

DOING OUR BEST FOR HEDONISTIC UTILITARIANISM REPLY TO CRITICS

PETER SINGER

Princeton University
University Center for Human Values
School of Historical and Philosophical Studies
University of Melbourne
psinger@princeton.edu

KATARZYNA DE LAZARI-RADEK

Instytut Filozofii UŁ
Uniwersitet Łódzki
kasialazari@gmail.com

ABSTRACT

In this response to the essays by Crisp, Parfit, Hooker and Nakano-Okuno on our *The Point of View of the Universe*, we focus on the following topics: whether egoism is more susceptible to an evolutionary debunking argument than universal benevolence; our defence of impartial rationalism; wide and narrow definitions of “ethics”; the role of moral rules; the extension of ethics to all sentient beings; how best to define and understand pleasure as an intrinsic value; and whether Ross’s ethic of prima facie duties is as defensible as utilitarianism.

KEYWORDS

Egoism, universal benevolence, ethics, impartial rationalism, rules, animals, pleasure, utilitarianism, Sidgwick, Ross, Crisp, Parfit, Hooker, Nakano-Okuno.

1. INTRODUCTION

In the preface to *The Point of View of the Universe* we wrote that our intention was to make the strongest possible case for the views that Sidgwick defended, while recognizing that there would remain many points at which one could object to his form of utilitarianism. Whatever the ultimate verdict on classical utilitari-

anism would be, we said, “there will be much to be learned from a discussion of its strengths as well as its weaknesses.” The four excellent essays to which we here respond show that whatever flaws our arguments may have, we at least got this right. We thank the authors for their careful attention to our work. We are especially grateful to Bart Schultz for having made this symposium happen, and to the editors of *Etica & Politica/Ethics & Politics* not only for publishing it, but for doing so at a speed that is, for an academic publication, extraordinary.

We begin with what we always expected would be the most controversial claim in our book: that all reasons for action are impartial. We support this claim by arguing that the principle of universal benevolence withstands evolutionary debunking in a way that egoism – the position Sidgwick saw as its most formidable rival – does not. Crisp, Hooker and Parfit discuss this argument. Parfit agrees with us that the debunking argument has some force but is not decisive. His discussion therefore focuses more on the plausibility of our claim that all reasons for action are impartial, independently of the debunking argument. Crisp and Hooker, on the other hand, attack the argument. We shall begin by discussing their objections to it and then consider whether it is plausible to hold that all reasons for action are impartial.

2. CRISP AND HOOKER ON THE EVOLUTIONARY DEBUNKING OF PARTIAL REASONS FOR ACTION.

Crisp accepts our claim that the principle of universal benevolence is not debunked by evolutionary arguments of the kind that Sharon Street invokes in seeking to reject the possibility of objective truth in ethics. Universal benevolence survives such arguments and so could be a moral truth that we can grasp by the use of reason. Crisp denies, however, that this insight enables us to resolve Sidgwick’s dualism of practical reason, because, in his view, “egoism is in as strong a position as universal benevolence in this context to resist debunking evolutionary arguments.” Evolution would, he points out, lead us to expect neither egoism nor universal benevolence, but something more like kin altruism. That view is summed up in J.B.S. Haldane’s quip when asked if he would lay down his life to save his brother: “No, but I would lay down my life to save two of my brothers, or eight of my cousins.” Haldane was referring to the fact that we share, on average, half of our genes with our brothers, and one-eighth of them with our cousins (assuming they are full brothers, and first cousins), and that therefore, the sacrifice

of his own life for the numbers of relatives specified would not cause his genes to be less likely to survive. Haldane might also have said he would lay down his life for two of his children, but Haldane's reference to siblings and cousins better suits the purposes of contrasting egoism and kin altruism, because when we speak of sacrificing for one's own children, it is less clear that this really is a sacrifice of one's own interests, given that commonly the wellbeing of parents is more closely tied to the wellbeing of their children.

Here we disagree with Crisp when he asserts that egoism "requires lack of concern for our own children (if we even have any), and a complete unreadiness to sacrifice anything, even something trivial in a case where it might prevent great suffering or death to those children." As parents – and knowing that Crisp is a father too – we are surprised by this comment. Our lives have been greatly enriched by having children. Even if we were to consider only our own interests, we would never agree that having them made us worse off than we would otherwise have been. Others may, of course, have different experiences, and we accept that there is some evidence to the contrary,¹ but in any case, once one has children, any normal parent will love them, be concerned about their wellbeing, and will unhesitatingly make sacrifices – and not only trivial ones – to save them from "great suffering or death." This does not hold, or holds to a lesser extent, for siblings and cousins, because some people, especially those from large families, have first cousins who are virtual strangers to them, and for whom they care very little. Strictly speaking, if we were following the rule "propagate your genes," Haldane would not have had to lay down his life for just two siblings or eight first cousins, because that is the break-even point – the rule would not forbid such a sacrifice, but would require it only for three siblings or nine first cousins. Egoism, on the other hand, would not require any such sacrifice, at least for those who do not care for their siblings or cousins. So we grant that "Do whatever is in your own interests" does not always prescribe the same actions as "Do whatever will best propagate your genes" and egoism is not the principle for which evolution would be expected to select.

