Indonesian parliament enacts its proposed new criminal code, which criminalizes "living together outside marriage" and "extramarital sex",[16] that will be a genuine criminal law, which reflects a conception of Indonesia's civil order as including matters concerning marriage and sexual relationships that, in western liberal societies, count as "private"; that code, and that conception of civil order, might not have been arrived at through a process of democratic deliberation, but that cannot disqualify them from counting as a conception of civil order and as a code that gives expression to it. Am I therefore committed to approving that criminal code, that conception of civil order, as acceptable—if not

14 And even codes that are purported thus to govern members' conduct, but are better understood as covers for the profession's unethical aims; to count these as codes of ethics is also to make clear that they are to be assessed in the light of their purported aims.

15 "Substantively unacceptable" to whom, or by what criteria? See s. 5 below.

16 See https://www.theguardian.com/world/2019/sep/20indonesia-hundreds-of-thousands-oppose-plan-that-would-outlaw-extramarital-sex.

universally or for "us", at least for Indonesia? Of course not—just as in recognizing that someone who believes that homosexual conduct is morally wrong, or that it is morally permissible to factory farm animals, holds a genuine moral view I am not thereby committed to endorsing that view as acceptable.

In theorizing about criminal law (as about morality), even if our aims are normative, we should distinguish different stages or elements, only some of which are (directly) normative. We may begin with a (relatively) formal account of criminal law, as I do in ch. 1; we might add, as I do, a relatively formal account of the role of criminal law, as thus understood, in the structure of a political community (this will include a formal account of the idea of a political community). This might lead us to the formal principle of criminalization that I advocate: one that is formal in that it leaves open the question of what substantive content is to be given to the concepts of "civil order" and "public wrong". The point of beginning with such a formal account is partly to clarify the nature of the practice we are theorizing; but also to bring into view the kinds of argument that are relevant to assessing, justifying or criticizing systems of criminal law. One who wants to justify a criminal code must, if my formal account is right, be able to argue that the conduct it defines as criminal is a public matter, and is wrongful, and must therefore also be able to articulate a conception of the polity's civil order in which such conduct will count as public and wrongful: but there are limits of intelligibility on what can even count as a conception of civil order or of public wrongs; and in offering such arguments the advocate will bring her claims within the reach of normative political, legal and moral appraisal—she will have to subject those claims to those kinds of appraisal.

A second stage or element is roughly procedural: we discuss the kind of process through which a polity's (self-)constitution, its conception of its civil order, and its criminal law are to be constructed. We cannot say, *a priori*, that those processes must be in any way democratic: monarchies and oligarchies are political communities, structured by some (non-democratic) conception of civil order, and maintaining systems of criminal law. But, changing gear from the analytical to the normative, we can argue that a decent polity, one we could recognise or accept as a respect-worthy political community, must be democratic in its self-constituting procedures (we will then need to go on to discuss which kinds of procedure should count as "democratic"). Familiar questions now arise about who "we" are to make such arguments, and what purchase they should rationally have on non-democratic rulers: about, for instance, whether we can rationally demonstrate that a polity *must* be democratic if it is to have any justified

claim to legitimacy or authority—and about what follows, in terms of the claims or judgments that we can properly make, if we must accept that no such rational demonstration is available (questions, that is, about "relativism"). We need not pursue such questions here: we need only note the familiar point that impeccably democratic procedures can produce conceptions of civil order, and of what counts as a "public wrong", that "we" (postponing again the question of who "we" are) find unacceptable.

If we take a positivist, detached stance, we can describe the conception of civil order that a polity espouses, and the conceptions of public wrongs and the proper scope of its criminal law that flow from that conception: in *P*, we might say, the civil order encompasses sexual relationships, and a heterosexual view of the kinds of sexual relationship that are legitimate; in *P*, accordingly, and consistently with that conception, other kinds of sexual relationship are criminalized as public wrongs. This does not commit us to endorsing *P*'s criminal law: once we replace our positivist hats by more committedly normative hats, and engage (as citizens, as external critics or as normative theorists) in substantive argument about how we, or they, should conceive their civil order and their criminal law, we can argue that consensual sexual relationships should be seen as falling, as far as the polity is concerned, within the "private" rather than within the "public" realm, and/or that non-heterosexual relationships should be accepted and recognized as valuable; and so on. Once we make this move to substantive normative arguments about the content of civil order or of the criminal law, we are at the third stage or element of our theorizing—a stage in which I did not seriously engage in this book (although in ch. 5 I sketched, as an illustration, what I take to be a plausible substantive conception of a republican polity). This was not because I think this stage unimportant; it was because I wanted to make clear the crucial initial stages of a theory of criminal law.

In brief, then, Brownlee is right to highlight the kinds of deficiency or pathology that can affect (infect) professional codes of ethics, and the criminal law; she is right to point out the ways in which such codes can distort or subvert the kind of ethical thinking that should guide the conduct of members of professions (and of citizens): that is why members of professions need to take a suitably (self-)critical stance towards their professions' self-justifications and their codes; that is why citizens need to maintain a (self-)critical stance towards their polities' conceptions of their civil order and towards the criminal laws that help to sustain that order. But none of this is a threat to the (quite limited) analogy that I drew between the criminal law and professional codes of ethics; nor does it undermine the claim that criminal law *can* serve an important and

valuable role in the life of a political community, just as a code of ethics can play an important, valuable role in the life of a profession.

5. MULTICULTURALISM AND DEMOCRATIC DIALOGUE

I have talked about a political community and its members; about how they can develop and articulate a conception of their civil order, and an account of the shared values that structure their civic enterprise of living together as fellow citizens. Such talk makes sense, it might be said, in societies with culturally homogenous populations: societies whose members can be said to share a form of life, and who can be expected to agree on the key values that inform that way of life. This is, however, not our present condition, as Roberto Gargarella forcefully reminds us: the societies about whose criminal laws contemporary criminal law theorists seek to theorize are characterized not by cultural homogeneity or agreement, but by deep diversity and difference. The pressing question then is whether and how we can normatively theorize criminal law in ways that do justice to the facts of "multiculturalism": Gargarella argues that my account fails to do adequately democratic justice to multiculturalism (Gargarella 2020). Indeed, his critique is broader than that: although, on my account, the voice of the criminal law should be the voice of the political community, my account does not make that possible, whether or not the community is a multicultural one.

