INTRODUCTION
Symposium on David Phillips's *Sidgwickian Ethics*

Henry Sidgwick's The Methods of Ethics has many admirers. In the foreword to the Hackett edition of the book, John Rawls claims that of the works devoted to defending classical utilitarianism The Methods of Ethics is "the most philosophically profound." More recently, in On What Matters, Derek Parfit states that The Methods of Ethics is the best book on ethics ever written, since it contains "the largest number of true and important claims."

Despite these accolades, there exist only a small number of attempts to systematically engage with the ethical doctrines in The Methods of Ethics, including F. H. Hayward's The Ethical Philosophy of Sidgwick: Nine Essays, Critical and Expository, chapter six of C. D. Broad's Five Types of Ethical Theory, and J. B. Schneewind's Sidgwick's Ethics and Victorian Moral Philosophy. Each of these rewards repeated study, but they are no longer current. In the nearly four decades since Schneewind's book was written, there have been many innovations in philosophical ethics. In addition, a good number of excellent article- and chapter-length discussions of specific doctrines in Sidgwick have appeared, as has Bart Schultz's monumental achievement, Henry Sidgwick: Eye of the Universe. An Intellectual Biography, that treats of Sidgwick's life and entire corpus.

This has for some time represented an opportunity for a new systematic book-length treatment of Sidgwick's ethics and its relation to contemporary work in meta-ethics and normative ethics. David Phillips's newly published Sidgwickian Ethics (2011) has exploited this opportunity. Phillips provides an interpretation and evaluation of the key meta-ethical and normative moral views found in The Methods of Ethics. Accordingly, he deals with Sidgwick's commitment to non-naturalism, to epistemic intuitionism, to the falsity of deontology and to the truth of both utilitarianism and rational egoism, and therefore to the dualism of practical reason.

Phillips does a number of things of note. He brings contemporary work in meta-ethics and moral epistemology to bear on Sidgwick's (now popular) views in these areas, in an effort to both understand and defend them. He provides a full account of Sidgwick's argument for utilitarianism, emphasizing the role played therein by the self-evident intuitions that Sidgwick accepts. He raises a number of astute criticisms of it, including that its argument against the deontological morality of common sense is unfair. He spends considerable effort making sense of and in part defending rational egoism, concluding that the argument that Sidgwick provides for egoism is more successful than his argument for utilitarianism. Using these insights, Phillips sheds important light on the nature of Sidgwick's dualism of practical reason.

In every case the arguments and analyses found in Sidgwickian Ethics are worthy of sustained study. This is, of course, not to say that what Phillips argues is beyond cavil. Hence this symposium, which provides an opportunity to determine to what extent Phillips's own contributions to Sidgwick scholarship survive rational scrutiny. It includes, in addition to a précis of Sidgwickian Ethics, critical reactions to its main interpretive and philosophical claims from Roger Crisp, Robert Shaver and Anthony Skelton, and a reply to these reactions by Phillips. Beyond providing an opportunity to extend the conversation on Phillips's understanding of Sidgwick's doctrines and their philosophical import, the symposium is meant to contribute more broadly to the scholarly and philosophical appreciation of Sidgwick's distinct ethical outlook.[1]

Anthony Skelton

University of Western Ontario

[1]The papers in this symposium were first read at the 12th Annual Conference of the International Society for Utilitarian Studies, Stern School of Business, New York University, August 8-11, 2012.

Précis: Sidgwickian Ethics

My aim in *Sidgwickian Ethics*[2] is to interpret and evaluate the central argument of *The Methods of Ethics*,[3] in a way that brings out the important conceptual and historical connections between Sidgwick's views and contemporary moral philosophy.

Sidgwick defines a "method of ethics" as "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" (ME 1). He finds just three such methods "implicit in our common moral reasoning": egoism, utilitarianism, and intuitionism (ME 14). According to egoism, as Sidgwick defines it:

> The rational agent regards quantity of consequent pleasure for himself as alone important in choosing between alternatives of action. (ME 95)

Sidgwick is well aware of the important structural parallels between egoism and utilitarianism. Though he does not himself give it, we can generate a definition of utilitarianism modeled on his definition of egoism. According to utilitarianism, thus defined:

> The rational agent regards the quantity of consequent pleasure for all sentient beings as alone important in choosing between alternatives of action.

The definition of intuitionism requires more care, for it is really a hybrid concept, with both an epistemic and a moral-theoretic component. And that hybrid character is crucial to Sidgwick's distinctive philosophical view. He characterizes intuitionism as the view that

> We have the power of seeing clearly that certain kinds of actions are right and reasonable in themselves, apart from [at least some of] their consequences. (ME 200)

---

[2] David Phillips, *Sidgwickian Ethics* (New York: Oxford University Press, 2011).

[3] Henry Sidgwick, *The Methods of Ethics* 7th edition (Indianapolis: Hackett, 1981; original publication: 1st edition 1874, 7th edition 1907). I will refer to it as "ME". Further page references will be placed in the text. Chapters will be referred to by (Roman) Book then Chapter Number.

The epistemic component here is the idea that some moral claims are self-evident; the moral-theoretic component is the idea that what is self-evident is something non-consequentialist.

Though I claim to vindicate much of what I find in Sidgwick, my approach doesn't fit everything he says, and my conclusions differ in some important ways from his. As to approach, Sidgwick often presents himself as primarily concerned with methods, not principles, and with developing and considering three methods neutrally and in isolation, rather than with "discussing the considerations which should…be decisive in determining the adoptions of ethical first principles…[or establishing]…such principles" (ME 14). But I think the most interesting and interpretively troubling parts of the methods are exactly those parts where Sidgwick is focused on conflicts between ethical first principles and on which such principles we should adopt: most importantly III XI, III XIII, IV II, and the Concluding Chapter. And I think that in order to understand these Chapters, the most crucial background is the metaethics and moral epistemology found most centrally in I.I, I.III, and I.VIII.

