

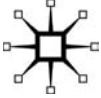
# The Palgrave Handbook of Philosophical Methods

Edited by

Chris Daly

*University of Manchester, UK*

palgrave  
macmillan



Introduction, selection and editorial matter © Chris Daly 2015  
Individual chapters © Contributors 2015

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The authors have asserted their rights to be identified as the authors of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2015 by  
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndsmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN: 978–1–137–34454–0

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

The Palgrave handbook of philosophical methods / [edited by] Chris Daly,  
University of Manchester, UK.

pages cm

Includes bibliographical references and index.

ISBN 978–1–137–34454–0

1. Methodology. I. Daly, Chris, editor.

BD241.P3155 2015

101—dc23

2014049660

12	Advice for Eleatics <i>Sam Cowling</i>	306
13	Pragmatism without Idealism <i>Robert Kraut and Kevin Scharp</i>	331
<b>Part III Intuition, Psychology, and Experimental Philosophy</b>		
14	Intuitions, Conceptual Engineering, and Conceptual Fixed Points <i>Matti Eklund</i>	363
15	Thought Experiments and Experimental Philosophy <i>Joachim Horvath</i>	386
16	Rationalizing Self-Interpretation <i>Laura Schroeter and François Schroeter</i>	419
17	Reclaiming the Armchair <i>Janet Levin</i>	448
<b>Part IV Method, Mind, and Epistemology</b>		
18	Placement, Grounding, and Mental Content <i>Kelly Trogdon</i>	481
19	Theory Dualism and the Metalogic of Mind – Body Problems <i>T. Parent</i>	497
20	Knowing How and ‘Knowing How’ <i>Yuri Cath</i>	527
21	Philosophy of Science and the Curse of the Case Study <i>Adrian Currie</i>	553
22	Three Degrees of Naturalism in the Philosophy of Science <i>Paul Dicken</i>	573
<b>Part V Metaethics and Normativity</b>		
23	Against Pluralism in Metaethics <i>Jens Johansson and Jonas Olson</i>	593
24	Directly Plausible Principles <i>Howard Nye</i>	610
25	Moral Inquiry and Mob Psychology <i>James Lenman</i>	637
26	The Methodological Irrelevance of Reflective Equilibrium <i>Tristram McPherson</i>	652
	<i>Index</i>	675

# 24

## Directly Plausible Principles

Howard Nye

In this chapter I defend a methodological view about how we should conduct substantive ethical inquiries in the fields of normative and practical ethics. I maintain that the direct plausibility and implausibility of general ethical principles – once fully clarified and understood – should be foundational in our substantive ethical reasoning. I argue that, in order to expose our ethical intuitions about particular cases to maximal critical scrutiny, we must determine whether they can be justified by directly plausible principles. To expose apparently plausible principles to maximal critical scrutiny, we must determine whether their direct plausibility can survive careful clarification of what they are really saying. This means that intuitions about cases are useful only in (a) suggesting principles that must stand on their own two feet, and (b) illustrating or otherwise helping us clarify what a principle is really saying. We should not reject principles that seem most directly plausible after we have fully clarified their content simply because they conflict with our intuitions about cases, because to do so is to side with uncritical prejudices over the teachings of critical scrutiny.

### 1 Principles and cases

We often find ourselves unsure about, or in disagreement concerning, the answers to important ethical questions. Familiar examples include questions about the moral permissibility of various cases of abortion, euthanasia, military action, conduct towards non-human animals, and omitting to help the global poor. In some cases our uncertainties and disagreements are primarily empirical, and may be resolved by gathering or more carefully evaluating empirical evidence such as that concerning the nature of fetal and non-human animal psychology, the effectiveness of palliative care, the risks of harm to civilians, and the effects of contributing to aid organizations. But in many cases our uncertainties and disagreements trace to *basic* ethical

uncertainties and disagreements. We may agree about (or stipulate) all of the merely descriptive features of the foregoing kinds of conduct, but still find ourselves uncertain about or in disagreement concerning whether they are permissible.

When we are faced with such basic ethical uncertainties and disagreements, there are two very natural things we do in attempting to reason our way to a resolution. First, we search for ethical principles – which seem plausible to ourselves and our opponents – that seem to support a specific answer. For instance, those arguing that abortion is impermissible might adduce the plausible principle that it is wrong to kill innocent human beings, and those arguing that it is permissible might adduce the plausible principle that people have a right to control their own bodies. Second, in the course of evaluating these principles and their applicability to the case at hand, we think about their implications for other cases, which may seem intuitive or counterintuitive to ourselves and our opponents. For instance, to support the claim that considerations of bodily autonomy can sometimes justify killing innocent human beings, Judith Thomson (1971, 48–9) asked us to consider the case of finding one’s kidneys connected to a famous violinist, where disconnecting from him any time in the next nine months will cause him to die. If it seems to us permissible to disconnect from the violinist in such a case, this appears to give us some reason to doubt that what truth there is in the claim that it is wrong to kill innocent human beings entails that abortion – at least in cases of rape – is impermissible.

In the course of further evaluating such principles and verdicts about cases, we are often led to consider principles of greater generality, which may concern related ethical domains. For instance, many of those who are inclined to think that abortion is typically wrong even in cases of rape are inclined to think it permissible to abort ectopic pregnancies, which will otherwise kill both the mother and the developing fetus. They may also be inclined to think that it would be morally wrong to kill innocent beings who are not human but psychologically identical to humans – like extraterrestrials portrayed in unrealistic science fiction. This may incline them to doubt that the ethical relevance of the fact that an act will kill an innocent human being is fundamental or underivative. Rather, it may be plausibly seen as at least largely derived from the harm it does to the being by depriving her of future goods, together with a general principle of non-maleficence according to which there are weighty moral reasons against inflicting harm (cf. Marquis 1989, 189–92). If this is what determines the strength of the objection to killing – particularly in cases where death cannot frustrate the victim’s autonomously chosen ends – then the moral reasons against an abortion may depend crucially on the extent to which death deprives the fetus of goods that otherwise would have counted ethically as hers. To determine this, we must consider principles

about what matters in survival – whether these are certain psychological relations (Shoemaker 1963; Parfit 1984), certain physical relations (Williams 1970; Marquis 1989), or certain combinations of the two (Unger 1990; McMahan 2002). Especially in cases of fetuses whose later lives may be atypical, such as those with intellectual disabilities, we may also need to consider principles about what makes life more or less of a benefit – whether this is determined by the balance of enjoyment over misery (Bentham 1789; Sidgwick 1907), or in part the extent to which one's experiences and activities are cognitively sophisticated (Mill 1863; McMahan 2002).

In this way, the enterprise of practical or applied ethics, which seeks to resolve our uncertainties and disagreements about concrete ethical issues, blends seamlessly into that of normative ethics, which works to determine the most general basic ethical facts there are about what falls under our various ethical categories.<sup>1</sup> Most philosophers working in the fields of practical and normative ethics use some version of our natural methods of ethical reasoning, adducing and evaluating both plausible ethical principles and intuitions about particular cases. But there are crucial differences in the specific ways they reason about cases and principles. Moreover, there are *radical anti-theorists* who eschew any basic role for principles in our ethical reasoning (see for instance Dancy 2004; 2013).

In this chapter I defend a particular methodological approach to substantive ethical reasoning. Let us say that a general ethical idea or principle is *directly plausible* if it seems true considered simply by itself – in the sense of independent of its inferential relations, such as what else it entails and what else entails it. Many would call such appearances of the direct plausibility of principles *intuitions* that they are true (see for instance Huemer 2005; Singer 2005; Sandberg and Juth 2011). Unfortunately the term 'intuition' has become so associated with directly compelling propositions about particular cases that, simply to avoid confusion, I will speak of finding a principle directly plausible instead of having the intuition that it is true.

The methodological approach I defend maintains that the direct plausibility or implausibility of principles about the ethical relevance of various factors is foundational in normative and practical ethics. This does not mean that appearances of direct plausibility are infallible. Principles often seem plausible only because we are making confusions and do not fully appreciate what they are really saying. On the approach I defend, much of the business of ethical reasoning consists in correcting erroneous appearances of plausibility by clarifying the content of principles, making crucial distinctions, and discovering alternatives with greater direct plausibility. Nor does the claim that the direct plausibility of principles is foundational mean that we should *begin* our ethical reasoning by considering only which principles seem plausible. The principles

that turn out to be most plausible on reflection might be suggested to us only by first considering our intuitions about a variety of cases and then seeing which of them can be justified by principles that, once formulated and clarified, are directly plausible.