Nevertheless, we remain unconvinced that egoism and universal benevolence are, as Crisp maintains, equidistant from what evolution would select for, and that

¹ Empirical research on whether people with children are more or less happy than those without children is inconclusive, but at least it does not confirm the contrary view stated by Crisp. See, for example, Angus Deaton and Arthur Stone. "Evaluative and Hedonic Wellbeing Among Those With and Without Children at Home" *Proceedings of the National Academy of Sciences of the United States of America* 111 (4) (2014), pp.1328-1333.

therefore when it comes to resisting evolutionary debunking arguments, “egoism is in as strong a position as universal benevolence.” The difference is that furthering one’s own interests will normally also increase one’s evolutionary fitness, whereas impartial benevolence towards strangers will normally reduce it. Here we can only seek to reinforce a point made in our book (p.194) and already quoted by Brad Hooker in his contribution to this symposium.² Saving one’s own life preserves the possibility of doing many things that enhance one’s reproductive fitness: assisting one’s children or siblings, having more children, and so on. Acting in one’s own interests against the interests of one’s kin therefore almost always involves, from the perspective of evolutionary fitness, a balancing of conflicting factors, where a complex and difficult to estimate set of probabilities will determine whether the course of action that will most enhance evolutionary fitness is to do what is in one’s own interests or to assist kin. There is no such balancing when it comes to impartially advancing the interests of strangers, at some cost to oneself, for that will seldom be the action that would do most to enhance one’s reproductive fitness.

We therefore disagree with Crisp when he writes:

We would expect evolution to produce *some* concern for others, and universal benevolence can be seen as an extension of that concern in the impartial direction in just the way that egoism might be taken to be an extension of concern for oneself in the direction of partiality.³

We would expect evolution to produce concern for ourselves, for our kin, for those with whom we can form mutually beneficial cooperative relationships, and concern for others in the small face-to-face social groups, typically of not more than 200 individuals, in which humans and our non-human primate ancestors spent most of our evolutionary history. We would not expect it to produce any concern for those outside our social group with whom there is no expectation of cooperation. On the other hand we would expect evolution to produce beings with a strong desire to survive and reproduce, and to value whatever will increase their chances of doing so. Thus, except in the special circumstances of a conflict between one’s own interests and the interests of one’s kin, pursuing one’s own interest would always be conducive to evolutionary fitness; advancing the interests of humans other than those in the categories mentioned would not be.

² Brad Hooker, “Wrongness, Evolutionary Debunking, Public Rules,” this issue, p. 141 above.

³ Crisp, p. 126.

Hooker appears to accept our “important insights” as he calls them, but then draws our attention to commonsense morality, at least some aspects of which are also, he suggests, immune from evolutionary debunking arguments. As an example, he offers the intuition that, when other things are equal, “one should provide a *large* benefit to an old and now infertile relative rather than a *small* benefit to a young and fertile relative.” We grant that such a principle is not readily explicable in evolutionary terms, but that fact, far from undermining our argument, supports it. For in seeking to show that some principles of commonsense morality are not readily explicable in evolutionary terms, Hooker has selected a principle that is derivable from the principle of universal benevolence, that is, from the very principle that we are seeking to show is not explicable in evolutionary terms, and therefore should be seen as a moral truth we grasp by using our capacity to reason. The principle of universal benevolence tells us that, when other things are equal, we should always provide a large benefit to one person rather than a small benefit to another.

Hooker would need to show, to make out his argument for the independent validity of his principle, that it essentially depends on the fact that it involves relatives, or that it involves two people, one old and infertile and the other young and fertile. He might, for example, modify the principle by adding a third option, so that the principle reads: “one should provide a *large* benefit to an old and now infertile relative rather than a *small* benefit to a young and fertile relative, or an *even larger* benefit to a stranger .” But now we have a principle that has another possible explanation: we learn our moral principles from our parents and elders, most of whom are also parents. It would not be surprising that the morality we absorb from them should be distorted in the direction of favoring their interests.⁴

For these reasons, we continue to hold that the principle of universal benevolence is less likely to have a non-truth tracking explanation than either egoism, or those principles of commonsense morality that lead us to favor our kin, whether the kin be young and fertile or old and infertile.

3. CRISP AND PARFIT ON IMPARTIAL RATIONALISM

Crisp correctly points out, following Sidgwick, that the availability of an evolutionary explanation for a principle does not necessarily overturn it. He gives the example of our capacity to do arithmetic, which doubtless has an evolutionary

⁴ Cf. Thrasymachus, in Plato, *The Republic*, 338c.

explanation, but that does not cause us to doubt that $7+5=12$. That's true, but the example doesn't offer strong support for taking the same view of moral principles, because if our ancestors had evolved a capacity for mathematics that was not truth-tracking, it would not have helped them to survive. As Street has pointed out, with moral principles, there is no evolutionary advantage in holding a morality that consists only of true moral principles.⁵ Nevertheless, as we have already noted, we regard our argument for universal benevolence and against egoism as persuasive but not decisive. We must therefore ask whether the conclusion that reasons for action are impartial can stand up against the apparent self-evidence of some partial reasons for action.

Sidgwick concluded that because egoism and universal benevolence are both rational, but conflict, it must be rational to follow either of them. This led him to despair of providing a rational basis for deciding what we ought to do. Crisp and Parfit point out that Sidgwick failed to consider the possibility that egoism and universal benevolence both provide reasons for action that carry weight, but are not necessarily decisive. If that is the case, then Sidgwick's despair over the dualism would have been unnecessary, because in some situations we have more reason to follow self-interest and in other situations we have more reason to do what is impartially best.