Sidgwick's own conclusions about the conflicts between, respectively, utilitarianism and intuitionism, and utilitarianism and egoism, are well captured in a passage from the short intellectual autobiography added by E.E. Constance Jones to the Preface to the 6th edition of the *Methods*:

> I had found…that the opposition between Utilitarianism and Intuitionism was due to a misunderstanding. There was indeed a fundamental opposition between the individual's interest and either morality, which I could not solve by any method I had yet found trustworthy, without the assumption of the moral government of the world. (ME xxii)

To unpack: Sidgwick thinks that utilitarianism and intuitionism can be reconciled because, first, there is no real epistemic contrast between utilitarianism and intuitionism: all sensible moralists will be epistemic intuitionists; and, second, the moral-theoretic conflict between utilitarianism

and intuitionism is won by the utilitarian. By contrast, utilitarianism and egoism cannot be reconciled (without theological help); they remain in fundamental opposition, an opposition which Sidgwick often characterizes as involving a contradiction.

I disagree in important ways with these conclusions of Sidgwick's. First, I think he does not succeed in arguing for utilitarianism as against the moral-theoretic component of intuitionism, primarily because his attack on the moral-theoretic component of intuitionism is subject to a charge of unfairness: he requires the intuitionist's putative first principles to satisfy criteria which are so stringent that his own favored, proto-utilitarian first principles do not satisfy them either. Second, I think his view about the conflict between egoism and utilitarianism is too pessimistic: while he typically presents the conflict as involving a fundamental contradiction, his arguments in fact do not properly generate any such contradiction, but instead provide support for a range of plausible and historically distinctive views of practical reason, according to which there are both agent-relative and agent-neutral reasons.

After the introductory chapter my book has four further chapters on, respectively, Sidgwick's Metaethics, Sidgwick's Moral Epistemology, Utilitarianism Versus Dogmatic Intuitionism, and Utilitarianism Versus Egoism. In this brief *précis* I cannot hope to articulate even in summary form all that is in those chapters. But I do hope to be able to say enough to introduce the most important claims I make with which some other interpreters, including in some cases some of my commentators here, disagree; to indicate the key textual sources for central interpretive issues in and objections to Sidgwick; and to indicate when I take my discussion to move in directions relatively underexplored in the Sidgwick literature.

In the Chapter on Sidgwick's metaethics, I claim that Sidgwick commits himself to, and defends with appropriate modesty, ethical non-naturalism. I locate non-naturalism within a

metaethical conceptual scheme with three other main options: naturalism, noncognitivism, and error theory. While I claim that this conceptual scheme (versions of which are found, *inter alia*, in Mackie and Parfit)[4] fits Sidgwick well and can be articulated in Sidgwickian language, I do not claim that it is a scheme with which Sidgwick himself explicitly operated. I argue (contra some fascinating work of Rob Shaver's)[5] that non-naturalism is a fundamental and strategic, rather than a peripheral and tactical, commitment of Sidgwick's. And I argue that, in consequence, though he did not encounter sophisticated contemporary versions of naturalism or noncognitivism, the right way to extrapolate Sidgwick's views into contemporary metaethical debates would involve rejecting any such noncognitivist or naturalistic view and embracing instead the sort of non-naturalism recently articulated, *inter alia*, by Nagel, Parfit, and Scanlon.[6] I focus finally on the interesting and underexplored relationship between Sidgwick and error theory, drawing in particular on Mackie's presentation, in his paper "Sidgwick's Pessimism,"[7] of a distinctive Sidgwickian argument for error theory.

In the Chapter on Sidgwick's moral epistemology I pursue three main themes. First, I interpret and defend Sidgwick's epistemic intuitionism, claiming (i) that Sidgwick offers an argument for intuitionism that is both crucial to him and has non-naturalism as a key premise, and (ii) that, appropriately recast and generalized, Sidgwick's argument for intuitionism is a powerful one. Here again, Shaver's work is my main foil.

Second, I articulate what I claim is a hitherto unrecognized, or under-recognized, puzzle in Sidgwick's moral epistemology. And I claim that this puzzle helps explain the striking level of

---

[4] J.L. Mackie, *Ethics* (Harmondsworth: Penguin 1977), pp. 31-35; Derek Parfit, *On What Matters*, Vol. 2 (Oxford: Oxford University Press, 2011), p. 263.
[5] Robert Shaver, "Sidgwick's Minimal Metaethics," *Utilitas* Vol. 12, No. 3 (2000), pp. 261-277.
[6] Thomas Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1986), Chapter VIII; Derek Parfit, *On What Matters*, Vol. 2 (Oxford: Oxford University Press, 2011), Part 6; Thomas Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), Chapter 1.
[7] J.L. Mackie, "Sidgwick's Pessimism," *Philosophical Quarterly* Vol. 26, No. 105 (1976), pp. 317-327.