What this claim that the direct plausibility of principles is foundational does mean is that:

- (1) The ultimate way to determine whether a basic ethical judgment is correct is to determine whether it can be subsumed under defensible ethical principles,
- (2) An ethical principle is defensible if it is either derived from or a member of a defensible set of axiomatic or fundamental ethical principles, and
- (3) The ultimate way of determining whether a set of fundamental ethical principles is defensible is by determining whether, after careful clarification of their contents, they are more directly plausible than all rival candidate sets of fundamental ethical principles.

On this view intuitions about cases can be useful in ethical reasoning, but only by (a) suggesting principles that must stand on their own two feet as independently plausible, and (b) illustrating or otherwise helping us clarify what a principle is really saying. The mere fact that a principle conflicts with our intuitions about cases is never in itself a reason to reject it. In this respect the approach differs from both that of radical anti-theory *and* many philosophers who recognize the importance of both cases and principles in ethical reasoning.

The approach I defend was, I believe, that of both Henry Sidgwick and W.D. Ross, although they differed significantly in their views about the scope and determinacy of the most general ethical principles that remain directly plausible after they have been fully clarified and understood.<sup>2</sup> Ross and some other early 20<sup>th</sup> century subscribers to this approach seemed to think that it justified commonsense ethical opinions. But I believe that the revolutionary potential of the approach was clearly demonstrated in the early 1970s (as befits revolutionary potential), in Peter Singer's groundbreaking works on "Famine, Affluence, and Morality" (1972), "All Animals are Equal" (1974b), and *Animal Liberation* (1975).

My contention below is that, in order to expose our ethical judgments to maximal critical scrutiny, we must treat the direct plausibility and implausibility of principles as foundational. To expose intuitions about particular cases [henceforth 'case-intuitions' or just 'intuitions'] to maximal critical scrutiny, we must determine whether they can be justified by directly plausible principles. To expose apparently plausible principles to maximal critical scrutiny, we must determine whether their direct plausibility can survive careful clarification of what the principles are really saying, and whether they can be

integrated with other such principles to form a maximally plausible set. But maximally critical methods will not allow us to reject principles that seem most directly plausible after we have fully clarified their content and that of their rivals simply because they do not fit our case-intuitions. To reject such principles for the sake of intuitions is to side with our uncritical prejudices over the teachings of critical scrutiny.

## 2 The insufficiency of considering cases and the relevance of factors in cases

If all we had to go on were our initial case-intuitions, then reasoning in the service of resolving basic ethical uncertainties and disagreements would seem impossible. If there were no ethical regularities, and each judgment had to be made on a case-by-case basis, then what we thought about one case could have no rational bearing on what we should think about any other. We would, for instance, have to dismiss both the intuitive impermissibility of killing older children who were conceived by rape and the intuitive permissibility of unplugging from Thomson's violinist as entirely irrelevant to the morality of abortion in cases of rape. If there were absolutely no regularities in how the features of various cases operated in determining their ethical status, we could not know whether such factors as depriving of someone of all future goods or controlling one's body so as to avoid a substantial burden played anything like the same role in determining whether such abortions are permissible. Consequently, if our initial case-intuitions about such an abortion were uncertain or in disagreement, we would seem to have no rational means to resolve this uncertainty or disagreement.

Radical anti-theorists seek ways of going beyond our initial case-intuitions *without* appealing to explicit principles that might connect our thinking about one case to our thinking about another. Their main suggestion is that we should seek to gain – or defer to individuals who have – more of a certain kind of experience or ability. Dancy (2004; 2013), for instance, suggests that the experience we should seek to gain and trust in the course of ethical reasoning is that of considering many similar cases. Surely this is sound advice. Ethical views that result from considering many related cases instead of just a few seem more critically informed. In practice, this method of exposing our initial case-intuitions to the simple scrutiny of related cases may suffice to resolve many uncertainties and disagreements.

Still, this “method of more cases” does not seem to go far enough. While the bare consideration of more cases can alter the influence certain features have on our intuitions, it may leave in place influences that would be rejected as distortionary by more careful reflection on the nature of the factors themselves. For instance, features like someone's race or sex can influence our judgments



even when we explicitly reject the idea that they should do so (Greenwald and Banaji 1995). While those of us who explicitly reject racism and sexism may try to purge our intuitions of this influence, there are actually two features of this practice which are at odds with the methodological stance of radical anti-theory. First, we do so in the service of our general ethical views, such as the principle that race and sex are utterly devoid of non-instrumental ethical relevance. Second, in determining how to revise our corrupted intuitions, we often compare our intuitions about some cases with our intuitions about others (cf. Schumacher 1996; “now imagine she’s white”). More importantly, the method of more cases does not by itself enable someone to reason her way to seeing these factors as illicit prejudices if she uncritically accepts them as relevant, even if she would reject their relevance if she thought more carefully about what they are.

For instance, the ethical intuitions of an unabashedly racist slaveholder about such questions as when it is permissible to force someone into slavery might be systematically sensitive to race *per se*. If this were pointed out to him, he might initially boast that this was the result of his superior moral discernment, which is evidently more refined and sensitive than that of crude abolitionists, whose ability to perceive the moral relevance of race must have been dulled. But his response could be quite different if it were pointed out to him that all race *per se* really amounts to is the presence of superficial phenotypic traits like skin color and hair texture as a result of one’s area of ancestry. This would probably put him on the defensive, making him fumble for the standard theoretical rationalizations of the time such as slaves’ having lesser intellectual abilities, which would make him vulnerable to Thomas Jefferson’s (1809) observation that “whatever be their degree of [intellectual] talent it is no measure of their rights. Because Sir Isaac Newton was superior to others in understanding, he was not therefore lord of the person or property of others.”

Upon finding it directly implausible that race *per se* – when it is distinguished and clarified – has non-instrumental ethical relevance, our slaveholder could revise his views and see his racist case-intuitions as problematic. Exposing one’s views about the ethical relevance of factors to such direct questions about whether they really should matter in the ways one implicitly or initially takes them to matter seems clearly to be an important method of exposing one’s views to critical scrutiny. Views that result from eliminating the influences of factors which – upon better appreciating what they are – seem indefensible are surely more critically informed than views that uncritically endorse our case-intuitions without asking whether the factors that influence them should matter.

This form of critical scrutiny clearly goes beyond the method of more cases. Moreover, the methodological stance of radical anti-theory appears to rule out such criticism of one’s views about the relevance of factors and one’s systematic

sensitivities to them. The radical anti-theorist seems committed to the view that the directly plausible idea that skin color and area of ancestry lack non-instrumental ethical relevance is precisely the sort of ethical generalization that cannot be had.

The extent to which a radical anti-theory cannot allow direct questions about whether factors should matter in the ways they matter to our case-intuitions may, however, depend upon exactly how radical the anti-theory is. The most radical anti-theory might hold that we must simply perceive the final ethical status of particular acts, and that we cannot criticize these perceptions in terms of any theoretical views about why the act should have this status. Dancy, however, seems to reject this, and to hold instead that we can rely upon our direct views about the exact ethical relevance a factor should have in a given case. So Dancy would seem to allow us to criticize our views by asking whether, once we appreciate what some factor really is, it should have the relevance in a given case that we initially took it to have. For instance, our slaveholder might initially intuit that it was perfectly permissible for some particular individual – say Kunta Kinte – to be enslaved. We might confront the slaveholder with the directly plausible thoughts that the great harm this did to Kunta Kinte and the violations of his autonomy it involved were powerful reasons against it, which could not be justified by the minor gains it conferred on those benefitting from the slave trade. If the slaveholder objected that the relevance of the harm and autonomy violation are disabled by Kunta Kinte's race, we could point out that race is nothing more than superficial phenotype resulting from area of ancestry, and it seems directly implausible that this should matter in itself.