In *The Point of View of the Universe*, we argue that although this view – which Parfit defends in *On What Matters*, where he calls it the “wide value-based objective view”⁶ – is less damaging to the importance of morality than Sidgwick's dualism of practical reason, it still undermines morality because it must allow that there are some cases in which I choose to act wrongly but am not acting contrary to any decisive reason [p.163].⁷ Crisp objects:

Consider a version of the dualism in which the principle of universal benevolence outweighs the principle of egoism in every case except those in which, if the agent is to produce the greatest good, she must produce a trivial amount of good for very many people who are already much better off than she is at huge cost to herself. Since such a version would require the same sacrifices of most of us most of the time

⁵ Sharon Street, “A Darwinian Dilemma for Realist Theories of Value,” *Philosophical Studies* 127 (2006) pp.109-66.

⁶ Derek Parfit, *On What Matters*, Volume One, p. 147.

⁷ We are here referring to morality in a narrower and more conventional sense than that which we use elsewhere; see the discussion of wide and narrow understandings of terms like “ethics” and “morality” in section iv of this response, p. 194-196 below.

as the principle of universal benevolence, it is hard to see why it can be said to ‘undermine’ morality.⁸

We accept that our statement that “any form of the dualism of practical reason undermines morality” was too sweeping. The version of the dualism Crisp describes would not undermine morality to any significant extent. To make the example something more than a merely possible position, however, Crisp would need to indicate why, once it is granted that both self-interest and universal benevolence offer reasons for action that carry weight, the scales are tilted so strongly towards universal benevolence.

The position we defend in *The Point of View of the Universe* is, in Parfit’s terminology, Impartial Rationalism: we have most reason to do what is impartially best. Parfit finds this view attractive in some respects – he has himself argued that if the egoist asserts, against the person who steeply discounts future pains, that *when a pain occurs* is not relevant to how bad the pain is, a similar challenge can be mounted against the egoist for holding that *who is in pain* is relevant to the badness of the pain. Parfit now thinks that this argument has some force but is not decisive. He then refers to the separateness of persons and observes: “Most of us would find it hard to believe that we have no reasons to be more concerned about our own well-being.”⁹ That remark is followed by the hypothetical “If we admit that we have these reasons...” but when Parfit introduces his *Case One*, in which after a shipwreck you can save your own life or that of a stranger with a longer life expectancy than you, the hypothetical nature of this claim is dropped, and Parfit writes:

Act Consequentialists would believe that you would be acting wrongly if you saved your own life. But since this act would be likely to make the outcome only slightly worse, you would have only a weak moral reason not to act in this way. As Act Consequentialists should admit, this weak moral reason would not be stronger than your strong self-interested reason to give yourself many more years of happy life. You would have a sufficient reason to act wrongly by saving your own life.

Should act consequentialists admit this? If they are impartial rationalists, they will object to Parfit’s framing of the choice. From an impartial standpoint, the reason for saving the stranger’s life is *not* a weak moral reason pitted against a strong self-interested reason. You can expect to live X more years of happy life, and this is reflected in your self-interested reason to save your own life, but the stranger

⁸ Crisp, p. 129.

⁹ Parfit, p. 174.

can expect to live Y more years of happy life, and Y is greater than X , so her self-interested reason to hope that you will save her is, impartially considered, stronger than your reason to save yourself. In other words, the talk of weak and strong reasons holds only from a partial standpoint. Admittedly, as Parfit has pointed out, this partial standpoint is also your actual standpoint, and it is more plausible to claim that you should ignore a partial standpoint than to claim that you should ignore your actual standpoint.¹⁰ Nevertheless, we think the crucial question is whether it is likely that adopting either standpoint is equally likely to be rational. Our evolutionary debunking argument suggests that the impartial standpoint is more likely to be rational. Parfit acknowledges that this argument has some, though not decisive, force. He says the same of his own argument against egoism that we described above. We therefore have two arguments pointing to impartialism. What do we have on the other side? Only an appeal to intuition, but this is the same set of intuitions that we have sought to debunk.

There is another reason why the intuitions on which Parfit appears to rely may be suspect. Most of us are influenced, to some degree, by an instrumentalist view of reasons for action – the view that all reasons for action start from something wanted or desired. Economic theory takes this for granted. Parfit, of course, rejects it and, along with Nagel, Scanlon and Dancy, has stressed the importance of distinguishing motivating reasons, which require a desire, from normative reasons, which do not. When Parfit describes his shipwreck case and says that it is not plausible to hold that a weak moral reason outweighs a strong self-interested reason, we may nod in agreement because we are thinking of motivating reasons rather than normative reasons – or are simply oblivious to the distinction between the two. If the question is explicitly raised about normative reasons, we may have no clear intuition about it at all.

Parfit buttresses his position with an analogy:

We cannot defensibly believe that these weak moral reasons would always outweigh even the strongest self-interested reasons. That would be like Newman's horrific view that sin is infinitely worse than pain, so that, if all mankind suffered extremest agony, that would be less bad than if one venial sin were committed.

Newman's view is horrific because of its content – its indifference to agony, and the importance it gives to avoiding sin at all costs – and not because it rejects the possibility of a trade-off between different standpoints. Later in his paper Parfit

¹⁰ Derek Parfit, personal communication to the authors, 29.2.2016.

refers to etiquette in a way that suggests he believes that we have no reason to do what etiquette requires. We can imagine Emily Post responding to this by saying: “We cannot defensibly believe that weak moral reasons would always outweigh even the strongest reasons of etiquette. That would be like Newman’s horrific view...” We would, of course, reject that analogy, but if it doesn’t work as an argument for allowing strong reasons of etiquette to prevail over weak moral reasons, it should also not persuade us to allow strong self-interested reasons to prevail over weak moral reasons.