scholarly disagreement as to how to understand that moral epistemology. The categories required to frame the puzzle are most readily found in Sidgwick's 1879 paper, "The Establishment of Ethical First Principles".[8] There Sidgwick begins with the problem of how a proponent of one putative ethical first principle can argue against a proponent of some other, conflicting, putative first principle. Sidgwick articulates two "quite different" forms of argument. In the first, which I label "bipartite argument," we begin with the claim the interlocutor thinks self-evident, and shows that it is not itself really self-evident, but at best a true consequence of some more general and genuinely self-evident claim; in the second, which I label "criterial argument," we develop general, topic-neutral criteria for first principles, and then apply these to determine which putative ethical first principles are genuine. The best initial way to see the puzzle is to contrast "The Establishment of Ethical First Principles" with the passage from *Methods* IV II where Sidgwick presents the same problem, but offers instead (I claim) just one solution, bipartite argument. Why? Why does the conception of independent criterial argument not appear in Methods IV II? Thus far, this may seem a minor puzzle: a conflict between a brief early paper and the much revised and much more authoritative text of the *Methods*. But, I suggest, the puzzle cannot be thus dismissed, for two reasons. First, the conception of criterial argument is not a conception that appears only in "The Establishment of Ethical First Principles"; instead, it appears in most of Sidgwick's work in general epistemology, including some of his last published work.[9] Second, a criterial argument for utilitarianism is apparently offered in Book III of the *Methods*. Hence (in part) the scholarly disagreement about Sidgwick's moral

[8] Henry Sidgwick, "The Establishment of Ethical First Principles," *Mind* Vol. 4, No. 13 (1879), pp. 106-111.
[9] Five relevant papers are reprinted in Part III of the collection of Sidgwick's papers edited by Marcus G. Singer, *Essays on Ethics and Method* (Oxford: Clarendon Press, 2000). See in particular "Criteria of Truth and Error" and "Further on the Criteria of Truth and Error," originally published respectively in 1900 and posthumously in 1905.

epistemology, even among those who have not drawn inspiration in just the way I have from "The Establishment of Ethical First Principles."

The third and final theme in the moral epistemology Chapter is the overall interpretation of Sidgwick's moral epistemology. I argue for a moderate foundationalist reading, and have a number of tactics, whose effectiveness some of my commentators contest, for largely explaining away the appearance that Sidgwick takes common-sense morality to be epistemically authoritative.

In Chapter Four, I focus on the first of the two key conflicts where my verdict differs from Sidgwick's: the conflict between utilitarianism and (dogmatic) intuitionism. Sidgwick argues that the dogmatic intuitionist's putative first principles can be rejected because they fail to satisfy the four conditions or criteria articulated on pp. 338-342 of the *Methods*. By contrast, he (implicitly) claims in III XIII that his favored "philosophical intuitions" do satisfy the four conditions. Sidgwick's argument is, I claim, vulnerable to a charge of unfairness. The basic version of the charge can be developed by contrasting the passage where Sidgwick claims that the maxims of common-sense morality fail to meet the four conditions with a passage introducing his favored philosophical intuitions. In the former passage, Sidgwick observes that

> [Common-sense morality's] maxims do not fulfill the conditions…So long as they are left in the state of somewhat vague generalities…we are disposed to yield them unquestioning assent, and it may fairly be claimed that the assent is approximately universal…But as soon as we attempt to give them the definiteness which science requires, we find that we cannot do this without abandoning the universality of acceptance. (ME 342)

In the latter, introducing his favored philosophical intuitions, Sidgwick observes that they are

> of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case (ME 379)

So, the objection has it, Sidgwick insists that common-sense principles meet a standard of determinacy from which he exempts his own favored principles. And that is unfair. I try, on Sidgwick's behalf, various responses to this objection, but find none satisfactory. And I argue that the problem becomes still worse for Sidgwick if the intuitionist is (properly) allowed to frame her view in a Rossian rather than a Whewellian form, employing the concept of *prima facie* duty.

Finally, in Chapter Five, I turn to the conflict between Egoism and Utilitarianism and the Dualism of Practical Reason. There is lots of scholarly debate as to how to interpret the dualism. I aim in part to characterize this debate, and to contribute to it. But the primary resource to which I appeal is not the text of the most explicit discussions of the dualism, inside and outside the *Methods*. Rather, I focus on Sidgwick's treatments of the self-evident bases of utilitarianism and egoism. I claim that Sidgwick offers an argument, premised on the "real and fundamental distinction" between any one individual and any other, which is supposed to supply the self-evident basis of egoism. I suggest that this argument does not properly support egoism, but does properly support an important, related, weaker claim: that there are agent-relative reasons. I claim that the argument Sidgwick offers, supposed to supply the self-evident basis of utilitarianism, is less successful than his argument for egoism; its key claims are, in a way Sidgwick himself worries about particularly in the first edition, self-evident only because tautologous. Thus Sidgwick offers no proper support for the idea that there are self-evidently-based and conflicting fundamental principles of egoism and utilitarianism. These reflections provide some support for what I label "conflict-mitigating" interpretations of the dualism of practical reason, though this support has to be balanced against the textual evidence for "conflict-enhancing" interpretations supplied by Sidgwick's explicit characterizations of the dualism as

involving a contradiction.[10] But they also make Sidgwick the most important model for a range of historically distinctive and plausible hybrid views of practical reason, according to which there are both genuine agent-relative reasons and genuine agent-neutral reasons.

David Phillips

University of Houston

[10] I articulate the distinction between "conflict-enhancing" and "conflict-mitigating" interpretations of the dualism on pp. 134-5 of *Sidgwickian Ethics*.

David Phillips on Sidgwick: Some Comments

The shape of contemporary ethics owes a great deal to Henry Sidgwick, through his influence on Rawls, Parfit, and others. No one who reads David Phillips's outstanding book can be left in the slightest doubt about Sidgwick's continuing significance for both metaethics and normative ethics.[11] Phillips's scholarship and his substantive arguments are powerful and insightful, and I find them largely persuasive. So in these remarks I intend merely to raise a few questions about each of his four main chapters, several of which may well amount to little more than requests for clarification.