In this way, I believe that Dancy's anti-theory can allow us to criticize our intuitions about the ethical status of particular acts by determining whether propositions about the ethical relevance of certain factors are plausible or defensible. Engaging in such criticism would be a significant advance over the sadly common practice, in both everyday life and philosophy, of uncritically accepting certain intuitions about the final ethical status of acts in particular cases, with little if any thought as to whether the factors present in those cases can, once clarified and understood, plausibly support that status. There are, however, two shortcomings of this restriction of thoughts about the ethical relevance of factors to particular cases. The first is that the most apparently defensible propositions about the ethical relevance of factors have a generality that extends beyond a particular case. Thus the propositions that:

- (G1) skin color and area of ancestry per se are ethically irrelevant, and
- (G2) all else held equal, the fact that an act would harm someone is a moral reason against it

are more directly plausible than the propositions that:

(P1) Kunta Kinte's skin color and area of ancestry do not disable the fact that slavery will harm him from disfavoring enslaving him, and

(P2) all else held equal, the fact that slavery will harm Kunta Kinte disfavors enslaving him.

(P1) and (P2) are surely plausible, but there seems to be nothing special about Kunta Kinte's case that deprives his skin color and area of ancestry of non-instrumental relevance or makes the harmfulness of enslaving him a reason not to enslave him. Consequently, the truths of (P1) and (P2) are most plausibly seen as grounded in the more general truths of (G1) and (G2). Moreover, (P2) might not initially seem so plausible or defensible to someone who, like the slaveholder, is not very sympathetic to Kunta Kinte. The force of (P2) might be apparent to such an individual only by considering (G2) and deriving (P2) from it.<sup>3</sup> Coming to appreciate the force of (P2) in this way seems clearly to be a way of coming to more critically informed beliefs as a result of exposing one's views to scrutiny.

A second and related problem is that it might seem directly plausible that a factor has a certain ethical relevance in a case, and the factor might come to be seen as irrelevant on reflection only when a more general proposition about its relevance is considered. Perhaps the least controversial examples of this are when the factor seems to have a certain kind of relevance primarily because it occurs together (and may be conflated) with another factor of more obvious irrelevance. For instance, I mentioned above the rationalization of slavery in terms of the allegedly lesser intellectual abilities of slaves. This was an empirical falsehood, but there surely were some slaves who happened to be born with mild to moderate intellectual disabilities and consequently had the intellectual abilities that slaveholders thought were typical of all slaves.<sup>4</sup> Suppose that Mildred was one such mildly intellectually disabled slave, who had been kidnapped from her homeland and purchased by our slaveholder, and that he routinely beats her severely for trivial reasons. Now our slaveholder might find it highly plausible that it is permissible for him to beat Mildred because "she's just a dumb slave" – that is, he might find it highly plausible that

(P3) Mildred's lesser intellectual ability *per se* weakens the strength with which the fact that beating her causes her pain disfavors beating her.

But if our slaveholder reflected, he might recognize that – when distinguished from such things as influencing how much pain an act causes someone – intellectual ability *per se* seems quite irrelevant to the strength of our reasons not to cause someone pain. That is, he might find it highly *implausible* that:

(G3) Someone's intellectual ability *per se* affects the strength of our reasons not to cause her pain.

The implausibility of this idea might be illustrated, *à la* Jefferson, by the plausible idea that

(P4) The lesser intellectual ability of the slaveholder's own daughter *per se* would not weaken the strength with which the fact that Isaac Newton's beating her would cause her pain counts against Isaac Newton's beating her.<sup>5</sup>

Appreciating the implausibility of (G3) – once its content is carefully clarified and illustrated by (P4) – could cause our slaveholder to revise his views about the plausibility of (P3). This would seem to be a clear instance of coming to have a more critically informed set of ethical beliefs as a result of having reflected more fully upon whether the factors to which they attribute ethical relevance really should have such relevance. But if Dancy's anti-theoretical stance were correct, then our slaveholder's revising (P3) in light of the general implausibility of (G3) and its illustration (P4) would be terrible reasoning. For according to Dancy, (G3) can be dismissed as false simply because it is general, making its implausibility irrelevant to whether there is anything wrong with (P3).<sup>6</sup>

I believe, however, that Dancy offers no good reasons for denying in this way the evident legitimacy of critically scrutinizing our beliefs about the relevance of factors in particular cases by reference to the plausibility and implausibility of thoughts about the general kinds of differences they should make. Dancy (2004; 2013) argues in favor of the *particularist* view that basic ethical facts cannot be derived from a set of exceptionless propositions about what factors favor and oppose acts without reference to context. But ethicists very often consider general principles about the relevance of factors that are implicitly and even explicitly understood to be *ceteris paribus* or all else held equal. As Little (2001) and Lance and Little (2006) argue, the idea that we should be able to support our claims about the ethical relevance of factors by reference to such *ceteris paribus* principles is perfectly compatible with Dancy's arguments in favor of particularism.

Dancy's rejection of this method of criticism would, moreover, allow us to shelter our prejudicial views about the relevance of certain factors from critical scrutiny by conflating them with other factors that, on reflection, seem clearly irrelevant. The only difference to which our slaveholder might be able to point between his beating Mildred and Newton's kidnapping and beating his daughter might be Mildred's race. However, while our slaveholder pre-reflectively thought race highly ethically relevant, he might now admit, upon realizing that race amounts to no more than things like skin color and area of ancestry, that race is devoid of non-instrumental relevance. Yet race may

still have a powerful influence on his intuitions and incautious judgments of plausibility – just as it does on ours. Consequently, the slaveholder may find both (P3) and (P4) plausible due largely to Mildred’s race – as well, perhaps, as other irrelevant factors irrelevant like bias towards his own daughter.

This might be sufficient to convince the slaveholder that either (P3) or (P4) is mistaken. But it might not be – he might insist that, although there is a danger of racism and personal bias, he is quite confident that it is his accurate ethical insight which tells him that Mildred’s intellectual abilities justify him in beating Mildred, but his daughter’s abilities would not justify Newton in beating his daughter. Moreover, even if the slaveholder agreed that there was a problem with his accepting both (P3) and (P4), how should he determine which to reject? Is he being insensitive to Mildred or biased in favor of his daughter? Critical reflection seems quite clearly to supply the answer by reference to the direct implausibility of (G3) once its content is carefully clarified and distinguished from such thoughts as that to the effect that intellectual ability can influence the pain caused by certain acts. Since someone’s intellectual ability *per se* seems in the abstract to be irrelevant to the case against causing her pain, it is the slaveholder’s judgment that Mildred’s intellectual abilities justify his doing so that needs to give.

### 3 The insufficiency of appeals to “practical wisdom”

As I mentioned above, radical anti-theorists tend to suggest that, in place of thoughts about which principles would be plausible or defensible, we should seek to revise and go beyond our initial case-intuitions by gaining or deferring to individuals who have more of a certain kind of experience or ability. Dancy’s fruitful – but as I have argued insufficient – suggestion that the relevant experience is that of considering many similar cases is not the only proposal. Most others can be understood as variants of the idea that the kind of experiences and abilities we should seek to gain and defer to are those that constitute *phronesis* or “practical wisdom.” Some anti-theorists suggest that the wisdom-making experiences are certain forms of engagement with the arts, or other ways of vividly imagining features of potential ethical significance (Diamond 1991). Others suggest that it is the possession of a set of character traits described by thick concepts like COURAGE, HONESTY, and JUSTICE (Hursthouse 1996).

These do, at least often, seem intended as methodological proposals about how to resolve our basic ethical uncertainties and disagreements in the unconstrained context of normative and practical ethical inquiry. But very often they are proposals about very different issues, such as: how ordinary people in fact make ethical judgments, how ethically ideal people would make ethical judgments, how we should respond to situations in the heat of the moment, what

will most effectively convince people of ethical conclusions the truth of which we have already established or simply stipulated, and under what conditions we should credit people with knowledge of these truths.<sup>7</sup> It is crucial to distinguish these questions from that of the appropriate methodology for normative and practical ethics. For instance, perhaps “the morally perfect person” would effortlessly intuit the basic ethical status of each act without any need for thoughts about generalities – or indeed any other kind of basic ethical reasoning. That’s ducky for the morally perfect person, but for those of us who are trying earnestly to determine the ethical facts of the matter regarding substantive issues about which we are uncertain or in disagreement, it is extremely unhelpful. It would be supreme intellectual arrogance to presume without argument that *we* are the morally perfect person whose intuitions are infallible, or the party whose views are in agreement with this omniscient ethical intuiter.