4. HOOKER ON WIDE VERSUS NARROW SENSES OF “ETHICS”

Hooker notes that Sidgwick took the fundamental question of ethics to be “What ought I to do?” and adds that he “sometimes” takes that question to be equivalent to “What is it rational to do?” It is true that Sidgwick is not entirely consistent in his usage, and so “sometimes” is, strictly speaking, accurate, but it is misleading. In the very first sentence of *The Methods of Ethics*, Sidgwick explains that by a “Method of Ethics” he means “any rational procedure by which we determine what individual human beings 'ought'—or what it is 'right' for them—to do, or to seek to realise by voluntary action.”¹¹ So the inclusion within the sphere of ethics of “any rational procedure” for deciding what we ought to do is not just Sidgwick’s casual and perhaps unthinking usage: it is his official position, and it is the departures from it that are likely to be unthinking lapses, or more probably, references to popular conceptions of morality that were not Sidgwick’s own.

Hooker’s first objection to taking ethics to be about what we have decisive reasons to do is that it “accords egoism more respect than it deserves.” The degree of respect that egoism deserves, however, depends on the truth of egoism’s claim that one always has decisive reason to do what is best for oneself: if this is true, egoism deserves the respect that we would give to any true answer to a fundamental practical question.

If we seek an answer to this question, there is nothing to be gained by defining ethics in such a way as to exclude egoism. The egoist will respond: “You may use the term ‘ethics’ as you please, but if you make it true by definition that egoism is not an ethical theory, so much the worse for ethics. You have not refuted my claim

¹¹ Henry Sidgwick, *The Methods of Ethics*, 7th edition, Macmillan, London, 1907, p. 1.

that I have decisive reasons to do what is in my interests, so that is what I will continue to do.”

If, on the other hand, what Hooker refers to as to “the commonplace that ethical considerations often conflict with egoistic ones.” is not merely a matter of terminology, but the expression of a widely held normative view, egoists are likely to respond that the masses accept this commonplace because they have been duped by the ruling class into accepting moral values that, beneath their altruistic veneer, serve the interests of the ruling class. That response can only be met by a substantive moral argument defending the commonplace as a justified normative view.

As the previous sections of this response indicate, we don’t think egoism provides the right answer to the question of what we have most reason to do. Whether or not you agree with us about this, however, is not the issue raised by Hooker’s criticisms of our wide understanding of ethics. We cannot escape the necessity of meeting the challenge posed by egoism either by stipulating as a matter of terminological fiat, that egoism is not an ethical theory, or by appealing to commonsense morality as if it were the standard against which egoism’s claims are to be judged.

In saying this we are turning against Hooker an argument he seeks to use against us. He writes:

... there is a real question whether, when moral reasons conflict with reasons of other kinds, the moral reasons always outweigh the other kinds of reasons. That question should not be eliminated by conceptual fiat. (Taking “morally required” to mean “what there is decisive reason to do” turns “there is decisive reason to do what morality requires” into a tautology.)

Hooker is right that the way we are defining “ethics” makes it a tautology that there is decisive reason to do what ethics requires, or what is right. (We don’t really object to Hooker substituting “morality” for “ethics” although we prefer Sidgwick’s term, which has less of the connotation of morality as a social institution.) No matter how we define our terms, some claims that would be substantive if we defined the terms differently will become tautologous. We have just seen that if we use the terms in the narrow sense Hooker prefers, it is a tautology that egoists are unethical. It doesn’t really matter whether one uses terms so as to make it a tautology that we have decisive reason to do what ethics requires, or use them to make it a tautology that egoists are unethical. What does matter is that we are clear about our terminology, and do not seek to achieve, by terminological fiat, a

conclusion that requires substantive argument. The substantive dispute is not about whether egoists are unethical, but about what we have most reason to do.

The lesson to be drawn from all this is that arguing about how we define terms like “moral,” “ethical,” “ought” and “wrong” doesn’t get us very far.¹² We do, however, acknowledge that the popular meaning of these terms often takes a narrower sense than that in which we and Sidgwick use it. That may be why Sidgwick sometimes uses these terms in a manner that is inconsistent with the position he states in the first sentence of *The Methods of Ethics*, and why we too have not been entirely consistent in our usage.¹³

We should not leave this topic without commenting on what Hooker seems to regard as a clinching argument against equating moral wrongness with what there is decisive reason to avoid: the case of the less preferred sorbet. It would not be morally wrong, Hooker says, for someone to choose the flavor she likes least. Hooker seems to have forgotten, however, that we are hedonistic utilitarians. Presumably if you choose the flavor you like least, you will get less pleasure from your sorbet than if you choose the flavor you like best. Pleasure is, for us, intrinsically good, so you are choosing to bring less intrinsic goodness into the universe. Nor is any other sentient being getting more pleasure as a result. Therefore what you are doing is morally wrong.