I do not think that Gargarella and I disagree, or need disagree, as much as he supposes we do. To show why that is so, I begin with four preliminary points, to clear up what might have been some misunderstandings.

First, Gargarella takes me to define "public wrongs as faults that 'are clearly inconsistent with, manifest violations of, any remotely plausible conception of civil order'", or "conducts that are 'intrinsically … inconsistent with civil order'" (Gargarella 2020: 195-6, quoting Duff 2018: 300). But I am not there discussing public wrongs in general: I am talking only about murder and rape as, I supposed, obvious examples of kinds of wrong that any polity ought to criminalize; I accept, as an implication of my account of the relationship of criminal law to civil order, that we will find differences both within and between polities about many public wrongs—about which kinds of conduct should be seen as public matters, or as wrongful. All I meant to claim was that it is hard, perhaps impossible, to imagine a polity whose members share a conception of their civil order according to which it is not a public wrong to kill any other person at will, or to force sexual penetration on others regardless of their will. That is not

to say that we cannot all too easily imagine a society in which members of some groups are treated as "fair game" for members of the dominant group: it is only to say that we cannot imagine a community (as distinct from a mere aggregate of individuals living within the same geographical territory) in which everyone is fair game for everyone else; and that a society in which members of one group were in that way fair game for members of another, dominant group would not be normatively plausible, or a society of *citizens* living *together*. Indeed, that claim is even stronger if we accept Gargarella's account of how a society ought to go about determining its voice and its laws—through a

> "public, collective conversation, where each person participates as an equal together with all the others. Every voice has then to be listened, every viewpoint has to be considered, every interest has to be taken into account" (2020: 205).

Such a conversation would predictably lead to different substantive conclusions in different groups or communities, but there are constraints on those conclusions. For if the participants are to view and treat each other as equals, that requires a mutual respect between them, which is inconsistent with seeing some other participants as fair game who can be killed or sexually coerced at will.

Second, even in cases in which I take it to be uncontroversial that any polity must count some wrongs, such as murder and rape, as public, criminalizable wrongs, I do not suppose that their precise definitions as public wrongs will be unproblematic. Any non-pacifist polity will allow, for instance, that killing another is sometimes permissible: so a polity will need to decide just how to define murder as a public wrong; and here there will be ample scope for disagreement. So too with rape: apart from the question of whether there should be a distinct crime of "rape", as opposed to various crimes of (sexual) assault,[17] there is a familiar range of difficult questions about the precise scope and definition of the crime. That is why I note that the process of defining even obvious *mala in se* as crimes is a process of construction (Duff 2018: ch. 7.4): we cannot simply reproduce in our criminal law some already available pre-legal understanding of just what the wrong consists in.

Third, one of my claims is that the criminal law's offence definitions must purport to be, or be justified as, definitions of kinds of conduct that are already wrongful independently of the criminal law. Some are (allegedly) pre-legal wrongs, wrongful independently of and prior to the

17 Compare ss 271-3 of the Canadian Criminal Code.

law: these are "*mala in se*". Others are wrongful only in virtue of the law, for instance in virtue of a legal regulation that prohibits them, and are thus not pre-legally wrongful: but they must still be wrongful independently of the criminal law that criminalizes them; these are "*mala prohibita*", which are pre-criminally, but not pre-legally, wrongful.[18] So when I say that the criminal law does not create the wrongfulness of the *mala in se* that it criminalizes, but rather "declares that these pre-existing, pre-legal wrongs" are public wrongs (Duff 2018: 123; see Gargarella 2020: 200), all that I mean by "pre-existing" is "pre-legal"—existing prior to the law. If we then ask whether such wrongs can be said to "pre-exist" not just the law, but the (or any) political community, we enmesh ourselves in questions about the metaphysics of value that neither Gargarella nor I need to tackle for our present purposes. We are neither of us relativists (Gargarella 2020: 195): thus while we can agree that in some sense communities "construct" the values by which they live, their constructions are not immune to criticism; dissident participants or interested observers can argue that they are wrong. We need not here tackle the further question of what kind of metaphysical foundation such criticism requires.

Fourth, my "oscillations between liberal and communitarian positions" are said to create "confusions" (Gargarella 2020: 192)—as if one cannot consistently be both a "communitarian" and a "liberal", which is what I claim to be. Now on Gargarella's understanding of liberalism, that is a fair point: for he takes the distinction between liberalism and communitarianism to be one "between the universal and the communal; the general and the particular; the critical view about the structures in which we live, and the attachment to the communities where we live"; between "shared, localized, communal values" and "more universal values" (2020: 198). Thus whilst "liberals" will insist on assessing the values espoused by particular communities in the light of (supposedly) universal values, "communitarians" cannot, it seems, take such a critical stance towards local values. But that is not the sense of "liberal" in which I claim to be a liberal, nor indeed the sense of "communitarian" in which I claim to be a communitarian —although given the variety of possible meanings for each of those terms, I should have made my usage clearer, which is what I

18 I thus defend (Duff 2018: ch. 2.3-4) a "Strong Wrongness Constraint". I should comment briefly here on the cautionary "(allegedly)" in "Some are (allegedly) pre-legal wrongs". Gargarella picks up on a cognate remark that the criminal law "must speak, or claim to speak, in the voice of" a political community (2020: 203; see Duff 2018: 110), and rightly notes that speaking and claiming to speak are very different things. All I meant was that if the law is to make a claim on our allegiance it must claim to speak in our voice, and we can then discuss whether that claim is justified; only if that claim is justified does the law *have* a justified claim on our allegiance.

will now try to do (see further Duff 2001: ch. 2).

Communitarianism is opposed not to liberalism (in all its forms), but to a particular type of "individualism", according to which political deliberations must always begin (and end) in the first person singular: each person must ask: "How am *I* to live?". For communitarians, the question must of course be answered, in the end, by each person: but the question to be asked is "How are *we* to live?"; and the "we" are the members of a particular, and inevitably local, community in which the asker finds herself or himself.[19] Thus the values by which a political community is to live are necessarily local: they cannot but be the values that the members of that community understand as theirs. (Democratic communitarians will go further than this, to argue that the values by which a political community is to live *ought to be* determined by its members by a process of democratic deliberation.) There is thus no "view from nowhere" (Nagel 1986), from which we can determine the values that should guide a polity; political deliberation must begin from, be grounded in, somewhere—some particular local community of thought. But two things, which worry Gargarella about my view, do *not* follow from this.