The main problem with these wisdom-based suggestions as methods of scrutinizing and going beyond our initial case-intuitions is that *which specific* experiences and character traits are wisdom-making is typically just as uncertain and controversial as the first-order uncertainties and disagreements that they are invoked to help resolve. An abolitionist would no doubt think that we can gain wisdom about the ethics of slavery by reading such literature as *Uncle Tom’s Cabin*, and attending vividly to the harms and unfreedoms inflicted upon the slaves. She would dismiss the Anti-Tom literature as crude propaganda, and suggest that focusing on such things as the greater life expectancy of slaves (if this were true) and their phenotypic and cultural differences from whites was distortionary. On the other hand a racist proponent of slavery would no doubt think the opposite – that we can gain wisdom about the ethics of slavery by reading the Anti-Tom literature, and attending vividly to the greater life expectancy of slaves and their phenotypic and cultural differences from whites. He would dismiss *Uncle Tom’s Cabin* as crude propaganda, and suggest that focusing on such things as the harms and unfreedoms inflicted on slaves was distortionary. One who was undecided might try reading and focusing on both sets of books and considerations, but would no doubt come to view as more distortionary whichever influences did not accord with her eventual conclusions.

This is not to say that experiencing the arts and more vividly attending to certain considerations is useless from the point of view of ethical inquiry. Crucial examples and ideas may be suggested to us by the arts, and we may not fully appreciate what a factor amounts to until we attend to it more vividly. But whether our ethical judgments about those examples, ideas, and factors can ultimately withstand critical scrutiny is not something that can be settled simply by experiencing the arts and vividly attending to the factors. The arts furnish important raw materials for ethical arguments, but they are no substitute

for ethical arguments. Vivid attention in the service of understanding what a factor amounts to is an important part of ethical argument, but it cannot on its own secure the defensibility of our ideas about the importance of that factor in relation to all other factors.<sup>8</sup>

Similar remarks go for the method of trying to improve our judgments by developing or deferring to the character traits described by thick concepts. Simply in virtue of her initial judgments about slavery, an abolitionist will think that the maximally just individual or maximally virtuous *phronemos* will be someone who condemns slavery. On the other hand a defender of slavery – like Aristotle himself – will, simply in virtue of his initial views about slavery, think that the maximally just person or *phronemos* will not condemn slavery.

I should emphasize that this does not have to be a problem for *virtue ethics* conceived as the normative ethical doctrine that what is ethically virtuous explains what it is ethically right to do (Hursthouse 1996). A proponent of virtue ethics so conceived is completely free to develop her view of what specific character traits are morally virtuous by whatever methods of ethical reasoning she finds appropriate. I am arguing, in effect, that in order for her account of the virtues to withstand sufficient critical scrutiny, she will need defensible (*ceteris paribus*) principles about what descriptive features of character traits make them virtuous. I see no reason why a virtue ethicist cannot try to articulate such principles – even if many people attracted to virtue ethics happen to be radical anti-theorists who are content to rely uncritically on their own intuitions about virtue and dismiss contrary views without argument as evident products of the “wrong” moral education.

#### 4 The insufficiency of unaided debunking explanations

There is one final method of critically evaluating our intuitions – which might not seem to involve ethical principles – that has received a great deal of recent attention. This is to examine the etiology of our intuitions to see whether it might render them suspect. Some authors, like Singer (2005), suggest that such “debunking explanations” render *all* case-intuitions suspect and in need of vindication by defensible principles. But others have suggested that these debunking explanations might be applied more selectively, allowing us to screen out some case-intuitions while continuing to rely upon others in a more or less foundational way in our ethical reasoning (Huemer 2008; Sandburg and Juth 2011; Liao et al. 2012; McMahan 2013). In this section I consider an extreme form of the second proposal that might be utilized by a radical anti-theorist. This is the idea that we can use selective debunking explanations to critically scrutinize and go beyond our initial case-intuitions *without* having to see whether they can be justified by defensible ethical principles.

Various of our case-intuitions have been empirically shown to be sensitive to the following kinds of influences.

*Framing and Ordering Effects.* How a situation is described or the order in which features are considered can influence our intuitions. For instance, Tversky and Kahneman (1981) found that a hypothetical policy which resulted in 200 out of 600 individuals living was more likely to be chosen if its outcome was described as “200 people being saved” than if it was described as “400 people dying.” Similarly, Liao et al. (2012) found that subjects’ intuitions about whether it is permissible to divert a trolley from killing five individuals to killing one whose being lethally struck is needed to save the five (known as the “loop variant”) are sensitive to which similar cases they considered before making the judgment.

*Emotional Influences.* Our intuitions can be influenced by our emotions. For instance, Schnall et al. (2008) showed that subjects are more likely to make harsher moral judgments about others’ conduct if they are exposed to disgusting stimuli such as bad smells and dirty desks.

*Social Influences.* Our intuitions are sensitive in non-rational ways to the judgments of those around us. Subjects’ intuitions have been shown to shift rather automatically to conform to those of their social groups, interaction partners, and intimates (Newcomb et al. 1965; Chen et al. 1996; Davis and Rusbult 2001). Haidt (2001) argues that this shapes many of our ethical intuitions, and Kuhse and Singer (1985) suggest that the historical influence of Christianity has come by this mechanism to shape contemporary western views about the impermissibility of infanticide.

Another point that is often made is that our intuitions are the result of processes of natural selection that shaped our psychology (Singer 2005; Huemer 2008; Lazari-Radek and Singer 2012). It is, however, not obvious why this feature of the etiology of our intuitions should be a reason to place less confidence in them without constituting an equally strong reason to place less confidence in all of our ethical cognitions, no matter how general or critically informed. For all of our ethical judgments result from our psychological processes, which have been shaped by evolution.

What Singer and Lazari-Radek essentially seem to suggest is that if we can show that some particular ethical judgment was biologically *adaptive* (i.e. such to increase the representation of the genes of its holder in future populations), then it is suspect, but that if it is simply a *spandrel* (i.e. a result of evolution that did not contribute to its holders’ genetic representation), then its evolutionary history casts no doubt upon its truth. But why should spandrels be in better order than adaptations? As Mason (2011) observes, Singer and Lazari-Radek seem to misunderstand the logic of evolutionary debunking arguments. The fact that our visual systems were adaptive is *not* a reason to doubt them because the explanation of why they were adaptive involves their being reliable trackers



of the facts they represent. On the other hand, if the mere fact that some ethical intuition was adaptive *is* a reason to doubt it, this is because the adaptive story does not make reference to its truth – and as Street (2006) argues, the simplest adaptive stories do not seem to make reference to intuitions' tracking mind-independent truths. But here is the problem for Singer and Lazari-Radek: even if some ethical judgments are spandrels, the simplest accounts of how they arose *still* do not seem to make reference to their tracking a mind-independent order of ethical truths – these accounts simply make reference to how our evolved psychologies interacted with our environments to produce the relevant judgments (Harman 1977). So if evolution debunks adaptive intuitions, evolution and social history debunk all ethical cognitions.

The apparent explanatory impotence of mind-independent ethical facts does not seem, however, to debunk all ethical cognitions. Our not seeming to need a particular kind of mind-independent fact to account for our judgments looks quite irrelevant to the reasonableness of trying to resolve our basic ethical uncertainties and disagreements by determining what ethical views can best withstand critical scrutiny. Each metaethical theory will have its own explanation of why this is the case, but almost all would agree that it operates as a constraint on metaethical theorizing.<sup>9</sup> Perhaps it is because ethical facts are simply facts about our own mental structures – not those embodied in our current attitudes, but those that would be manifest in our attitudes after exposing them to maximal critical scrutiny (Brandt 1959; Street 2008). Alternatively, all normative judgments – including both ethical judgments and judgments about their epistemic status – may express non-cognitive attitudes, like our acceptance of certain principles for what to believe, feel, and do (Gibbard 1990; 2003). If this is right, then our claim that ethical views formed by critical scrutiny are more likely to be correct simply expresses our basic commitment to putting more stock in such views, rather than, like Hermann Göring, having a megalomaniacal faith in our own uncritical reactions. Or perhaps there is some other way in which there can be a mind-independent order of ethical truths that, in virtue of their very different nature from scientific truths, need not play any similar role in explaining our ethical judgments (Parfit 2011).