We admit that this remark will strike many people as absurd. Suppose you prefer strawberry to toffee. Are we going to tell you that you have a moral obligation to choose the strawberry sorbet? We don’t normally use moral language in such trivial matters, and we don’t normally tell people that they ought to do what they most want to do, because usually are all too ready to do that. We especially avoid using moral language to criticize people’s choices in trivial matters that only affect themselves. Perhaps we think people will, on the whole, be happier if there are areas of their life where they can feel free to choose whatever they want, without criticism. Another reason may be because we want to reserve the heavy artillery – moral criticism – for battles where more is at stake. These reasons could explain

¹² As one of us argued long ago: see Peter Singer, “The Triviality of the Debate over “Is-Ought” and the Definition of “Moral”, *American Philosophical Quarterly*, vol. 10 (1973), pp. 51-56.

¹³ One such example occurs in our discussion of whether morality is undermined if we do not have decisive moral reasons to avoid doing what is wrong. We had in mind “morality” and “wrong” in the positive sense, referring to a set of principles or rules accepted in society. We did not, of course, mean that for morality to have its full importance, we must have decisive reasons to avoid doing what we have decisive reasons not to do. We thank Derek Parfit for pointing this out, in an unpublished earlier version of his contribution to this symposium.

why we find it strange to say that the person who knowingly chooses to give herself less pleasure from her sorbet has done something morally wrong. But she has done something that is, other things being equal, wrong – not very wrong, of course, but still wrong.

5. HOOKER ON RULES AND MORAL WRONGNESS

The case of the less preferred sorbet figures again in Hooker's response to our (and Sidgwick's) defense of esoteric morality. Hooker holds that moral wrongness should be determined not by its consequences, but by optimal public moral rules and optimal procedures for making everyday decisions. As the rules need to be public, there is no scope for doing what is right in rare circumstances by breaching the optimal public rules, but keeping the breach secret. Hooker needs the sorbet example here because he is willing to concede that "if moral wrongness is equated with what there is decisive reason to avoid" then we should *not* assume that moral wrongness should be determined by optimal public rules and procedures. If it is not wrong to choose the less preferred sorbet, however, then we should not equate moral wrongness with what there is decisive reason to avoid. Obviously, this answer will not persuade utilitarians like us who think that it is wrong not to choose the sorbet that will give you the most pleasure. We are therefore able to resist Hooker's claim that moral wrongness should be determined by optimal public rules.

Hooker has succeeded in persuading us, however, that we were in error when we said that although we should blame others only when doing so maximizes utility, we should feel the reproaches of our conscience only when we have done something wrong. He imagines himself on his deathbed, paralyzed and incapable of acting or of communicating with others, but finally realizing that he has failed to maximize utility. We fervently hope that Hooker will never find himself in such dire circumstances (and will realize much sooner that he should maximize utility), but we agree that anyone in the condition he describes would be justified in making use of whatever pleasant distractions will turn his thoughts away from his moral failings.

Hooker next questions Sidgwick's view, and ours, that we should not blame those who do more than others, even if what they do still falls short of what will maximize utility. We should be clear that we are not putting this forward as a rule without exceptions, but rather as a reminder that, for utilitarians, praising and

blaming are themselves actions, and as with all actions, we ought not to do what will have worse consequences than an alternative action, or inaction. (This doesn't apply to *believing* someone to be blameworthy, without acting on that belief. That raises different issues.) In a world in which many have much more than they need while others are living in extreme poverty, we hold that people who have more than they need ought to be giving substantial sums to the most effective charities they can find.¹⁴ Yet we live in a society in which very few give even one percent of their income. If we learn that a wealthy person has given 10 percent of her income to effective charities, should we blame her for not giving 20, 50 or 80 percent of her income? Should we do that even if there is good social science research showing that praising people who exceed the norm for giving will encourage others to give more, whereas blaming such people has the opposite effect?¹⁵

We think it is clear that in these circumstances, we should not always blame people for doing what is, objectively speaking, wrong. Instead we should, as Sidgwick suggests, use praise and blame to raise the general ethical standard. This may, as Hooker says, give rise to some counter-intuitive cases, but this is another area in which we should not trust our intuitions. Our readiness to praise and blame no doubt developed during our long evolution as social mammals, and for most of that time we would not even have been capable of asking ourselves whether to blame someone for doing something we regard as wrong would, in the long run, have better consequences than not blaming her. Even today when, as children, we learn to praise and blame, we are not likely to think about the long-term consequences. So it is not at all surprising that these less calculating reactive attitudes continue to generate the intuitions to which Hooker appeals.

Hooker gives a brief but interesting response to the “million imbeciles/one genius” objection to his view that the right thing to do, whether you are an imbecile or a genius, is always what is in accordance with the optimal public rules. He concedes that what is right for someone who has far more intelligence than everyone else may not be right for someone whose intelligence is roughly similar to that

¹⁴ See Peter Singer, *The Life You Can Save*, Random House, New York, 2009 (Italian translation, *Salvare una vita si può: agire ora per cancellare la povertà*, Il Saggiatore, Milan, 2009); Peter Singer, *Famine, Affluence and Morality*, Oxford University Press, New York, 2015.

¹⁵ We know of no research establishing this specific conclusion, but there is research indicating that to mention a demanding level of giving as the norm can deter potential donors and lead to a lower total sum being raised. See Jen Shang and Rachel Croson, “Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods,” *The Economic Journal*, 119 (2009) 1422-1439.

of most others and adds that rule-utilitarians have thought of themselves as asking what is right for people of roughly similar intelligence.” That appears to stake the soundness of taking optimal public rules as the criterion of right and wrong (and therefore as the basis for rejecting esoteric morality) on whether we live in a world in which people are of roughly similar intelligence. Granted, we don’t live in a world bifurcated between into imbeciles and geniuses, but nor do we live in a world in which everyone is of “roughly similar” intelligence. Sidgwick thought that the differences are sufficient to justify some people, some of the time, practicing esoteric morality. It isn’t obvious that he was mistaken.