Metaethics

1. Phillips attributes to Sidgwick what he calls the 'realist conceptual thesis' (12-14):

> It is part of our moral concepts that there is such a thing as moral truth and error.

His evidence comes from the preface and the first chapter of the first edition of The Methods of Ethics.[12] As Phillips himself says, these passages 'treat ... realism [that is, the view that there is moral truth and error] as a presupposition of all moral enquiry'. But a presupposition does not have to be conceptual. So when Sidgwick says in the passage quoted by Phillips from the first chapter that the view that rightness depends on the agent's beliefs about rightness implies that the 'common notion of morality must be rejected', he need not be taken as making a conceptual

[11] David Phillips, *Sidgwickian Ethics* (Oxford: Oxford University Press, 2011). All unattributed page references are to this book.
[12] Henry Sidgwick, *The Methods of Ethics* (London: Macmillan, 1874), pp. v-vi; 5-6. References to *ME* in the text below are to the 7th edition (1907).

claim. What he is rejecting is not a particular understanding of the concept of morality, but a conception of it.

2. G.E. Moore is Phillips's paradigm of a philosopher who claims to have a 'master argument' against any form of ethical naturalism. Sidgwick, in contrast, is said to offer only piecemeal, negative arguments against particular versions of naturalism (17). So 'we should not expect to find in Sidgwick any diagnosis for [the] ... failure [of these particular analyses] beyond the general idea that they fail to capture something central to the fundamental concept expressed by "right", "ought", etc.' (18).

> This 'general idea' is of course Sidgwick's 'fundamental distinction thesis':
>
> The fundamental notion represented by the word "ought" or "right", which [ordinary moral or prudential] judgments contain ... [is] ... essentially different from all notions representing facts of physical or psychical experience. (*ME* 25; see 12)

I fail to see, however, why this view does not itself amount to a 'master argument'. Sidgwick's claim is that any form of naturalism will be found to contradict the fundamental distinction thesis. This appears to be structurally analogous to Moore's suggestion that any form of naturalism will close certain questions which we believe are open. So in the case of Bernard Williams, for example, we can predict that his failure will consist in unacceptably denying the fundamental distinction thesis (20-21); and in the case of Stephen, we find Sidgwick wheeling out that thesis using the very same terminology as in the key statement in *ME* (cited by Phillips at 27).

3. Phillips closes his chapter on metaethics with the suggestion that Sidgwick's dualism might imply that he ought to have been, and perhaps sometimes was, an error theorist:

> The central thought is that if we have a body of beliefs that we take to express positive truths about a certain subject matter, but these beliefs turn out to involve a fundamental contradiction, then we are not entitled to think that there are positive truths about this subject matter. (45)

I am inclined to think that Sidgwick would have considered this position both too quick, and too pessimistic. But there is also an internal question. If he accepted the realist conceptual thesis and the realist substantive thesis (the view that there is moral truth and error), and on the basis of these would have denied non-cognitivism (28), then why would this not have led him to reject error theories for the same reasons?

One further question. I wonder whether Phillips might garner further support for his interpretation here from the final paragraph inserted into the *Methods* after the first edition. There, in accepting the possibility that ethics may turn out to be analogous to science in our taking certain claims to be true merely because we have a strong disposition to believe them and they establish coherence, Sidgwick could be read as allowing for the 'invention' of ethical principles.

Epistemology

1. According to Phillips, Sidgwick fails to draw 'fully or explicitly' a distinction between weaker and stronger self-evidence – that is between a conception of self-evidence as less or more conclusive (59). In a footnote, he continues:

> On the interpretation I favor, [Sidgwick] does in effect make the distinction in part. As I see it, Sidgwick has the idea of *apparent self-evidence*, the idea that the fact that a proposition seems compelling without reference to evidence beyond itself does not provide conclusive justification for believing that proposition, and that additional independent evidence in favor of believing it can be supplied by its passing the third and fourth tests [Phillips is here referring to the four tests for the highest degree of certainty at *ME* 338-42]. The additional idea he does not clearly have is the idea that *some but not*

> *apparently conclusive* evidence in favor of a proposition might be supplied by contemplating that proposition without reference to evidence beyond itself. (86 n9)

I am inclined to think that Sidgwick is fully aware of the distinction between more and less conclusive conceptions of self-evidence, and in fact distances himself from the idea that any view held as self-evident can *ever* be said to be conclusive: 'By cognition I always mean what some would call "apparent cognition"' (*ME* 34 n2; see *ME* 211). So all we have is the idea of a proposition's being *maximally conclusive*, in the sense of its having passed all four tests as fully as possible in the circumstances. But here I may be misunderstanding what Phillips had in mind.

2. Phillips states Sidgwick's fourth test as: 'The proposition must be universally accepted' (60). This positive condition seems to me significantly stronger than the negative condition Sidgwick in fact gives at *ME* 442:

> [I]f I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality.

To pass the fourth test, a proposition I believe to be self-evident must not be one I know some epistemic peer of mine disagrees with. I do not even have to ask the question whether some epistemic peer *might* disagree with it, let alone check on whether *everyone* accepts the proposition (a test which no even slightly substantive proposition could ever meet).

3. Phillips raises an important and interesting puzzle about Sidgwick's strategy in his chapter on the proof of utilitarianism (4.2), where he makes no reference back to the independent 'criterial' arguments of Book 3 and focuses entirely on what Phillips calls 'bipartite' arguments, which are

ad homines and rely on the idea that the homines addressed have beliefs which are not self-evident but provide a route towards beliefs which are self-evident (65-78).