So, whatever the explanation as to why, it seems that we should not treat the *mere* fact that an intuition has an adaptive evolutionary history that does not make reference to mind-independent ethical facts as a reason to doubt it. This seems clearly to be appropriate in the case of intuitions that can be justified by directly plausible principles. For instance, we intuit that we have most practical reason not to do things like hit ourselves with hammers when this would serve no further purpose.<sup>10</sup> Part of the explanation of why we have these intuitions may well be that such judgments were evolutionary adaptations. But it seems that we should not care, because these intuitions can be justified by defensible ethical principles, such as:

(G4) The fact that an act will cause one pain is (at least *ceteris paribus*) a practical reason for one not to perform it.

(G4) seems to have an overwhelming degree of direct plausibility, which seems extremely unlikely to be overturned by further scrutiny of the concepts it involves. This might well be in part because accepting (G4) was adaptive, but this seems irrelevant.

All of this having been said, when we are not sure whether an intuition can be justified by defensible principles, it *does* seem that an explanation of its origin as an adaptation should make us worry. Such worries appear justified, because the adaptive explanation of the intuition decreases the probability that it will be justifiable by defensible principles. The reason for this is that it increases the probability that the intuition has been generated by forces other than a tacit grasp on defensible principles that we cannot yet articulate. This is of course consistent with the possibility that the intuition will prove justifiable by some other defensible principle. But it seems less likely that an intuition should by luck turn out to be justifiable by principles, our grasp on which played no role in its origin, than by principles that played some role in its genesis.

But if this is correct, the selective application of evolutionary debunking arguments actually *presupposes* the view that we should not accept an intuition unless it can be subsumed under defensible ethical principles. This is sufficient to render such selective evolutionary debunking arguments useless to radical anti-theorists who might wish to use them to scrutinize our intuitions in the place of attempts to see whether these intuitions can be justified by defensible principles.

I believe that useful criticisms of intuitions in terms of their resulting from framing and ordering effects and social and emotional influences also presuppose that, in order to be defensible, intuitions must be justified by defensible principles. It might seem that, because intuitions about cases that stem from framing and ordering effects are affected by features like salience to the intuiter, which are not genuine features of the cases, we need presuppose no substantive ethical principles in criticizing them on this basis. There might, however, be a coherent view according to which the actual ethical status of each act is relative to the judge and dependent upon the salience of various features to her. If so, then because such a view would sanction intuitions that result from framing effects, our rejection of these intuitions as problematic presupposes the falsity of this view. But how would we know that the view is false? Surely it would be due to the overwhelming direct implausibility of its principle:

(G6) How salient a feature is to a judge influences the genuine ethical status of acts relative to that judge.

Perhaps more importantly, the sensitivity of intuitions to framing and ordering effects tells us that we are making a mistake, but it does not by itself tell us *which* mistake. We know, for instance, that we should not both accept and reject the policy that results in 200 out of 600 individuals living, but learning about the sensitivity of our intuitions to framing effects does not by itself tell us whether we should accept or reject it. Similarly, we know that it is not both permissible and impermissible to divert the trolley in the “loop variant,” but learning about the sensitivity of our intuitions to ordering effects does not by itself tell us whether it is permissible. Nor can our intuitions about other cases that are uncorrupted by framing and ordering effects help tell us determine which intuition about these cases is mistaken *unless* there are defensible principles that can connect our thinking about these cases to our thinking about others. Consequently, criticizing some intuitions as corrupted by framing and ordering effects will be of little if any use to the radical anti-theorist who wishes to revise and go beyond our initial intuitions without having to rely upon defensible ethical principles.

Similarly, the fact that an intuition is a result of social or emotional mechanisms can be a reason to suspect that it might be mistaken, but it cannot be a very conclusive reason to think that it is in the absence of considerations about whether it can be supported by defensible ethical principles. For instance, views about the impermissibility of infanticide are not the only distinctive legacy that Christianity left to western society. Christian ideas about human equality have with equal likelihood inspired our “deviant” views about the wrongness of slavery and obligations to the poor, unfortunate, and oppressed. While Nietzsche seemed to think that this is a good reason to throw these beliefs over too, the genetic fallacy is so called for good reasons. As Levi (1980, 1) observes, beliefs that turn out to be perfectly defensible are very often “born on the wrong side of the blanket.”

Or consider the intuition that, if you saw a child drowning in front of you, and saving her would cost you nothing but the ruin of your \$200 shoes (‘Pond case’; Singer 1972), it would be morally wrong not to save her. This intuition has very likely been influenced by emotional mechanisms, which may play a role in explaining why we do not similarly intuit that it would be wrong not to give an alliance of aid organizations \$200 which could be expected to do even more life-saving good (‘Envelope case’; Unger 1996). If we are to criticize our intuitions by simply purging those that have been influenced by emotional mechanisms, we should presumably retain our intuition that it is permissible to give nothing in the Envelope case but eliminate our intuition that it would be wrong not to save the child in the Pond case. This might make certain extreme libertarians happy, but such purging of cognitions simply because of their emotional origin would, as Mason (2011) observes, be a commission of what Antonio Damasio (1994) calls “Descartes’ error.” As Damasio argues, our emotions very often help us in

our practical reasoning by drawing our attention to features which would not in their absence receive sufficient attention – like the fact that children really will die if we do not sacrifice our mere luxuries.

Of course, one must somehow determine which socially influenced beliefs are perfectly defensible and which features highlighted by emotion really do deserve the highlighting. The answer as to how we must do this is as obvious as it is unavailable to radical anti-theorists: we must determine which beliefs and intuitions can be justified by defensible ethical principles.

## 5 The supremacy of ethical principles

I have thus argued that, in order to expose our initial case-intuitions to maximal critical scrutiny, we must determine whether they can be justified by defensible ethical principles. As my discussion has suggested, a natural and potentially powerful way to determine whether a principle is defensible is to determine whether it either follows from or is itself among those principles that are most directly plausible once we have fully understood their content. In this section I argue in favor of this criterion of defensibility by contending that appearances of direct plausibility are more foundational than case-intuitions. While we can use case-intuitions to suggest principles and help us understand what principles are really saying, maximally critical methods will not allow us to abandon a principle simply because it conflicts with our case-intuitions.

As I noted above, most normative and practical ethicists are not radical anti-theorists; they agree that there is some important role for principles in our ethical reasoning. But in practice many come close to giving case-intuitions a methodological role that is nearly as central as that given to them by radical anti-theorists, namely that of rock-solid data that our ethical principles must fit. Of course, if we were to literally fit our principles to our case-intuitions *whatever* these case-intuitions happened to be, the principles could play absolutely no role in critically scrutinizing or justifying our intuitions. For any pattern of intuitions whatever, we could find some gerrymandered set of general propositions about the ethical relevance of factors to fit them.

Perhaps the most intuition-centric criterion for when a principle is defensible that retains some critical force is that a principle is defensible if it is part of or derivable from a *best systemization* of our case-intuitions – *viz.* a set of principles that gives us the best trade-off between simplicity and fit with our case-intuitions (cf. Rawls 1951, 184–6; 1971, 46–8; Kagan 1989, 11; Unger 1996, 94). One problem with this best systemization approach is that how simple our ethical principles should be seems to be something that cannot be dictated at the outset of ethical inquiry, but should be determined instead by a more direct examination of which proposals regarding the fundamental ethical relevance of factors will bear scrutiny.

More importantly, the best systemization approach seems clearly to give us insufficient critical purchase on the intuitions being systemized. Such systemization will primarily remove outlying intuitions that are at variance with the general pattern of the bulk of our judgments. But it will leave in place the more pervasive influences of certain factors that we would reject as irrelevant upon a more careful examination of what they amount to. For instance, the best systemization of the intuitions of someone who is deeply racist, sexist, or elitist might leave in place the influences of race, sex, and social status. But, as we have seen, clarification of what these factors really amount to might make them seem directly implausible.