First, it does not follow that a community's values cannot be criticized either by its members or by outsiders. Even if we can imagine a community whose modes of thought, whose normative and conceptual structures, are so monolithic, and so insulated from others, as to leave no room for internal dissent to even make sense, and no scope for external critique via shared concepts or values,[20] the political communities that actually exist are not like that: there is conceptual room, even if in more totalitarian polities there is no normative welcome, for internal critiques; and there is room for critiques from outside appealing to values that are, if not currently espoused by the polity, at least accessible to it.

Second, it does not follow that a political community cannot espouse liberal values—or that critics cannot argue that it should do so: values such as individual autonomy or freedom, privacy, equality, "equal concern and respect" (Dworkin 1986: 190)—although those values would, on a communitarian account, be given a communitarian interpretation as values that structure *our* lives together. That is how I can be a liberal

19 I say "inevitably local" for two reasons. First, the scope of any community is contingent, however large it might in fact be, whereas the "universalism" with which Gargarella is concerned is *a priori*, concerning all rational beings. Second, as a matter of fact, the political communities in which humans live are relatively local rather than global: whatever power there is to the idea of a cosmopolitan community of "humanity", political communities as we have them are much more local.

20 Though even Oceania, in Orwell's 1984, did not manage to achieve this.

communitarian: I can hope to live in a polity structured by liberal values; as a member of this polity, I can argue that we ought to espouse liberal values; I can argue with members of other political communities, which do not espouse liberal values, that they ought to do so—not by appealing to some supposedly *a priori* universal values binding on all rational beings, but by appealing to the conceptual links that exist between their existing values and liberal values. Thus, to take one of Gargarella's examples (2020: 197-8), someone who mocks the traditions and way of life that Catalans regard as important to their identity as a political community, *might* be seen by Catalans as attacking their *res publica*. But, first, this is not necessary, since their conception of their *res publica* might include a commitment to robust political debate of which this kind of critique is a part. Second, even if it is an attack, they can be committed to a liberal value of freedom of speech, given which such attacks must be tolerated, if not welcomed—and we could argue, either as Catalans or as outsiders, that they ought to be thus committed.

With these preliminaries out of the way, we can turn to Gargarella's central concern: that my account cannot do suitably democratic justice to multiculturalism, or to the manifest fact that we do not live in normatively homogenous communities. The problem here might take two forms. First, we can imagine that several culturally and normatively distinct groups, or a large number of individuals with very different values, share a territory and have to work out how they can live together, given that they are roughly equal in power: the question then will be whether they can come to agree on enough in the way of values and procedures to make it possible to live together as a political community—as distinct from living beside each other in a state of uneasy co-existence. That is not, I think, the version that worries Gargarella—or the version that we typically find in our contemporary world. The more usual version is that an existing polity contains a dominant cultural group (dominant perhaps in numbers, certainly in power), and one or more weaker groups; or a dominant majority and a dissenting minority who do not share the majority's core values. What is to be done in such situations?

A tempting way to approach this question is to ask what "we" should do about such non-powerful minority cultures, or such dissidents: what can "we" say to "them"; how far should "we" tolerate "their" different values and customs, or allow "them" to live by their distinctive values? But, as I think Gargarella and I agree, that is a dangerously misleading way to put the question: for it implies that those who must answer the question, the

"we", are the members of the dominant majority; and that they, *and only they,* must decide what to do about or to or with these others. That is what all too often in fact happens: but if the question is what *should* happen in a polity that aspires to democracy, the "we" who decide what to do must include all those in the polity, in a democratic deliberation in which every voice can be heard, and is heard with attention and respect—in, as Gargarella puts it, "a public, collective conversation" (2020: 205). What form that conversation should take, how it should be institutionalized (since any large polity needs institutions to actualize its democratic aspirations), who is in the end to speak for "the polity"—these are familiar issues for democratic theory, on which I won't take a view here. But whatever form that conversation is to take, it must end with a decision (or, if no decision can be reached, with the collapse of the polity): one that we cannot realistically expect to be unanimous, even if we can hope for something close to unanimous agreement on the kind of procedure by which it is to be reached. We (the "we" who constitute the polity) must address and listen to each other respectfully: we must try to understand the values by which others want to live, which might require us to try to understand the distinctive way of life in which those values are embedded; we must see whether disagreements can be resolved through discussion, so that those who initially disagreed about how to live can come to agree; if we cannot agree on those substantive questions, we must try to agree on where differences should be accepted—you live your lives, we live ours (bearing in mind that the "you" and the "we" are still fellow members of a larger "we"). I think that Gargarella and I also agree on two further points.

First, we are not relativists who would say that each culture must be left free to determine its own way of life (see Gargarella 2020: 195):[21] it must be open to "us" to come in the end to the view that a particular culture should not be allowed to continue with this or that culturally embedded practice (with a practice of "female genital mutilation", for instance, or of "honour killings" of disobedient daughters),[22] even if that is not a view that members of the particular culture can be persuaded to accept. Second, as should already be clear, we do not believe that the dominant group should claim the right simply to enact (to "impose") its own values: even if they have the

21 I am talking here of how "we" are to deal with multiculturalism: the issues raised by individual dissenters or rebels have the same form, and can be addressed in the same way (see further Duff 2018: chs 3.4, 5.7).

22 It is worth noting how disagreement about the permissibility of certain practices is also disagreement about how they should be described—hence the scare quotes in this bracket.

effective power to do so, they must engage as citizens with all their fellows, seeking to build a "we" that includes everyone in the "public, collective conversation".[23]

In the end, however, and assuming that quite often we will be unable to agree either on the substantive issue or on whether we should simply accept difference, there will need to be a decision procedure, from which will emerge (if the polity can survive at all) something that can count as the polity's view, and therefore as "our" collective view; and there will then be people (officials of some kind) who are entitled to speak and act in the community's name. If there is to be a criminal law (something that requires argument, but that Gargarella does not seem to deny), it will then define as public wrongs some kinds of conduct that some members of the polity will not see as wrongs. I offer no view here on what those should be, but again emphasize the importance of distinguishing the procedural from the substantive stages of the argument. We can argue first about the procedures through which a democratic polity should determine its laws; and we can argue secondly about what the substantive outcome of those procedures should be—for instance when we engage, as citizens, in the public conversation that will determine our laws: but we must recognize that the outcome of a democratically impeccable procedure might be one with which we substantively disagree—that is a familiar truth about democracy. All of which is to say that both Gargarella and I recognize the serious problems raised for would-be democrats by the facts of deep cultural differences and deep normative disagreements within contemporary polities; we agree that such problems must be addressed within the polities themselves, through a democratic deliberation in which every citizen has an equal voice; we recognize that such deliberations may result, not in unanimous agreement, but in a decision with which some will be unhappy; and we therefore realize that there can be no clean and simple route to a resolution of such disagreements—only the messy processes of democratic deliberation.