6. NAKANO-OKUNO ON EXTENDING ETHICS TO ALL SENTIENT BEINGS

Mariko Nakano-Okuno suggests that we have reconstructed Sidgwick’s argument for utilitarianism in a manner that leads to the version of utilitarianism we favor, that is, a version that requires us to have equal concern for the interests of people who near to us or far away, and for nonhuman animals too.¹⁶ This implies that one could equally well reconstruct Sidgwick’s argument for utilitarianism in a manner that did not require us to consider people in remote places, or to treat animals equally. Yet Nakano-Okuno is a careful Sidgwick scholar, and so recognizes that Sidgwick himself thought we should include all sentient beings in our consideration of what we ought to do. She quotes him as saying that it would be “arbitrary and unreasonable” to fail to take into account “any pleasure of any sentient being.”¹⁷ We are therefore puzzled as to why she should think that we have somehow tailored our interpretation of Sidgwick to fit our particular version of utilitarianism. It is, of course, a separate question whether we and Sidgwick are justified in holding that we ought to take into account, in deciding what to do, the interests of every sentient being. Nakano-Okuno argues that this conclusion does not follow from Sidgwick’s Axiom of Justice, his Axiom of Benevolence, or his acceptance of hedonism.

We accept what Nakano-Okuno says about Sidgwick’s Axiom of Justice. It leaves open the question of what differences between two individuals are grounds for difference of treatment. Although we doubt that a mere difference of species could be stated as a reasonable ground for difference of treatment, differences

¹⁶ Nakano-Okuno, p. 153.

¹⁷ *ME*, 414, as quoted by Nakano-Okuno, p. 155.

such as the possession of the capacity to reason, or self-awareness, or autonomy, are more plausible candidates. In any case, the Axiom of Justice says nothing about which differences provide reasonable grounds for difference of treatment, and hence does not, on its own, require the extension of our concern to all sentient beings.

Does the Axiom of Benevolence requires us to be concerned about the wellbeing of all sentient beings? Nakano-Okuno quotes one sentence in which we say that it does. It comes, as she points out, from our argument against an evolutionary explanation of that axiom. To reinforce our argument, we say that even if there could be an evolutionary explanation of altruism in terms of its benefit for our own species, this would not suffice to explain the axiom of benevolence because it “bids us to have concern not only for the good of our own species, but for all sentient beings.”¹⁸ This sentence was not as precisely phrased as it should have been. We should have written that the axiom of benevolence “bids us to have concern not only for the good of members of our own species, but for whatever members of other species are capable of having a good at all.” This phrasing leaves open the question of what is of intrinsic value, whereas the sentence as published implies that all sentient beings are capable of realizing intrinsic value in some way. That should not be assumed because the Axiom of Benevolence does not specify what is of intrinsic value. If intrinsic value were to consist only in, say, autonomous action by moral agents, then the axiom would not require us to have concern for all sentient beings, but only for those who are moral agents.

Sidgwick’s claim that in deciding what we ought to do, we should take into account, “any pleasure of any sentient being” is based not on the Axiom of Benevolence alone, but on that axiom combined with his theory of value, namely hedonism. Towards the end of her essay, Nakano-Okuno recognizes that this may be our position. To this she replies that Sidgwick’s proof of hedonism fails to convince us that “all kinds of positive internal feelings of all sentient beings” are to be included in the ultimate good. She elaborates:

In his proof of hedonism, Sidgwick proceeds by asking himself and us what we, upon reflection, would regard as ultimately desirable. It seems to me that, during this process, I could claim that I cannot regard the presumably positive conscious state of insects, amphibians, birds and probably swine as ultimate good, simply because I cannot really imagine what it is like to feel like them. I might be able to better imagine and represent in my mind the internal feelings of such communicative

¹⁸ *The Point of View of the Universe (PUV)*, 187.

animals as primates and dolphins, as well as pets such as dogs and cats, but I can hardly imagine what it is like to feel like a fly or a cockroach, so I would possibly exclude the internal feelings of arthropods from the category of desirable feelings; then following the same logic, I may possibly exclude the internal feelings of fish, whose brains seem also simple enough, by refusing to regard them as desirable, and then those of birds, reptiles, and so on.

There are three serious flaws in this passage. First, Sidgwick's argument for hedonism does not depend on our ability to imagine what the conscious states of various beings are like. Rather, it asks us to think about what is ultimately desirable. Sidgwick believes that if we do this, we will agree that only conscious experiences have intrinsic value, with pleasures having positive intrinsic value and pains having negative intrinsic value. This leads us to a hypothetical judgment about insects, amphibians, birds, pigs, and, for that matter, human beings with various kinds of brain damage or profound intellectual disability: *if* they experience pleasures or pains, *then* their experiences have intrinsic value; and if they do not, their experiences, and indeed their very existences, have no intrinsic value. (They might, of course, have great instrumental value for other sentient beings.)

Second, Nakano-Okuno may well be justified in taking the view that arthropods have no desirable conscious experiences, but she then moves very swiftly through completely different kinds of beings. She cannot do this by, as she puts it, "following the same logic" because this is not a question of logic, but of empirical investigation into the anatomy, physiology and behavior of each species, genus, family, phylum, or whatever else the category might be. There is no reason to think that what holds for insects also holds for fish or birds. The evidence is different in each case, and it must be examined.