Phillips's own explanation of this is that Sidgwick himself believed that the criterial argument had failed, but thought that he could nevertheless, using bipartite arguments, show that utilitarianism is superior to dogmatic intuitionism. It seems to me that one could argue that Sidgwick *should* have taken this pessimistic view of the criteria argument, primarily because of the large amount of disagreement with utilitarianism by epistemic peers. But as far as I can see, there is no evidence that he in fact does, except for the 'puzzle' of 4.2 itself.

I myself am inclined to accept something like Phillips's second solution of the puzzle, according to which Sidgwick's aim in 4.2 is to focus only on arguments for one method as against another (73). His objection to this suggestion is that Sidgwick believes that in that context criterial arguments can still be helpful. But this assumes that criterial arguments have not already been rejected. As I understand the context of 4.2, Sidgwick does indeed see the criterial argument as having failed – but only rhetorically, not as an argument in itself. He is considering what might be said to someone who is unpersuaded by the (perfectly good) arguments of book 3 to shift them from their current beliefs to a conclusion which, had they considered things properly in the first place, they might have arrived at using the criterial argument (see esp. *ME* 419).

4. Does Sidgwick give any credence to common-sense morality? Like David Brink and others, I am inclined to think he does. Consider in particular the passage from *ME* 373, cited by Phillips at 75:

> [T]he truth of a philosopher's premises will always be tested by the acceptability of his conclusions: if in any important point he be found in flagrant conflict with common opinion, his method is likely to be declared invalid.

Phillips disagrees (75-6), claiming that Sidgwick 'clearly has ... a conception of criterial argument which gives no evidential role to common-sense morality'. This claim, however, is ambiguous. It could mean that criterial argument is in itself independent of common-sense morality; or that the claim that common-sense morality plays no such evidential role is part of criterial argument itself. Phillips I presume intends the stronger claim, and yet he provides no evidence for it. And he explains passages such as that on 373 as involving a confusion on Sidgwick's part between philosophical and dogmatic intuitionism, so that his conception of the former there mistakenly includes a commitment to the latter.

Given Sidgwick's own formidable intellect, and the fact that the *Methods* was so carefully revised so many times, this strikes me as an implausible interpretation. More plausible is that his epistemology includes both elements of foundationalism and a commitment to some kind of Aristotelian dialectic, in which the results of philosophically intuitionist theorizing are checked against common-sense morality as a whole. (Recall Sidgwick's claim in the preface to the 6th edition that he deliberately set out to 'imitate' Aristotle.)

5. There is a particularly interesting discussion of the implications of the four tests for Sidgwick's epistemology at 81-4. Here Phillips argues that Sidgwick is not what he calls an 'extreme foundationalist' – that is, a foundationalist who would deny that a proposition's passing the test could be part of our justification for accepting it. Phillips sees Sidgwick as a moderate foundationalist who does not deny this role to the tests, but who nevertheless is not a coherentist

because the criteria are justified by appeal to what the opponent Sidgwick has in mind – the natural sceptic – already accepts.

This seems to me an ingenious and suggestive interpretation. But I wonder whether Sidgwick might also be understood as an extreme foundationalist (as far as the tests are concerned) who sees the passing of the tests as itself part of what it is for a proposition to be *genuinely* as opposed to merely *apparently* self-evident. Then self-evidence itself would still be doing all the epistemic heavy-lifting

Utilitarianism versus Dogmatic Intuitionism

1. Phillips claims that Sidgwick does not see hedonism as self-evident (97, 111 n7). His evidence is that Sidgwick treats it in 3.14 rather than in 3.13 (the chapter on philosophical intuitionism) and does not list it as an axiom. And in 3.14 he notes the disagreement about hedonism.

I agree that Sidgwick's not discussing hedonism in 3.13 requires explanation. I suspect that he wanted to dedicate a whole chapter to the good and hedonism, and it might be said to be telling that this chapter follows immediately on from 3.13 and that 3.13 ends with an explicit link to the 'more indirect' argument of 3.14. It is also true that Sidgwick is aware of the disagreement about hedonism, and tries to deal with it. But the same is of course true of utilitarianism and dogmatic intuitionism.

We should expect hedonism to be self-evident for Sidgwick, since it is required for his version of utilitarianism, and he takes utilitarianism to be justified by intuition. And early in the book (*ME* 97-8) Sidgwick implies that the alternative to inductivism as a support for hedonism is intuitionism. Further, at *ME* 400, Sidgwick asks his reader to consider his own intuitive

judgement, in the same way as he was asked to do 'in considering the absolute and independent validity of common moral precepts' (see also *ME* 406-7).

2. Phillips rightly sees Rossian deontology as a major competitor to Sidgwickian utilitarianism. But the claim that 'Sidgwick, of course, never directly encountered Ross's position' (105), though obviously true, strikes me as potentially misleading. Whewell's position is similar in various significant ways to Ross's,[13] and even more importantly so is that of Aristotle (Ross himself sees his own view as Aristotelian[14]).

I think also that Phillips is correct in thinking that Ross's account of promising is the best potential counter-example to Sidgwick's utilitarianism. But Sidgwick also fails to recognize the important role played by the capacity for practical judgement in the ethics of Aristotle, Whewell, and Ross, and – I believe – in any plausible moral theory. Sidgwick held a scientistic view of ethics, according to which moral theories must aim to guide action in each particular case, without reliance on judgement. But this goal is a chimera. Any theory must itself be judged to be correct in the first place, and any plausible one, including utilitarianism, will require judgement, not explicitly guided by the theory, to be exercised in its application to particular cases. This seems to me another major failure in Sidgwick's position which Phillips might have said more about.