23 There are then further familiar problems about the possibility of such conversation: suppose that one group refuses even to listen to the kinds of argument that another group offers (one aspect of the problem of "public reason"), or denies that certain kinds of people (identified by gender, or race, for instance) should have an equal voice, or any voice, in the conversation. Although the appeal to democratic conversation generates a largely procedural account, the procedure itself involves certain substantive pre-commitments about the standing of those who are to engage in it.

6. PRINCIPLES OF CRIMINALIZATION

I must turn now to issues more directly concerned with criminalization. Tatjana Hörnle and I agree that we cannot expect to articulate a single thick master principle of criminalization—a principle providing determinate substantive criteria by which to decide what kinds of conduct we have good reason to criminalize: any such principle will either be too narrow, as failing to capture kinds of wrongful conduct that we clearly have reason to criminalize; or become so diluted, as it is revised or extended so as to capture every kind of criminalizable conduct, that it can no longer offer substantive guidance (Hörnle 2020; see Duff 2018: ch. 6). My response to this critique of "single principle" theories of criminalization is still to offer a single master principle, but one that is avowedly very thin: the "public wrongs" principle explained in s. 2 above. Hörnle's response is different: she argues that we can still identify determinate, thick principles of criminalization; but that we need two (or even three) such principles, rather than just one, if we are to capture all the kinds of conduct that there is good reason to criminalize.

I discuss Hörnle's proposed principles below, but should first make two brief comments on the relationship between her account and mine. First, her substantive principles could be seen as complementary to, rather than as competing with, the thin principle that I offer—as spelling out in more substantive terms the criteria by which we are to identify relevant public wrongs: a type of conduct constitutes a public wrong if it is incompatible with an important collective interest, or violates another's right to non-intervention. Second, a merit of the kind of thin principle that I offer is that, whilst it does not provide substantive guidance about what kinds of conduct we have reason to criminalize, it shows what kinds of argument are relevant for, or against, criminalization: thus whilst it might not *directly* address the "genuine puzzlement" of legislators or civil servants deliberating about criminalization (Hörnle 2020: 212), by offering them direct guidance on what they should or should not criminalize, it can help them to see what kinds of reason they need to consider, what kinds of argument they need to offer.

Now for Hörnle's proposed principles. Two are responsive principles concerning conduct that violates others' rights "to non-intervention": we categorically *must* criminalize all violent attacks that violate "important rights of others"; and we have good reason to criminalize other kinds of

violations of the right to non-intervention.[24] How thick are these principles; and how complete, as regards kinds of criminalizable conduct that do not violate collective interests? As Hörnle points out, if our topic is criminal law, we must be concerned with rights as they figure in our relations to each other not simply as moral agents, but as citizens: with, I would say, rights as they figure within, or bear on, the civil order in which we live as citizens.[25] As she also points out, there is no serious dispute about the criminalizability of the most obvious kinds of right-violation in this context, such as murder, rape, and other kinds of violent attack on the person (although there is room for argument about just how we should define these as crimes). She also argues that we can distinguish a "core" set of rights-violations that we *must* criminalize from other kinds of violation that we have only *pro tanto* reason to criminalize, and that I do not do justice to that core set, since I say only that we must criminalize them *if* we have a system of criminal law—whereas she argues that we have "a categorical duty to create criminal law in the core area" (2020: 216; see Duff 2018: 277). This is not a significant disagreement, since I explain later (2018: 299-301) why such wrongs must count as serious public wrongs, which we have very good reason to criminalize. I do not, it is true, insist with Hörnle that we must create criminal law *in order* to criminalize them: for I would not rule out *a priori* the possibility that a polity could find suitable ways to mark and respond to such wrongs that might not count as "criminal law"; but I agree that they violate values that must be central to any plausible polity's conception of its civil order, and are inconsistent with the most basic requirements of civil order. However, first, to see why these are such obviously, indisputably, central kinds of public wrong, we need to appeal to a conception of a polity—of, as I would put it, civil order; or, as Hörnle puts it, of what it is to be a citizen living with other citizens. Second, how are we to distinguish "core" rights-violations, which must be

24 Hörnle seems to think (2020: 216) that I would deny such essentially responsive principles, since I think that the criminal law is "essentially preventive", but my view is the reverse of this. In the passage from which she quotes (Duff 2018: 261), I begin by noting that on my account criminal law is "essentially responsive", but then add that it is "also essentially preventive, or perhaps … persuasive". Thus I agree that in the case of violent rights-violations the primary reason to criminalize is responsive; I differ from her, as we will see, in that I would say the same about our reasons to criminalize conduct that violates important collective interests.

25 Another point of clarification. Hörnle argues, criticising my account, that "[c]rimes against persons are not foremost conduct that violates the civil order. They are this, too, but this is not the most salient description. The main feature is interpersonal" (2020: 214). I agree: the wrong that we have reason to criminalize, what I call the object of criminalization, consists precisely in the violation of another's rights; that it violates the civil order is relevant purely as a condition of criminalization, as making that wrong the collective business of the polity and its citizens (see Duff 2018: 216-7).

criminalized, from other kinds of violation of the right to non-intervention? Hörnle specifies "individuals' most important rights (rights to life, bodily integrity, sexual autonomy, freedom of movement, freedom from coercion)", violent attacks on which will then form the core of criminal law (2020: 215):[26] but it is not clear, at least to me, why *only* violent attacks should count; or whether *every* kind of violent attack (including for instance minor physical assaults) must count; or why only these rights, and the interests that they protect, should be included. More significantly, it is not clear why we should try to draw a sharp distinction between the core and the non-core—rather than saying, more plausibly, that we can distinguish along a spectrum (or two spectra) more from less important rights, and more from less significant kinds of violation of them; and that the more important the right, and the more significant the violation, the stronger our reason to criminalize it.