It is difficult to see how a principle could be defensible unless it either follows from or is itself a principle which is to some extent directly plausible (at least once it is clarified and fully understood). Most ethicists seem to accept that a fundamental principle's having some degree of such direct plausibility is a necessary condition for it to be defensible (Rawls 1951, 188–9; 1971, 48; Kamm 1993, 6–7; 2007, 5; McMahan 2013, 112). But there are significant differences among ethicists in how seriously they take this necessary condition, and the role that they allow intuitions to play in selecting among rival candidate principles, each of which have some initial degree of direct plausibility.

Some ethicists seem to pay only lip service to the requirement that principles must be directly plausible, and in practice simply fit their principles to their case-intuitions with very token regard for the direct plausibility of the results. For instance, after acknowledging the requirement of direct plausibility in the introductions to her books, Kamm (1993; 2007) proceeds to spend almost their entirety determining – in an admirably rigorous and systematic way – what principles would fit her intuitions about a great many cases. But the resulting principles are manifestly implausible, attributing complicated kinds of non-instrumental ethical significance to such factors as bare physical distance and the relative directness with which events cause benefits and harms. Although these seem as directly implausible as principles that attribute non-instrumental relevance to skin color, Kamm spends only a few sentences saying extremely hazy things in half-hearted attempts to make them seem plausible (2007, 164–7, 386–7). In some places, she openly admits that she cannot see how even these obscure remarks fit the content of her principles (2007, 166–7).

Along these lines, in order to fit her intuitions about variants of the Pond and Envelope cases, Kamm proposes a principle according to which someone's physical proximity *per se* affects our duties to aid her. In defense of this she suggests that if we think that there are special moral permissions to give our own interests more weight than they have from an impartial perspective, we might as well take physical distance to make a non-instrumental moral difference because "After all, we are locatable beings, positioned at the center of our world in virtue of taking an agent-centered perspective. We also identify

with our means that are locatable" (2007, 387). These remarks shield Kamm's intuitions behind an obscure haze of words rather than successfully show them to withstand critical scrutiny. For almost any pattern of intuitions, we could probably mutter some similarly pseudo-relevant-sounding muddle over them. For instance, if our slaveholder were desperately trying to defend the non-instrumental relevance of skin color to the permissibility of kidnapping someone, he might say, "rights not to be kidnapped are located in the person. But after all, persons are colored beings, colored by and coloring the rights they have. We also identify with those of our color."

I believe that some of the considerations that motivate Kamm's method are laudable, and that she seems easy to criticize only because she pursues this method more rigorously and successfully than others who are in effect pursuing it. Kamm is rightly motivated by the excellent point that we may not have a very good idea what the most defensible principles might be until we examine those suggested by our intuitions in a great variety of cases. Where Kamm seems to go wrong is in her confidence in the final victory of her case-intuitions, no matter how badly her principles seem to be losing their battles for direct plausibility. Just as we have no reason to think that initial appearances of the plausibility of principles will always withstand careful scrutiny of their contents, we have no reason to expect that our case-intuitions will always be justifiable by principles the plausibility of which withstands such scrutiny.

Probably most normative and practical ethicists take the requirement that principles be directly plausible somewhat more seriously than Kamm, and would describe themselves as following the method Rawls (1971, 19–21, 48–52) called that of seeking a 'reflective equilibrium' between our case-intuitions and those principles that seem to us directly plausible. This method holds that (i) to be defensible fundamental principles must have some degree of direct plausibility, (ii) we should often revise our case-intuitions to fit such principles, but also (iii) in choosing among such principles we may sometimes look to our case-intuitions. There are, however, two very different ways in which we can use case-intuitions to influence our choice among principles that seem initially plausible, which tend not to be clearly distinguished.

The first method is to treat the fact that a principle fits more case-intuitions as a reason *in itself* to accept the principle. This can result in our accepting principles that are less directly plausible but better aligned with our case-intuitions over principles that are more directly plausible but more at odds with them.<sup>11</sup> The problem with this, however, is that such sacrifices of principles that seem more directly plausible after careful clarification to mere case-intuitions seems to allow our less critically informed, systematic biases to overwhelm what critical scrutiny has revealed to be more defensible. For instance, consider our slaveholder's great-grandfather, who was an English aristocrat. This aristocrat

often beat the peasants who lived on his large rural estate for trivial reasons, but was extremely punctilious in his treatment of other aristocrats. He found direct plausibility in the following principle:

(G7) The fact that someone is not a gentleman (or lady) *per se* weakens one's moral reasons against beating him.

We might ask our aristocrat to think more carefully about what this is really saying – that to be a gentleman is simply something like to have been born to certain parents, had a certain education, and have certain speech patterns and other mannerisms. Why should who someone's parents were, how educated he is, and the aesthetic qualities of his mannerisms have any bearing in themselves on our moral reasons not to cause him pain? Our aristocrat might see the force of this, finding the following more directly plausible:

(G8) All else held equal, the strength of our non-instrumental moral reasons not to harm someone depend on the degree of harm. Who his parents were, how educated he is, and his aesthetic qualities *per se* are irrelevant to our moral reasons not to harm him.

Our aristocrat might, however, still find *some* degree of plausibility in (G7), even after it has been clarified to him what being a gentleman amounts to. Moreover, (G7) – perhaps together with some modification of (G8) that accommodates it – would surely fit many of his intuitions much better than (G8) in its unmodified form. If the aristocrat were to follow the practice of rejecting principles simply to fit case-intuitions, his reasoning might look like the following:

If we had special obligations to someone's parents, and they had special obligations to him, perhaps they transfer by a kind of transitivity. Also someone's greater education and more refined aesthetic properties make him cooler and thus in some sense "better" – and we may owe more to those who are better. I grant that it's hard to see why we would have special obligations to every gentleman's ancestors, and the transference idea looks a little hokey. I also grant that the kind of "betterness" here seems pretty orthogonal to our reasons not to harm someone. So I grant that (G8) seems much more luminously self-evident than (G7). But many of my strongest case-intuitions suggest that it matters morally whether someone is a gentleman, and Rawls said that I could work from both ends.

If we could be just a little less smug, I think that we could see that our favoring less directly plausible principles simply because they fit our case-intuitions is equally an instance of our letting our less critically informed prejudices get the better of us.

There is, however, a second way of using case-intuitions to help select principles that seem initially to be directly plausible. This is to use the cases to help us understand what the principles are really saying, and to illustrate the genuinely direct plausibility or implausibility of the principles. For instance, when Singer (1972, 231) first introduced the Pond case, he did so only to illustrate the direct plausibility of the following principle regarding the obliging force of reasons of beneficence:

(G9) All else held equal, if you can prevent someone from suffering a serious harm at relatively trivial cost to yourself, it is morally wrong not to do so.<sup>12</sup>

The point of the Pond case was not to show that (G9) fits our case-intuitions, but rather to give us a more concrete understanding of how plausible the idea contained in (G9) really is – in itself and independent of what else it supports or entails. In considering the Pond case, we can see how the fact that we would lose \$200 looks like a terrible justification for not saving the drowning child – precisely because the harm to the child is so great and the cost to us is so trivial. By giving a concrete illustration of what relatively trivial costs really look like in relation to serious harms when we pay sufficient attention to them, the Pond case helps us better appreciate what (G9) is really saying, in a way that amplifies its *direct* plausibility. After we have clarified (G9) and its plausibility in this way, its conflicts with other intuitions, like that to the effect that it is permissible to spend money on luxuries rather than donate them to aid organizations, seem to count much more obviously against the intuitions than against the principle.

The contrast between using cases as data that principles should fit and using cases to illustrate and clarify principles may be even clearer when it comes to the criticism of initially plausible principles. For instance, in thinking about whether there is a morally relevant difference between Pond and Envelope, an initially plausible difference seems to be that in Envelope there are many other individuals who could help, and if everyone helped the serious harm could be prevented at much less cost to us (Singer 1972, 233; Unger 1996, 39; Murphy 2000). Along these lines, something like the following principle might seem plausible:

(G10) You cannot be morally required to do more at greater cost to yourself just because others are not doing their fair share – even if the additional cost would be trivial in comparison to the significant harm you would prevent.

Singer and Unger, however, consider a case like the following:

*Multiple Children.* One child is drowning near you and another is drowning across the stream near someone else. If you were both to wade in and save



the child near you, you would each ruin only your \$200 shoes. But the other person walks away. Now, if you save the child across the stream after saving the child near you, you will ruin your \$200 designer jacket in addition to your \$200 shoes.