[13] See esp. his *The Elements of Morality, Including Polity*, 2 vols. (London: John W. Parker, 1845).
[14] See e.g. *The Foundations of Ethics* (Oxford: Clarendon Press, 1939), 5.

Utilitarianism versus Egoism

1. Phillips is inclined to commit Sidgwick to what one might call a 'permissibility' conception of the dualism of practical reason. He cites a passage from the first edition of the *Methods* which expresses the axiom of justice with reference to what is 'right', 'reasonable', 'the dictate of reason', and 'my duty'[15], and notes that 'reasonable' 'in ordinary language tends to suggest permission'. But if Sidgwick were here mixing concepts of requirement with a concept of permission, he would of course be deeply confused. 'Reasonable' for him is, in such contexts, a technical term: the reasonable action is the one there is strongest ultimate reason to do. Nor does it make a difference that he uses 'reasonable' elsewhere in the ordinary sense (Phillips ibid.).

My own view is that Sidgwick tends to use too many concepts to express his position, the passage here quoted by Phillips being an excellent example of that. That position could be stated purely in terms what we have ultimate reason to do, with no reference to permission, requirement, 'ought', duty, or whatever. So – *pace* Phillips 153 n17 – I would not want to attribute a 'permissive' interpretation to Sidgwick, nor even to advocate myself such a version of a 'dual source' view of practical reasons. For me to know what to do, all that is required is that I know what I have strongest ultimate reason to do. To ask whether I am permitted or required to act in that way is to ask an unnecessary and potentially confusing question.

2. I do find the version of permissivism Phillips advocates, and ascribes to Sidgwick (149-51), a lot more plausible than Sidgwick's own pessimistic and 'chaotic' dualism, and I find it a little mysterious that Sidgwick never considers such a position. I suspect one reason may be that such view relies on judgement about how 'serious' consequences are for the various individuals in the

[15] *The Methods of Ethics*, 1st edn., cited at 116.

relevant outcomes of actions (147). Sidgwick would have seen such a view as a version of 'perceptional intuitionism' and thus to be rejected out of hand. But of course one might attempt to make such a view somewhat more precise, by attaching weightings to one's own interests as against those of others, and it is not obvious that such a view would be any more imprecise than the egoistic and universalistic versions of hedonism Sidgwick does discuss.

Roger Crisp

St Anne's College, Oxford

Oxford Uehiro Centre for Practical Ethics

Utilitarianism and Egoism in *Sidgwickian Ethics*

In his excellent Sidgwickian Ethics, David Phillips argues that Sidgwick's argument for utilitarianism from the axioms is less successful than Sidgwick believes. He also argues that Sidgwick's argument for egoism is more successful than this argument for utilitarianism. I disagree. I close by noting, briefly, a possible solution to an epistemological puzzle in Sidgwick that Phillips raises.

I

Phillips takes the argument for utilitarianism to have two premises:

(U) The good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other.

(R) As a rational being I am bound to aim at good generally…not merely at a particular part of it (121).[16]

Phillips thinks (R) is analytic and that Sidgwick "should be happy to admit" this (124). (R) is analytic because for Sidgwick, "to say something is good just is to say there is reason to aim at it" (124).

I think that Sidgwick would not be happy to admit that (R) is analytic, nor need he admit this. As Phillips notes, Sidgwick is very concerned to show that his axioms are not tautologies (123). And Sidgwick's account of goodness does not commit him to saying that (R) is analytic. Sidgwick defines "'ultimate good on the whole,' unqualified by reference to a particular subject" as "what as a rational being I should desire and seek to realize, assuming myself to have an equal

[16] Bare parenthetical references are to David Phillips, Sidgwickian Ethics (New York: Oxford University Press, 2011).

concern for all existence" (ME 112).[17] Suppose that "ultimate good on the whole, unqualified by reference to a particular subject" is the same as "good generally." "I ought to aim at good generally" becomes "I ought to aim at what I ought to desire and seek to realize, assuming myself to have an equal concern for all existence." That is not analytic.

But Phillips does not think Sidgwick's argument fails because (R) is analytic. He takes (U) to be the problem. The first problem is that an egoist can deny "that there is such a thing as universal goodness" (125).

I do not think that (U) unpacks the concept of "universal goodness." Phillips argues that it does mainly by citing the first edition version of (U) (121-2). But even if he is right about the first edition, in the later editions Sidgwick seems to unpack not "universal goodness" but rather what it is to take up the point of view of the universe (or at least has this as the antecedent of (U)). And putting the egoist's disagreement in terms of denying that there is universal goodness is puzzling. Phillips's idea is that the "egoist can admit…that some people have the concept of universal goodness; but the egoist will deny that that concept is ever instantiated" (152n3). The egoist, then, does not find the concept confused (like "round square"), but rather uninstantiated (like "unicorn"). But this seems an odd thing to say: surely there is something I ought to desire, assuming myself to have an equal concern for all existence (or, alternatively, surely there is something that is the set of the goods of each individual). It seems preferable to say instead that the egoist is uninterested in this thing—or, as Sidgwick says, refuses to take up the point of view of the universe.