If we abandon, as I suggest we should, Hörnle's sharp distinction between core and non-core, we are left with just one principle in place of her latter two: that the "state has reason to criminalize conduct if it violates another person's right to non-intervention".[27] But how thick or substantive is this principle? We need, of course, an account of what kinds of conduct are to count as "intervention", and of what kinds of intervention violate rights. I cannot pursue this in detail here, but would suggest, first, that any account of "wrongful intervention" must appeal to a conception of civil order, since only on that basis can we give an account either of what counts as intervention (since intervention must be into a sphere one can count as one's own), or of which kinds of intervention are wrongful: but in that case, in giving an account of rights to non-intervention we are already giving an account of public wrongs, and the notion of "rights to non-intervention" proves to be hardly thicker than that of "public wrongs", save that it focuses our attention on wrongs with individual victims. For we are giving an account of the kinds of conduct that we are entitled to expect or demand of each other, as citizens—or of the kinds of conduct about which we are entitled, as citizens, to complain: whether we call such complaint-grounding conduct victimizing public wrongs, or rights-violations, we are in effect saying the same thing. Thus to ask whether, for instance, we have reason to criminalize conduct that violates dignity would be, on Hörnle's account, to ask whether citizens have a right not to be treated with indignity—whether this constitutes a wrongful intervention. On my

26 Though she should presumably, given what she and I agree on (see above), talk of the most important rights of individuals *as citizens*.

27 2020: 220. She adds "(however, there might be countervailing reasons)", but that seems unnecessary: if a principle says only that we have reason to criminalize, that leaves open the possibility that there might be other, countervailing reasons in.

account, it is to ask whether our civil order should be defined, in part, in terms of respect for dignity: these seem to me to be in effect the same question, in which case Hörnle's rights-violation principle is hardly thicker than my "public wrong" principle.

Second, we must also ask whether all victimizing crimes could be portrayed as violating a right to non-intervention—without depriving that right of any substantive content. Familiar examples will surface here: crimes of omission—of failing to help someone in desperate need (can such failures be portrayed as wrongful interventions?); or crimes committed against the dead such as corpse desecration (see Jones 2017); or crimes against non-human animals.[28] The key question is again whether these are wrongs that should concern us as citizens, in virtue of their incompatibility with some value internal to our civil order; I do not see that it helps much to translate that question into one about rights.

I turn now to Hörnle's other principle, concerning the protection of collective interests, and begin with a clarification. Hörnle insists that this principle, unlike those which deal with rights-violations, is "straightforwardly preventive" (2020: 213), but she conflates two different types of (apparently) preventive principle. One is indeed "straightforwardly preventive": "We have reason to prohibit X if this will prevent harm".[29] The other, by contrast, looks initially responsive: "The state has reason to criminalize conduct if it is incompatible with important collective interests"—though it adds "that cannot be adequately protected by other means" (2020: 219-20). The former principle is purely preventive and forward-looking: we are to ask simply whether prohibiting this type of conduct will prevent harm to collective interests; and although the obvious way to prevent such harm is to prohibit conduct that is likely to cause such harm, we might sometimes secure preventive benefits by introducing regulations that also prohibit conduct that is neither harmful nor dangerous. The latter principle, by contrast, is responsive, since it requires criminal sanctions to be a response to conduct that itself violates or threatens a collective interest, even if the purpose of that response is (in part) to prevent future harms. The difference between these two types of principle can be illustrated by driving offences. If we criminalize dangerous driving or driving when rendered unfit through drink or drugs,[30] we criminalize conduct that is itself incompatible with important collective interests—and is for that reason wrongful. If instead

28 See e.g. (Animal Welfare Act 2006). Hörnle notes the possibility of appeals to animal rights (2020: 213, n. 5); but it seems to me unnecessary to go down that road.

29 Hörnle 2020: 213. She cites Duff 2018: 238, where I talk of having reason to *criminalize* a type of conduct if by doing so we will prevent harm; as we will see, the replacement of "criminalize" by "prohibit" matters.

30 See e.g. Road Traffic Act 1988, ss 2, 4.

we prohibit driving at more than a specified speed, or with more than a specific proportion of alcohol in one's blood,[31] we are likely to prevent harm more efficiently, but we do this by including within the prohibition conduct that is not itself harmful or dangerous, since not all of those who break the speed limit, or drive with a higher than permitted level of alcohol, will be driving dangerously or be impaired. Such prohibitions are not yet *criminal* laws, since they do not mark the prohibited conduct as being wrongful (as criminal laws mark the conduct they define as criminal); the criminal law comes into the picture only when we criminalize, as wrongful, breaches of these kinds of regulation.[32]

Of these principles ("We have reason to prohibit X if this will prevent harm", "The state has reason to criminalize conduct if it is incompatible with important collective interests"), the latter is indeed a principle of criminalization, so long as we can say that conduct that is incompatible with important collective interests is for that reason wrongful: for if it is thus wrongful, it constitutes a public wrong in a very straightforward way, as impinging directly on a public interest. But our reason to criminalize is then, as in the case of rights-violations, as much responsive as preventive, since in both cases the publicly wrongful character of the conduct gives us reason (makes it appropriate) to respond to it by calling the wrongdoer to formal public account through a criminal process. In both cases, we also have a preventive interest; but what is intrinsic to criminalization is the responsive dimension. What is true is, admittedly, that when it comes to the question of whether we should criminalize all things considered (as distinct from the initial question of whether we have reason to criminalize at all), the balance between responsive and preventive considerations might differ as between different kinds of wrong. Perhaps responsive considerations will weigh more heavily in the case of serious wrongs, whilst preventive considerations will be more prominent with less serious wrongs: we might see strong or even compelling reasons to criminalize very serious wrongs, even if by doing so we do not expect to achieve much in the way of prevention;[33] whereas with (relatively) minor wrongs we might think that we should criminalize (rather than responding in some other way) only if this would be preventively efficient. Even so, this does

31 See Road Traffic Regulation Act 1984, ss 81-9; Road Traffic Act 1988, s. 5.

32 On these two kinds of principle, and the distinction (important to understanding the character of so-called "*mala prohibita*") between criminalization and regulation, see Duff 2018: 21, 65-8, 237-48. Another issue (Duff 2018: 16-19, 284-6) is whether we should maintain a (formally) non-criminal system of regulations whose violations do not (formally) count as criminal, like the German system of *Ordnungswidrigkeiten*.