(G10) implies that it would be permissible to save only the child near you and then walk away, allowing the second child to drown. But this seems worse than counterintuitive. I do not believe that Singer and Unger are suggesting that we should reject (G10) simply because it conflicts with our case-intuition about Multiple Children. They are instead using the case to help illustrate more clearly what it really means to think that the unfairness of slightly greater trivial costs to morally conscientious helpers is a more important consideration than the fact that someone will die if these trivial costs are not incurred. The fact that it is unfair for one to have to save the second child because the other person walked away seems like a horrible justification for walking away oneself. Multiple Children thus illustrates the directly implausible narcissism and pettiness embodied in (G10), and in this way diminishes its direct plausibility. Consequently, after we have considered Multiple Children, (G10) seems in itself to be a much poorer justification of the alleged permissibility of not giving aid in cases like Envelope.

I believe that this stands in stark contrast to the way our intuitions about Envelope conflict with (G9) and the aristocrat's intuition about the permissibility of beating his peasants conflicts with (G8). Our initial intuition that it is permissible not to give to aid organizations does not seem to diminish the direct plausibility of the idea that we are required to make relatively trivial sacrifices to prevent others' suffering serious harm. The case does illustrate such things as how (G9) will make demands on *us*, but none of these things seem to diminish (G9)'s direct plausibility. This, I think, is why it seems like self-serving rationalization to be more hesitant to accept (G9) in light of its implications about Envelope.<sup>13</sup> Similarly, our aristocrat's intuitions about the permissibility of beating his peasants did nothing to diminish the direct plausibility of the idea that such things as who someone's parents were and one's aesthetic properties *per se* are irrelevant to our moral reasons not to harm him. Those intuitions do show how (G8) conflicts with many of his assumptions, but they do not seem to reveal anything inherently implausible about (G8)'s points regarding the lack of plausible connection between who someone's parents were and how cool he is on the one hand and the morality of inflicting harm on him on the other. This, I think, is why the Aristocrat's abandoning (G8) in light of its implications about beating his peasants looks like his allowing his uncritical prejudices to overwhelm what critical scrutiny has taught him.

## 6 Conclusion

I have thus argued that, while case-intuitions can help suggest and clarify the content of principles, it is the direct plausibility of the principles, once so proposed and clarified, that must be taken as foundational. To rest content with less directly plausible principles simply because they better fit our case-intuitions would be to side with our less critically informed prejudices over our more critically informed determinations. We would be like slaveholders or aristocrats who found the direct plausibility of the ethical relevance of race or class questionable on reflection, but concluded that these factors must be relevant because they influence their case-intuitions. We are of course confident that our unreflective responses tend to be more reliable than those of these characters. But we cannot justify this confidence simply by claiming that we were raised in what we regard as “more enlightened” societies if our sole evidence for the greater enlightenment is that our societies’ ethical views better comport with our unreflective responses. Moreover, just like the slaveholder and aristocrat, we cannot know which features of our intuitive responses still constitute unenlightened prejudices until we gain the enlightenment of seeing whether they can be justified by principles that seem most directly plausible after we have fully clarified their content.

Most philosophers working in normative and practical ethics acknowledge that ethical principles should be directly plausible. But it is extremely common to see much effort devoted to investigating what principles fit our case-intuitions and relatively little effort devoted to determining whether these principles are directly plausible – let alone *most* directly plausible – upon careful inspection of what they are really saying. In some respects a good deal of attention to the question of which principles are suggested by our case-intuitions makes perfect sense, as this may be the only way to discover the principles that ultimately prove most directly plausible. But if ethicists wish, as surely they do, not only to generate candidate principles but to actually defend substantive conclusions, I believe that many of them need to spend a bit more energy on the tasks of (i) explaining what directly plausible ideas are embodied in their principles, and (ii) showing these ideas to be more directly plausible than their rivals after we have gotten as clear as possible about precisely what they are saying.<sup>14</sup>

## Notes

1. While I speak freely of ethical facts and truths, I am not assuming the truth of cognitivism. I am assuming only that if non-cognitivism is true, it is a sufficiently quasi-realist version so as to allow me to speak coherently of ethical facts and truths in *some* minimal sense, in which to call something an ethical fact or truth is to express some non-cognitive attitude.

2. For this reading of Sidgwick, see Singer (1974a) and Hurka (2014). In addition to Sidgwick and Ross, Hurka identifies G.E. Moore and H.A. Prichard as subscribers to this approach.
3. Of course, the slaveholder might actually infer (G2) from his intuitions about other cases – for instance about how it is permissible to treat various individuals who are not slaves – before inferring (P2) from (G2). In section 5, I consider two very different ways of “inferring” general principles from case-intuitions, and argue that while we may use case-intuitions to suggest principles and clarify their content, we should treat the direct plausibility of principles – once so suggested and clarified – as foundational. My primary aim at this point is simply to argue against the radical anti-theoretical view that we need not invoke general principles at any stage in our basic ethical reasoning.
4. Slaveholders presumably did not think slaves less intellectually able than this, lest they think them unfit for performing the tasks set to them. Moreover, the bans on teaching slaves to read and write would have been superfluous if slaveholders had not thought them capable of these abilities.
5. Of course, if our slaveholder’s daughter has typical intellectual abilities, she might have correspondingly lesser intellectual ability than Newton but not lesser intellectual ability in an absolute sense. But we could imagine our slaveholder to have already thought about such things as the arbitrary nature of setting a baseline for protected intellectual ability at our absolute level as opposed to Newton’s.
6. Dancy (2013, §7) even suggests that according to some anti-theorists, (P3) and (P4) may both be in good order, even if one cannot point to a relevant difference between them. I am not sure if Dancy means to endorse this last suggestion, but it seems to be a glaringly unacceptable way of allowing people to shelter their uncritical prejudices from critical scrutiny.
7. These other issues seem primary in some of Dancy’s (2004; 2013) discussions, as well as those of Diamond (1991) and Hursthouse (1996).
8. Diamond (1991, 301–4) considers the question “how else can we judge the strength of a moral view [except by argument],” and asserts that it is “not an unanswerable question.” But her answer seems extremely evasive. She suggests that “the shoddy thought can be shown up by being placed alongside the genuine,” but if after such placing we are uncertain which thought is genuine, or wish to explain why, I see no substitute for careful analysis and argument. She also suggests that “a work may [by means other than philosophical argument] invite the reader to elaborate and develop a way of looking and to respond critically to it then as a possibility.” This seems very similar to my suggestion that the arts can be a source of ethical ideas that we can go on to critically evaluate. It is, however, unclear how the most careful critical evaluation can take place except by means of philosophical arguments that seek to determine which general proposals about the relevance of some factors as against others are most directly plausible once we have made them as precise and clear as possible.
9. As Parfit (2011) observes, while some authors are willing to be error theorists about morality in particular, the epistemological and metaphysical considerations that lead to such error theory seem equally to support error theory about all practical reasons, and indeed a kind of self-defeating error theory about epistemic reasons.
10. I understand the category of ethics broadly, to include judgments about what there is practical reason for us to do. Even if one is inclined to use the word ‘ethics’ more narrowly, the epistemic and methodological issues seem to be the same when it comes to questions about what there is basic practical reason to do.

11. Examples of this may include McMahan's (1993, 279; 2013, 115) attitudes towards principles like the Doctrine of Doing and Allowing [DDA] (that there are much stronger moral reasons against doing harm than allowing harm) and the Doctrine of Double Effect [DDE] (that there are much stronger moral reasons against doing or allowing harm as a means than as a side-effect). He seems to concede that these principles, once carefully clarified, have less plausibility than the view that the prevention of serious harm outweighs the importance of such factors as doing vs. allowing and means vs. side-effects. I am not sure if he proposes to retain the DDA and DDE even if all future attempts to show them to be more plausible fail, but if he is, these would be particularly explicit examples of a philosopher retaining what he regards as less plausible principles because they better fit his case-intuitions.