In 2012, a prominent international group of cognitive neuroscientists, neuropharmacologists, neurophysiologists, neuroanatomists and computational neuroscientists gathered at The University of Cambridge agreed to *The Cambridge Declaration on Consciousness*, which includes the following statement:

The absence of a neocortex does not appear to preclude an organism from experiencing affective states. Convergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors. Consequently, the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Nonhuman animals, including all

mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates.¹⁹

Other parts of the statement reinforce the impression given by this passage, that the authors of this statement regarded the evidence for consciousness in mammals and birds, and octopuses, as stronger than in, say, insects.

Third, Nakano-Okuno's criterion of what she can imagine allows her to include "pets such as dogs and cats" but not "swine." We know of no scientific grounds for believing that pigs are less capable of experiencing pleasure and pain than dogs and cats. Nakano-Okuno's position here may be based on the fact that many humans live with dogs and cats and are often able to know what their companion animal is experiencing. That is not, however, a reason for thinking that dogs and cats have mental lives that are richer or more intense than the mental lives of pigs. We need a more objective basis for deciding which beings are capable of conscious experiences and therefore, for a hedonist, of experiences that have intrinsic value.

Nakano-Okuno thinks she has shown that Sidgwick's proof of hedonism does not lead to the requirement for expanding the circle of ethics that we advocate; but this must be based on a misunderstanding of what it is that we advocate. We are happy for people to raise doubts about *whether* certain beings can feel pleasure or pain. That does not challenge our goal of expanding ethics to all beings who *do* feel pleasure or pain. If Nakano-Okuno has misunderstood this, the fault may lie in our use of the term "sentient being." We use this term as equivalent to "a being capable of conscious experiences, including the capacity to experience pain and/or pleasure." One might think that because mosquitos are obviously capable of sensing the presence of blood, they must be sentient beings. For those who understand the term "sentient being" in this way, whether mosquitos experience pleasure when they bite someone and suck in blood, is a separate question, not answered by acknowledging that they are sentient beings. When we call for the expansion of ethics to all sentient beings, however, we do *not* mean all beings capable of sensing something, as the mosquito is capable of sensing the presence of blood and as, for that matter, a plant is capable of sensing light; we mean beings capable of experiencing pleasure or pain. Notwithstanding Nakano-Okuno's

¹⁹ *The Cambridge Statement on Consciousness*, July 7, 2012, <http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf>. For a useful discussion of the evidence regarding fish, see Victoria Braithwaite, *Do Fish Feel Pain*, Oxford University Press, New York, 2010.

arguments, we continue to regard this expansion of ethics as a requirement of hedonistic utilitarianism.

7. CRISP ON PLEASURE

Crisp is right that it is hard to figure out which of Sidgwick's definitions of pleasure we should treat as canonical. Not everything he says on this topic is consistent. In our book we regard his final view as what Crisp calls "the apprehension account," but this label could be misleading. The key sentence is:

I propose therefore to define Pleasure – when we are considering its 'strict value' for purposes of quantitative comparison – as a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable or – in cases of comparison – preferable. (*ME* 127)

It is important to understand that this definition does not restrict pleasure to intelligent beings capable of apprehending it as desirable. Rather, it defines pleasure as that feeling which, *when* apprehended by intelligent beings is at least implicitly apprehended as desirable. (Compare a definition of "poisonous" as the property of things that causes illness in beings who ingest them. That would not mean that a mushroom that is never ingested cannot be poisonous.) Crisp is right to say that Sidgwick also (just two pages later) refers to pleasure as "desirable feeling, apprehended as desirable by the sentient individual at the time of feeling it," without any reference to intelligent beings. Perhaps he himself was unclear whether "apprehension as desirable" requires intelligence, or merely sentience. Would he have said that a lizard who moves to a sunny spot on a rock implicitly apprehends the warmth of the sun as desirable? It's hard to say, but it seems possible, given that Sidgwick is here talking about pleasure qua feeling, and specifically contrasting that apprehension with the judgment of the "moralist of stoical turn" who he thinks ought to admit that the feeling of pleasure is *per se* desirable, while at the same time "holding that sound philosophy shows the illusoriness of such judgments." (*ME* 129)

Crisp asks whether by "desirable" Sidgwick here has in mind the understanding of that term reached at the end of his earlier discussion of a desire-based view of what is good for someone. That understanding was that something is desirable if it is "what I should practically desire if my desires were in harmony with reason, assuming my own existence alone to be considered." (*ME* 112) If he does mean

that, his position would be difficult to reconcile with the distinction that, as we have just seen, Sidgwick draws when he contrasts the feeling of pleasure as desirable with the stoic's judgment that sound philosophy shows the illusoriness of such judgments. If that is what sound philosophy shows, then surely if my desires were in harmony with reason, I would not desire to have such feelings. In other words, to include in the definition of pleasure a reference to what is desirable means, on this view of what is desirable, that we cannot know what pleasure is until we have an answer to the question of what is ultimately good. But as pleasure is itself under consideration for being ultimately good, that introduces circularity into the account of pleasure. We find it difficult to imagine Sidgwick failing to realize this. We find it more plausible to believe that when he uses the term "desirable" in his definition of pleasure, he has in mind the more ordinary sense of the term, that is, that pleasure is a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as something they desire, qua feeling. On this view, the reference to "intelligent beings" makes very little difference, since we could quite naturally say that the lizard desires the warmth of the sun, and so "implicitly apprehends" that warmth as desirable.