33 Even if we expect no preventive effect? As Hörnle notes (2020: 216) it is hard to imagine such a case; and if criminal law as a whole has some preventive efficacy (and this is necessary to its justification), consistency requires us to criminalize serious public wrongs even if in particular cases this is not preventively effective.

not support Hörnle's claim that the principle concerning conduct that violates collective interests is a preventive principle, whereas that concerning conduct that violates others' rights is responsive (Hörnle 2020: 213-216). It would support her claim if violations of individual rights were always more serious than violations of collective interests, but that is surely not true: compare treason, or large scale tax evasion or fraud against the public purse, with damaging another's property or a minor physical assault.

We must also ask of Hörnle's "collective interests" principle, as of her "rights-violations" principle, how thick it is; and the answer is again that it is almost as thin as my public wrongs principle. For we must first identify the collective interests of the members of the polity (and assess how important each is): which we can do only by getting clear about the character of the polity, and the interests that its members share in virtue of their membership—about, that is, the polity's civil order. Hörnle is scathing about the German *Rechtsgutslehre*, according to which we have good reason to criminalize conduct that violates or threatens a *"Rechtsgut"*, a legally protected good: the concept of a *Rechtsgut* is, she says, "not only a thin concept, but so thin as to be an empty concept" (2020: 212). Now I would not myself call it "empty", since there will be constraints at least of intelligibility on what one could count as a good worthy of recognition or protection by the law; but if it is empty, so too is that of a collective interest, since to work out the interests of either an individual or a group, we must first have some idea of the good of that individual or group. We can try to distinguish criminal laws that protect, or respond to violations of, individual interests or rights from those that protect, or respond to violations of, collective interests,[34] and in both contexts there will be plenty of obvious cases of conduct that we of course have reason to criminalize (because, I would say, in such cases it is obvious that any plausible polity must count them as public wrongs): but when the cases become controversial or unclear rather than easy, I do not think that Hörnle offers a principle that is both thick and plausible in either the individual or the collective context can.

7. A RIGHT TO BE PROSECUTED—AND A DUTY TO FACE TRIAL?

Finally, I turn to Gustavo Beade's interesting discussion of "the right to be prosecuted" (Beade 2020). Beade's leading example is Mr Lamb, who makes some (to put it mildly) intemperate remarks at a promotional event,

34 Though this will sometimes involve controversy, for instance about the normative grounding of property crimes (on which see Simester and Sullivan 2005).

about the legal age of sexual consent, about rape, and generally about law. Given their context and their topic, his remarks are clearly a public matter: they are, it seems, intended as public contributions to public debate about matters of public concern. Beade also takes it that they clearly constitute a public wrong, as an instance of "incitement to hatred", of a kind that we therefore have reason to criminalize (2020: 232): but that is a bit quick. To argue, as Mr Lamb does, that the legal age of sexual consent (which we must suppose is specified in his polity as something over sixteen) is "an atrocity" does not look like incitement to crime, especially since he talks only of consensual intercourse in that context: it is an overdramatically rhetorical way of arguing that the law ought to be changed. His remarks about women who "need to be raped", and about the "rights of women" are more clearly wrong, as expressing unacceptable moral views: they might or might not constitute an offence of "incitement" in English or American law, for instance (we would need to inquire more closely both into his intentions and into the likely effects of his words) but they might amount to an offence in German law (to which Beade appeals), as "assault[ing] the human dignity of others".[35] Let us agree, however, that we can frame the example so that Mr Lamb's remarks clearly constitute a public wrong: they are, that is, wrongful in a way that properly concerns his fellow citizens, in virtue of their impact on or their implications for the polity's civil order.

If remarks of this kind constitute a public wrong, we therefore have on my account good reason to criminalise them. However, this is not yet to say that they should be criminalized, all things considered: for although we have reason to criminalize them, we must ask whether we should criminalize them *rather than* responding in one of the various other ways that are available to us, including doing nothing formal or legal, and relying solely on informal kinds of social response to do justice to the wrong (see Duff 2018: ch. 7.2). I won't discuss most of the other possibilities here,[36] but should say something about the significance of the kinds of informal social response that Mr Lamb suffered—or more precisely, since we are talking so far about whether to criminalize this kind of conduct, rather than about whether to prosecute Mr Lamb in particular, about the kinds of informal response that those who commit this kind of wrong are likely to suffer. The central question is this: should we see such responses as informal, not centrally organized, versions of the kind of response that is formally, publicly, provided by a criminal prosecution; or as something quite different? If the former, we must ask whether we should criminalize as

35 See German Criminal Code *(Strafgesetzbuch)* § 130(1).1; Beade 2020: 232, n. 14; for English and American law, see Ormerod and Laird 2018: 474-94; LaFave 2017: ch. 11.1.

36 But see in particular Beade's comments on why a "restorative justice" response might not be appropriate (Beade 2020: 229-30).

something additional to such informal responses (for instance because they are thought not to be adequate); or as a formal replacement for such responses, perhaps because they are liable to be misdirected or excessive. If the latter, then it is not clear why the likelihood of such informal responses should be relevant to the question of criminalization.

I am not quite sure how Beade would answer these questions. He talks of the informal responses as species of "public blame", which itself is "the foundation on which a certain type of retributive-communicative punishment is based" (Beade 2020: 224): this implies that a criminal prosecution would provide a formal, perhaps more serious or more proportionate, version of roughly the same kind of censorial response. On the other hand, he also talks of "public humiliation", and of "'lynching' in the mass media" (2020: 224), which seem rather different. I would myself say that blame (blaming another to his face) is a mode of calling to account: that is how it can provide the basis for "retributive-communicative punishment", and for criminal prosecution as a formal mode of calling to public account; and such a calling to account is an appropriate response to the wrongdoing of others whom we are still to see and address as responsible fellow members of our normative community (of, in this case, the polity). But public humiliation and media "lynching" are not modes of calling to account; they are therefore not species of public blame (at least of the kind that criminal prosecution formalizes, or of the kind that is appropriate between citizens).