But this disagreement is not so important, since Phillips notes that Sidgwick, given the dualism, might admit this first problem (125). Whether the problem is put in terms of taking up

[17] ME = Henry Sidgwick, The Methods of Ethics (Indianapolis: Hackett, 1981); GSM = Lectures on the Ethics of T. H. Green, Mr. H. Spencer, and J. Martineau (London: Macmillan, 1902); FEC = "Some Fundamental Ethical Controversies," Mind 14, 1889, 473-87.

a point of view or instantiating a concept does not affect this admission. The more serious, second problem Phillips raises for (U) is that, even if true, and even if one does take up the point of view of the universe (or think "universal good" is instantiated), (U) cannot justify utilitarianism. Consider a view on which "(a) any person's happiness is good; i.e., there is a requirement of reason for any agent, ceteris parabis, to promote it; and (b) there is a special requirement of reason for any agent to promote his own happiness three times as much as he promotes the happiness of anyone else" (125). On this view, (U) is true, but what one ought to do is not act as a utilitarian, but rather "promote the weighted sum of her own good and everyone else's" (125). Phillips puts his objection this way: "Sidgwick slips from the (proper, but nonsubstantive) idea that if there is such a thing as universal goodness there is some reason to aim at it, to the (substantive, but not properly supported) idea that it is the only thing that ought to be aimed at or promoted" (125) The real problem, he thinks, is that (U), like (R), is analytic, and so cannot rule out aiming at other things. (Phillips does not say explicitly that (U) is analytic, but he does write that "the argument…contains only tautologies" (125).)

I have three comments on this objection.

(1) Insofar as the objection is that Sidgwick cannot go directly from the axioms to utilitarianism, Phillips is surely correct. Sidgwick thinks common sense moralists agree with his axioms (ME 421, GSM 331-2). He also thinks Clarke and Kant agree with them (ME 384-6). So he cannot understand the axioms as ruling out reasons to act in non-utilitarian ways. But the obvious fix is to think that Sidgwick supposes that he has already, in Methods III.XI, ruled out any additional axioms (such as Phillips's (b)). Whether that argument works is another matter (considered by Phillips in his pretty convincing chapter 4). But the issue is the success of that further argument; Sidgwick is not defeated simply by noting the possibility of a view like (a)-(b).

(2) The objection in (1) can be made without thinking that (U) is analytic. One might think that (U) says that "from a point of view from which I give no special weight to myself, I ought to be indifferent between distributions that result in the same amounts of good." This is not analytic, in part because giving no special weight to myself is consistent with favouring one distribution over another on grounds other than the amount of goodness. But if so, (U) by itself (or with (R)) does not secure utilitarianism.

(3) I am not sure how Phillips understands (U). He seems to gloss (U) as saying that "there is such a thing as 'goodness' period" (124), or as saying that "his happiness is good, not just good for him" and "his happiness cannot be a more important part of good, taken universally, than the equal happiness of any other person" (125). He also thinks of "the premise" of the argument for utilitarianism as saying that "there is such a thing as…good-from-the-point-of-view-of-the-universe" (125). But (U) seems to say only that from the point of view of the universe, I ought to be indifferent between equal amounts of good. (U) does not, then, entail that one does take up the point of view of the universe. Nor is it analytic, even on what I take to be Phillips's reading: if (U) claims that "universal good" is instantiated, Phillips says the egoist disagrees, but presumably not with an analytic truth.

II

Phillips takes the argument for egoism to have two premises:

1.The distinction between any one individual and any other is real and fundamental.

2. If the distinction between any one individual and any other is real and fundamental, then I ought to be concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I ought not to be concerned with the quality of the existence of other individuals.

Therefore,

3. I ought to be concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I ought not to be concerned with the quality of the existence of other individuals (127-8).

Phillips notes that 3. is weaker than egoism, since special concern for myself is not exclusive concern for myself. But he thinks the argument does better than the argument for utilitarianism, for two reasons: 2. is "both substantive and self-evident," and 3., though not egoism, is the significant claim that there are agent-relative reasons (129).

Phillips glosses 1. as follows: "I have a special connection to certain goods and bads (e.g., I directly experience certain pains and pleasures)….The key idea is just that of a kind of special connection: that the fact that a certain pleasure or pain is mine means I experience it in a way others do not" (129, 131). 2. is compelling because if 1. is true, "that fact gives me special reason to want and pursue those goods, and to not want and avoid those bads….My reasons are reasons <u>for me</u>. How could the fact that something has a special effect on <u>me</u> not affect <u>my</u> reasons?" (129)

I have four comments on this argument.

(1) Phillips describes 2. as a normative premiss (130). This lets him avoid the objection that the argument deduces a normative conclusion (3.) from wholly non-normative premises (1.). But this still violates Sidgwick's rule that normative conclusions cannot be deduced from wholly non-normative premises. For 2. says that a wholly non-normative claim entails a normative claim.

(2) As it stands, the argument seems to depend on a particular view of the good. The special connection is that I directly experience the things that are good. This may be plausible if

the things that are goods are mental states. But I would not have thought that the argument for agent-relative reasons depended on holding any particular view about what is good. For example, some think that what is good is the state of affairs (whether that is a mental state or not) that satisfies my preference. I have no special connection to the obtaining of that state of affairs—I might not know that it obtains, and no state of me need be a part of it. Phillips might reply that for such goods, there is no case for agent-relative reasons. But that has odd results: for example, I then have no agent-relative reason to care about how my children fare after my death.

(3) I take it that Phillips's idea in 1. is this: we are creatures such that if I experience a pleasure or pain, it does not follow that anyone else experiences it. In this sense, we are distinct individuals, as would not be true if, for example, we were creatures with direct empathetic connections, or there was a sea of pleasures and pains without distinctions between individuals. The question is why this gives me a special reason to care about my own pleasure or pain, over and above any reason I have to care about the occurrence of pleasure or pain somewhere. Phillips's thought seems to be that since it is a state of me (and not you), it must give me a special reason. But this seems awfully close to simply asserting what was supposed to be the conclusion of the argument, viz. that I have a special reason to care about states of myself.