Another possible interpretation – lying midway between the two we have so far considered – is that by "desirable" Sidgwick means "is implicitly apprehended as something worthy of desire, qua feeling" where "worthy of desire" refers merely to the fact that the intelligent beings do so apprehend it, not that their apprehension is necessarily in accordance with the desires they would have if their desires were in harmony with reason. This interpretation has the advantage of explaining why Sidgwick might have thought it necessary to refer to intelligent beings, since such apprehension is presumably beyond lizards.

Crisp thinks that the "moralist of stoical turn" – who he refers to as the ascetic – could deny that pleasure is desirable, even as a feeling, and could instead give a descriptive, non-evaluative account of it. Crisp revisits Sidgwick's objection to this view, that there is no common quality in the many and very varied feelings we describe as pleasure, other than that we find them desirable, qua feeling. As Crisp notes, in our book we approve of this "heterogeneity objection," and cite the work of the neuroscientists Kent Berridge and Morten Kringelbach in support of the view that pleasure is not a sensation, but rather is a kind of "hedonic gloss" that our hedonic brain systems paint on certain sensations. Crisp thinks that it is better to interpret Berridge and Kringelbach as saying that pleasure is not merely a sensation, but allowing that the hedonic gloss may itself be a sensation, or a sin-

gle type of feeling. We are now willing to concede that Crisp may be right on this point.²⁰ If so, the heterogeneity objection can be met, to the extent that we can say that the hedonic systems of the brain put the same kind of hedonic gloss on various different sensations. The question is, however, whether this does not simply transfer the problem of saying what pleasure is to the problem of saying what the hedonic gloss is. How is it to be distinguished from other feelings? If we answer this question by saying that it is just a particular kind of feeling, in the way that sweetness is a particular kind of taste, then anyone who thinks that pleasure is intrinsically good needs to explain why just this kind of feeling has intrinsic value. If some people do not desire this kind of feeling, even qua feeling, then hedonistic theories of the good will fail to satisfy what, following Peter Railton, we have called “the resonance requirement” – we will be telling such people that pleasure is good for them, but this will not resonate with anything they desire or value.²¹ Obviously, the same problem recurs if we describe the hedonic gloss in terms of certain brain states. If, on the other hand, we say that the hedonic gloss is to be distinguished from other feelings by the fact that we do desire it, qua feeling, we are back to where we started, with pleasure being whatever mental state we desire, qua mental state. These problems remain in need of more work.²²

8. CRISP ON SIDGWICK VERSUS ROSS

We argue, in our book, that hedonistic utilitarianism has, when compared with Ross’s pluralistic intuitionism, the advantage that it is, at least in principle, able to give a definite answer to the question of what I ought to do, whereas on Ross’s view, when prima facie duties conflict, one can only intuit, in each specific set of circumstances, which duty or duties carry more weight. Crisp counters this argument by saying that even the hedonist must balance pleasure and pain against each other, so the same problem arises for hedonistic utilitarianism too.

Sidgwick was aware of this objection, and responded to it by saying pain is “the negative quantity of pleasure” and so they are not distinct qualities, but different

²⁰ We thank Adam Shriver for discussions after our book was published which had already inclined us to think that our reading of Berridge and Kringelbach was mistaken.

²¹ On the possibility of not desiring pleasure, understood as a particular feeling, see Adam Shriver, “The Asymmetrical Contributions of Pleasure and Pain to Subjective Well-Being,” *The Review of Philosophy and Psychology*, 5 (2014) 135-153, especially pp. 140-146. On the resonance requirement see *The Point of View of the Universe*, pp.214-5, 252-3.

²² One of us (KLR) is currently carrying out research in this area.

points on the same scale. (ME 125) At the time of writing our book, we would have endorsed this view. Adam Shriver has recently argued that current research in neuroscience does not support it. There are, he says, huge differences between pleasure and pain as to where they take place in the brain and this in turn influences the ways in which pleasure and pain are related to motivation and hedonic evaluation.²³ This is not the place to decide whether these differences really mean that hedonistic utilitarianism has two potentially conflicting intrinsic goods, but Shriver's arguments leave us unsure whether we have an adequate response to the objection.

Crisp does not discuss another reason why we think utilitarianism has an advantage over Rossian intuitionism. Ross starts by assuming that, as he puts it, "the main moral convictions of the plain man" are "not opinions which it is for philosophy to prove or disprove, but knowledge from the start."²⁴ As we have already indicated both in our book and in this response, we are skeptical about the validity of most of our moral intuitions. Debunking arguments, in other words, play a role here again in persuading us that utilitarianism is more defensible than Ross's intuitionism, whether the explanations of "the main moral convictions of the plain man" are evolutionary or cultural.

²³Adam Shriver, "The Asymmetrical Contributions of Pleasure and Pain to Subjective Well-Being," *The Review of Philosophy and Psychology*, 5 (2014) 135-153. See also Adam Shriver, "The Asymmetrical Contributions of Pleasure and Pain to Animal Welfare," *Cambridge Quarterly of Healthcare Ethics*, 23 (2014) 152-162.

²⁴W.D. Ross, *The Right and the Good*, ed. P. Stratton-Lake, Clarendon Press, Oxford, 2002, 20-1, n.1