However, whatever we should say about public humiliation or media lynching, there are clearly informal modes of calling to censorial public account that wrongdoers like Mr Lamb might face; these are indeed a feature of a number of recent social and political campaigns, part of whose aim is to "call out", or to call to properly public account, those who commit various kinds of wrong that have not hitherto been taken seriously enough. The question of criminalization is then in part the question of whether such informal callings to account can be expected to be adequately reliable and proportionate; if they cannot, that gives us good (which is not yet to say conclusive) reason to criminalize—both to ensure that an appropriate response is available *and*, perhaps, to pre-empt the possibility of dangerously disorganized and excessive informal responses.[37]

If we turn now from the question of whether we should criminalize conduct of the type in which Mr Lamb engages, to the question of whether he should be prosecuted, similar issues will arise. In England, for instance, the prosecutor would need to decide not only whether she could lead

37 Compare the idea that a system of formal criminal punishment can serve, in part, to pre-empt vigilante attempts at retribution or revenge: see e.g. Gardner 1998.

evidence at his trial that would give a reasonable prospect of conviction, but also whether it would be "in the public interest" to prosecute him; even in jurisdictions in which prosecutors are supposedly bound by a "legality principle" that requires them to prosecute if there is sufficient evidence of the person's guilt, they actually have a considerable degree of discretion as to whether to prosecute (or investigate) in individual cases. [38] I think Beade and I roughly agree on the kinds of consideration to which prosecutors should attend (and to which legislators should attend in deciding whether to criminalize such conduct in the first place): the importance of calling to formal public account those who commit public wrongs—which requires that the calling be done by some person or body who can properly claim to speak for the polity (something that the often self-appointed leaders of informal responses might claim, but not with reliable authority); and how and why this is owed to the wrongdoer as well as to the victim(s). As Beade rightly emphasizes, it follows that wrongdoers have a categorical but non-absolute right to be prosecuted: that is, not only a right to be prosecuted rather than being subjected to other kinds of legal coercion that may fail to treat them as responsible agents, but a right to be prosecuted rather than having their wrongdoing ignored. To put the point simply, if we are to treat each other as responsible fellow members of a normative community (of, in this case, a polity), we must be ready and willing to, inter alia, call each other to account for the wrongs that we commit against each other.

I want to focus here, however, on a question that Beade addresses only briefly: does the offender also have an "obligation", as Beade puts it, to explain himself—to answer for his conduct, or to provide an account of that conduct. Beade is sure that we should not assert any such obligation: "[w]e cannot claim that a polity may demand an explanation, nor therefore maintain that the wrongdoer is under an obligation to explain himself" (Beade 2020: 228). But this is, I think, much too quick.

If we focus first, as Beade appears to focus, on whether defendants should have a *legal* obligation to account for themselves, we need to be clear that to impose a legal obligation is not to "force" the person to answer (Beade 2020: 228)—unless the sanction for not answering is really drastic;[39] even if a failure to discharge such an obligation rendered the person liable not merely to formal censure, but to punishment, that need not amount to forcing them. We should also distinguish different obligations that might

38 On the two tests that English prosecutors must apply, see Ashworth and Redmayne 2010: 194-214; on the legality principle and its actual operation, see e.g. Perrodet 2002, Boyne 2017; see further Duff 2018: 45-9.

39 Defendants in English trials used to be *forced* to enter a plea, since anyone who refused to do so was subjected to the "*peine forte et dure*" to coerce a plea (see McKenzie 2005); but a legal obligation need not be thus coercive.

figure here: for instance, an obligation to appear for trial when summoned; an obligation to enter a formal plea of "Guilty" or "Not Guilty" (and, separately, an obligation to plead honestly); and an obligation not merely to answer formally to the charge by entering a plea, but to engage in the trial and to answer or account for one's conduct.[40] Thus in English law, for instance, defendants do have a legal obligation to appear for trial, and can be arrested and brought (forcibly) to court if they fail to do so;[41] they are asked, but are not legally required, to enter a plea (if a defendant refuses to plead, a plea of "Not Guilty" is entered on his or her behalf); and they are not legally required to play any part in the trial, or to offer any account of their actions—though exercising such a right of silence might of course have consequences, if it involves failing to alert the court to a possible defence. Now I do not myself find the first of these legal obligations, to appear for trial, normatively worrying (assuming, of course, a tolerably just criminal process as part of a tolerably just system of criminal law, and a non-oppressive police force), but I do agree with Beade that defendants should not be legally obligated either to enter a plea or to answer for their conduct, on pain of further sanctions if they refuse to do so—essentially for the reason that he gives, that if defendants are to be treated as responsible agents, they must retain the legal right to refuse to answer to the charge, and to refuse to account for themselves. A rebel or dissident who rejects the authority of the court, and who regards his trial as an oppressive travesty, will not be able to avoid being tried (the law cannot allow dissent to bar the trial): but he should be allowed, without further penal consequences, to make clear his denial of the court's authority by refusing to play any active part in the process.

However, it does not follow from this that the polity may not "demand an explanation"; nor that we cannot say that a defendant has a *civic*, as opposed to a *legal*, responsibility or obligation to enter a plea and to answer for his conduct in the trial. Civic obligations lack the formal force of law: they are nothing more (or less) than aspects of what we hope is a shared understanding of what it is to be a citizen, of what we owe to each other as citizens; they are enforceable only by informal persuasion from our fellow citizens; but they are crucial aspects of civic life—there would be something sadly amiss with a polity whose members recognized only legal obligations or responsibilities towards each other. If we are to ask whether those who have committed criminal wrongs, or those who are accused on reasonable grounds of committing such wrongs, have a civic obligation or responsibility

40 My comments here apply to "adversarial" trials; the account would need to be adjusted for trials in more "inquisitorial" systems.

41 At least for trials "on indictment", for more serious offences; see Bail Act 1976, s. 7.

to answer such accusations, and to answer for their conduct, in a criminal trial, the answer must surely be that they do—assuming, again, a tolerably just system of criminal justice, which addresses them properly as members of the polity.