(In McMahan's case there does seem to be another strand of argument, namely that unless we accept something like the DDA or DDE we cannot be said to be accepting a *moral* view. But, as Smart (1956) observed, the main thing most of us care about in ethics is determining what there is basic reason to do, and we do not particularly care whether the reasons we identify get called 'moral'. In any event I cannot think of any reasons that have a better claim on being distinctively moral than fundamental reasons of beneficence to promote the well-being of others. How could one's simply coming to think of reasons of non-maleficence as a species of such reasons (viz. to promote others' well-being passively, by not decreasing it) entail that one has ceased to think in moral terms?)

12. Singer did not actually speak of relative triviality but of 'comparable moral importance' and 'no moral importance', which I think are more opaque. But I believe that, somewhere between Singer's two explicit formulations, (G9) was essentially intended as the logically weaker idea on which his argument depends. The logically stronger idea, on which the bulk of Singer's argument does not depend, is that if you can prevent serious harm by incurring anything less than a roughly equal harm to yourself, it is wrong not to do so. I do not believe that *this* principle has much if any direct plausibility at the outset of inquiry. If it is true, its truth can only be established by the kind of detailed undermining of the initial direct plausibility of agent-centered ideas of the kind undertaken by Kagan (1989) and Bennett (1995).
13. The mere fact that one is inclined to reject a principle upon seeing certain of its implications, which would be contrary to one's interests, is insufficient to make the inclination seem like self-serving rationalization. For instance, being more inclined to reject the view that intellectual ability *per se* affects one's rights because of what it would imply about Isaac Newton's rights to enslave you would not seem like such rationalization.
14. This chapter has benefitted from discussions over the years with too many people to remember or name. I am particularly grateful to John Ku, Alastair Norcross, and David Plunkett for especially helpful and formative discussions, and to Chris Daly for very helpful feedback on an earlier draft.

## References

- Bennett, J. (1995) *The Act Itself* (Oxford: Oxford University Press).  
 Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation* (London: T. Payne and Son).  
 Brandt, R. (1959) *Ethical Theory* (Englewood Cliffs, NJ: Prentice Hall).

- Chen, S., D. Shechter, and S. Chaiken (1996) 'Getting at the Truth or Getting Along: Accuracy- Versus Impression- Motivated Heuristic and Systematic Processing', *Journal of Personality and Social Psychology*, 71, 262–75.
- Dancy, J. (2004) *Ethics without Principles* (Oxford: Clarendon Press).
- (2013) 'Moral Particularism', in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2013 Edition, URL = <<http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>>.
- Damasio, A. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: G.P. Putnam's Sons).
- Davis, J., and C. Rusbult (2001) 'Attitude Alignment in Close Relationships', *Journal of Personality and Social Psychology*, 81, 65–84.
- Diamond, C. (1991) 'Anything But Argument?', in C. Diamond, *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind* (Cambridge, MA: MIT Press), pp. 291–308.
- Gibbard, A. (1990) *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press).
- (2003) *Thinking How to Live* (Cambridge, MA: Harvard University Press).
- Greenwald, A., and M. Banaji (1995) 'Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes', *Psychological Review*, 102, 4–27.
- Haidt, J. (2001) 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review*, 108, 814–34.
- Harman, G. (1977) *The Nature of Morality: An Introduction to Ethics* (New York: Oxford University Press).
- Huemer, M. (2008) 'Revisionary Intuitionism', *Social Philosophy and Policy*, 25, 368–92.
- Hursthouse, R. (1996) 'Normative Virtue Ethics', in R. Crisp (ed.) *How Should One Live?* (Oxford: Clarendon Press), pp. 19–36.
- Hurka, T. (2014) *British Ethical Theorists from Sidgwick to Ewing* (New York: Oxford University Press).
- Jefferson, T. (1789) Letter to Henri Gregoire, February 25, 1809.
- Kagan, S. (1989) *The Limits of Morality* (Oxford: Clarendon Press).
- Kamm, F. (1993) *Morality, Morality: Volume I: Death and Whom to Save From It* (New York: Oxford University Press).
- (2007) *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (New York: Oxford University Press).
- Kuhse, H. and P. Singer (1985) *Should the Baby Live? The Problem of Handicapped Infants* (New York: Oxford University Press).
- Lance, M. and M. Little (2006) 'Particularism and Antitheory', in D. Copp (ed.), *The Oxford Handbook of Ethical Theory* (Oxford: Oxford University Press), pp. 567–94.
- Lazari-Radek, K., and P. Singer (2012) 'The Objectivity of Ethics and the Unity of Practical Reason', *Ethics*, 123, 9–31.
- Levi, I. (1980) *The Enterprise of Knowledge* (Cambridge, MA: MIT Press).
- Liao, S., A. Wiegmann, J. Alexander, and G. Vong (2012) 'Putting the Trolley in Order: Experimental Philosophy and the Loop Case', *Philosophical Psychology*, 25, 661–71.
- Little, M. (2001) 'On Knowing the "Why": Particularism and Moral Theory', *Hastings Center Report*, 31, 32–40.
- Marquis, D. (1989) 'Why Abortion is Immoral', *The Journal of Philosophy*, 86, 183–202.
- Mason, K. (2011) 'Moral Psychology and Moral Intuition: A Pox on All Your Houses', *Australasian Journal of Philosophy*, 89, 441–58.
- McMahan, J. (1993) 'Killing, Letting Die and Withdrawing Aid', *Ethics*, 103, 250–79.
- (2002) *The Ethics of Killing: Problems at the Margins of Life* (New York: Oxford University Press).

- (2013) 'Moral Intuition', in H. LaFollette and I. Persson (eds) *The Blackwell Guide to Ethical Theory*, Second Edition (Malden, MA: Blackwell Publishing Ltd.), pp. 103–20.
- Mill, J.S. (1863) *Utilitarianism* (London: Parker, Son, and Bourn).
- Murphy, L. (2000) *Moral Demands in Nonideal Theory* (Oxford: Oxford University Press).
- Newcomb, T., R. Turner, and P. Converse (1965) *Social Psychology* (New York: Holt, Rinehart, and Winston).
- Parfit, D. (1984) *Reasons and Persons* (Oxford: Oxford University Press).
- (2011). *On What Matters*, Two Volumes (Oxford, Oxford University Press).
- Rawls, J. (1951) 'Outline of a Decision Procedure for Ethics', *The Philosophical Review*, 60, 177–97.
- (1971) *A Theory of Justice* (Cambridge, MA: Harvard University Press).
- Sandberg, J., and N. Juth (2011) 'Ethics and Intuitions: A Reply to Singer', *The Journal of Ethics*, 15, 209–26.
- Schnall, S., J. Haidt, G. Clore, and A. Jordan (2008) 'Disgust as Embodied Moral Judgment', *Personality and Social Psychology Bulletin*, 34, 1096–109.
- Schumacher, J., dir. (1996) *A Time to Kill* (Warner Bros.) (Film.)
- Shoemaker, S. (1963) *Self-Knowledge and Self-Identity* (Ithaca: Cornell University Press).
- Sidgwick, H. (1907) *The Methods of Ethics*, Seventh Edition. Reprinted 1981 (Indianapolis: Hackett Publishing Company).
- Singer, P. (1972) 'Famine, Affluence, and Morality', *Philosophy and Public Affairs*, 1, 229–43.
- (1974a) 'All Animals Are Equal', *Philosophical Exchange*, 1, 103–16.
- (1974b) 'Sidgwick and Reflective Equilibrium,' *The Monist*, 58, 490–517.
- (1975) *Animal Liberation* (New York: Harper Collins).
- (2005) 'Ethics and Intuitions', *The Journal of Ethics*, 9, 331–52.
- Smart, J.J.C. (1956) 'Extreme and Restricted Utilitarianism', *Philosophical Quarterly*, 6, 344–54.
- Street, S. (2006) 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies*, 127, 109–66.
- (2008) 'Constructivism About Reasons' in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics*, vol. 3, pp. 207–45.
- Thomson, J.J. (1971) 'A Defense of Abortion', *Philosophy and Public Affairs*, 1, 47–66.
- Tversky, A., and D. Kahneman (1981) 'The Framing of Decisions and the Psychology of Choice', *Science* (New Series), 211, 453–8.
- Unger, P. (1990) *Identity, Consciousness, and Value* (New York: Oxford University Press).
- (1996) *Living High and Letting Die: Our Illusion of Innocence*. (New York: Oxford University Press).
- Williams, B. (1970) 'The Self and the Future', *The Philosophical Review*, 79, 161–80.