The justification for this answer is that if I am a citizen of the polity whose criminal law I am accused of violating, I have a civic responsibility to assist the criminal law's enterprise, including the enterprise of calling criminal wrongdoers to public account: if I know that I am a criminal wrongdoer, I must recognize my crime as a public wrong, for which I should now answer, to my fellow citizens, whose business it is; and the criminal trial is the public forum in which criminal wrongdoers are formally called to answer. I must therefore be ready to face trial—and indeed, if I know that I am guilty, to plead guilty at my trial. Even if I believe (or know) that I am innocent, I have a responsibility to answer to the accusation, to explain why I am not guilty—and to offer evidence to back that explanation up: for I must recognise that the enterprise of calling public wrongdoers to censorial account requires a practice of calling *alleged* public wrongdoers to account, in order to establish whether they are guilty or not; and as a citizen whose law this is, I have a responsibility to assist in that enterprise, and to answer to my fellow citizens if I am accused (on reasonable grounds) of committing a public wrong. This is one of the ways in which citizens of a democratic republic should see themselves not merely as subject to its law, but as agents of the law—as active participants in the enterprise of democratic self-governance.

However, I should emphasize that this civic responsibility to participate in my criminal trial ought not (for the reasons indicated above) be turned into a sanction-backed legal duty; a citizen must retain the legal right to refuse to take part in the process. I should also emphasize that citizens can have such a civic responsibility only in a polity and a criminal justice system that is tolerably just: one that treats them as equal members of the political community, with the respect and concern that is therefore due to them; one that imposes humane, proportionate punishments on convicted offenders; and one whose criminal process enables those accused of crimes to answer for themselves as responsible citizens. I leave it to readers to ask how far our existing systems of criminal law meet these conditions.

8. CONCLUDING REMARK

I am painfully aware of at least some of the flaws in my book: in particular of the number of questions that it leaves unanswered, of various ways in which its arguments are under-developed or unclear, of how much work

remains to be done to articulate more adequately and persuasively a normative understanding of criminal law as helping to sustain and to constitute a polity's civil order, by providing an appropriate response to wrongs that violate that order. My commentators here have highlighted some of those flaws: but I hope that in my response I have shown that they have not cast doubt on the central features of this account of criminal law.

BIBLIOGRAPHY

Ashworth, A. J., and Redmayne, M., 2010: *The Criminal Process* (4th ed.), Oxford: Oxford University Press

Beade, G., 2020: "Should Public Blame Replace the Criminal Trial?" *Law, Ethics and Philosophy 7*: 222-233.

Boyne, S., 2017: "German Prosecutors and the Rechtsstaat", in *Prosecutors and Democracy,* ed. M. Langer and D. Sklansky, 138-74, Cambridge: Cambridge University Press

Brownlee, K., 2020: "Professional Ethics and Criminal Law" *Law, Ethics and Philosophy 7*: 174-190.

Chiao, V., 2019: *Criminal Law in the Age of the Administrative State*, New York, Oxford University Press

Dempsey, M. M., 2009: *Prosecuting Domestic Violence: A Philosophical Analysis*, Oxford: Oxford University Press

—2011: "Public Wrongs and the 'Criminal Law's Business': When Victims Won't Share", in *Crime, Punishment, and Responsibility,* ed. R Cruft et al, 254-72, Oxford: Oxford University Press

—2020: "The Public Realms: On How to Think About Public Wrongs" *Law, Ethics and Philosophy 7*: 158-173.

Duff, R. A., 2001: *Punishment, Communication, and Community*, New York: Oxford University Press

—2013: "Relational Reasons and the Criminal Law", *Oxford Studies in Legal Philosophy* 2: 175-208.

—2018: *The Realm of Criminal Law*, Oxford: Oxford University Press

Duff, R. A., and Marshall, S. E., 2016: "Civic Punishment", in *Democratic Theory and Mass Incarceration*, ed. A. Dzur et al, 33-59, Oxford: Oxford University Press

Duff, R. A, Farmer, L., Marshall, S. E., Renzo, M., and Tadros, V., 2014: "Introduction: Towards a Theory of Criminalization?", in *Criminalization: The Political Morality of the Criminal Law*, ed. R. A. Duff et al, 1-53, Oxford: Oxford University Press

Dworkin, R. M., 1986: *Law's Empire*, London: Fontana

Farmer, L., 2016: *Making the Modern Criminal Law: Criminalization and Civil Order*, Oxford: Oxford University Press

Gardner, J., 1998: "Crime: in Proportion and in Perspective", in *Fundamentals of Sentencing Theory,* ed. A. J. Ashworth and M. Wasik, 31-52, Oxford: Oxford University Press

—2007: "Complicity and Causality", in Gardner, *Offences and Defences,* 57-76, Oxford: Oxford University Press

—2011: "Relations of Responsibility", in *Crime, Punishment, and Responsibility*, ed. R Cruft et al, 87-102, Oxford: Oxford University Press
Gargarella, R., 2020: "Democratic Dialogue, Multiculturalism, and 'Public Wrongs'" *Law, Ethics and Philosophy 7*: 191-208.
Hörnle, T., 2020: "One Master Principle of Criminalization—Or Several Principles?" *Law, Ethics and Philosophy 7*: 209-221.
Jones, I., 2017: "A Grave Offence: Corpse Desecration and the Criminal Law", *Legal Studies* 37: 599-620
LaFave, W., 2017: *Criminal Law* (6th ed.), St Paul MN: West Academic Publishing
McKenzie, A., 2005: "This Death Some Strong and Stout Hearted Man Doth Choose", *Law and History Review* 23: 279-313.
Moore, M. S., 1997: Placing Blame: *A Theory of Criminal Law*, Oxford: Oxford University Press
Nagel, T., 1986: *The View from Nowhere*, New York: Oxford University Press
Ormerod, D., and Laird. K., 2018: *Smith and Hogan's Criminal Law* (15th ed.), Oxford: Oxford University Press
Perrodet, A., 2002: "The Public Prosecutor", in *European Criminal Procedures*, ed. M. Delmas-Marty and J. R. Spencer, 415-58, Cambridge: Cambridge University Press
Simester, A. P., and Sullivan, G. R., 2005: "On the Nature and Rationale of Property Offences", in *Defining Crimes: Essays on the Special Part of the Criminal Law*, ed. R. A. Duff and S. P. Green, 168-95, Oxford: Oxford University Press
Thorburn, M., 2010: "Criminal Law as Public Law", in *Philosophical Foundations of Criminal Law*, ed. R. A. Duff and S. P. Green, 21-43, Oxford: Oxford University Press
Wenar, L., 2013: "The Nature of Claim-Rights", *Ethics* 123: 202-29.

LAW, ETHICS AND PHILOSOPHY