I can see one way in which 1. is part of an argument for 3. Phillips goes on to endorse Roger Crisp's "Two Doors 3":

> You are confronted with two doors. If you do not pass through one or other of them, you will suffer an extremely painful electric shock. If you pass through door A, you will experience a less painful but significant shock. If you pass through door B, you will not experience this shock, but some other person, a stranger and out of sight, will suffer a

> shock of the same intensity….Surely you have a reason—a strong reason—to choose door B over door A grounded in the fact that it is that door which will significantly promote your well-being (148-9).

1. makes this kind of case possible, and our intuition about this case then constitutes an argument for 3.

I do not share this intuition. Since some others do, I see how this might be an effective argument for agent-relative reasons. It is not really, however, the argument of 1.-3.. That argument promised a justification for agent-relative reasons that goes beyond just giving cases in which some think we have them.

(4) Phillips reconstructs Sidgwick as giving a deductive argument for egoism. I think Sidgwick intends something weaker. In the Methods, after noting 2., he writes that "I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual" (ME 498). The point seems to be that if someone thought that the distinction between one individual and another did matter to choosing between egoism and utilitarianism, it is hard to see how to show that such a person is wrong. It does not follow that one must think it matters. Similarly, in "Some Fundamental Ethical Controversies," where the argument first appears, Sidgwick writes that "the proposition that this distinction is to be taken as fundamental in determining the ultimate end of rational action for an individual cannot be disproved" (FEC 484). Of the utilitarian position, Sidgwick writes "I do not deny this position to be tenable; since, even if the reality and essentiality of the distinction between one individual and another be granted, I do not see how to show its fundamental practical importance to anyone who refuses to admit it" (FEC 485). Sidgwick's view seems to be that the distinction between individuals is something to which more than one response is

permitted. He, and egoists, take it to be relevant to setting ultimate ends. Others, such as utilitarians, may disagree.

III

Phillips notes that in Methods IV.II, the "proof" of utilitarianism proceeds by ad hominem arguments directed at the egoist and the common sense moralist. The application of the tests for highest certainty in III.XIII, which seemed at least to have established consequentialism, and which elsewhere Sidgwick endorses as the best way of justifying beliefs, seems to have been forgotten (65-76). Earlier I suggested that there is a gap between the axioms and utilitarianism. If so, Sidgwick has reason to think the tests are insufficient: they secure the axioms but not utilitarianism. This is slightly different than Phillips's solution to the puzzle, according to which Sidgwick simply saw that the appeal to the tests failed. On my view, the appeal to the tests succeeds in establishing the axioms—which is why Sidgwick does not note that the argument of III.XIII fails.

Robert Shaver

University of Manitoba

Some Remarks on David Phillips's *Sidgwickian Ethics*

David Phillips's Sidgwickian Ethics is a penetrating contribution to the scholarly and philosophical understanding of Henry Sidgwick's The Methods of Ethics.[18]

This note focuses on Phillips's understanding of (aspects of) Sidgwick's argument for utilitarianism and the moral epistemology to which he subscribes. In § I, I briefly outline the basic features of the argument that Sidgwick provides for utilitarianism, noting some disagreements with Phillips along the way. In § II, I raise some objections to Phillips's account of the epistemology underlying the argument. In § III, I reply to the claim that there is a puzzle at the heart of Sidgwick's epistemology. In § IV, I respond to Phillips's claim that Sidgwick is unfair in his argument against the (deontological) morality of common sense.

I

One aim of The Methods of Ethics is to provide an argument for classical utilitarianism. For Sidgwick, this involves, in part, showing that there exist certain non-derivatively warranted propositions, i.e., intuitions, on which the truth of utilitarianism depends. He arrives at these in the penultimate chapter of Book III after a long and exhaustive survey of common-sense morality. The most important of these are:

(U): "The good of any one individual is of no more importance, from the point of view…of the Universe, than the good of any other; unless, that is, there are special grounds for believing that more good is likely to be realized in the one case than in the other" (ME 382), and

[18] Abbreviations: SE = David Phillips, Sidgwickian Ethics (Oxford: University Press, 2011); ME = Henry Sidgwick, The Methods of Ethics, seventh edition (Indianapolis: Hackett, 1907); PC = "Professor Calderwood on Intuitionism in Morals," Mind 1 (1876), 563-566; EP = "Establishment of Ethical First Principles," Mind 4 (1879), 106-111; LK = Lectures on the Philosophy of Kant and Other Philosophical Lectures and Essays, ed. James Ward (London: Macmillan, 1905); LE = Lectures on the Ethics of T. H. Green, Mr. Herbert Spencer, and J. Martineau, ed., E. E. Constance Jones (London: Macmillan, 1902); FC = Henry Sidgwick, "Some Fundamental Ethical Controversies," Mind 14 (1889), 473-487.

(R): "It is evident to me that as a rational being I am bound to aim at good generally, – so far as it is attainable by my efforts, – not merely at a particular part of it" (ME 382).

From (U) and (R), Sidgwick infers the maxim of benevolence: "that each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him" (ME 382). This expresses the idea that one is required to give each sentient being's good weight proportionate to its quantity in one's reasoning about what to do. It is clear that for Sidgwick the maxim or principle of benevolence consists in and is equivalent to the requirement that one promote or pursue (surplus) aggregate good even at the expense of promoting one's own (surplus) good (ME xxi, 382, 385, 392, 400).

Phillips is not clear on what he thinks Sidgwick gets from (U) and (R). He suggests at times that Sidgwick infers from them the bulk of utilitarianism (SE 64, 68, 102, 125). At other times, he suggests that Sidgwick infers something more like the principle of rational benevolence (SE 118). This is understandable. Sidgwick is unclear. Soon after arriving at (U) and (R) he says "Utilitarianism is thus presented as the final form into which Intuitionism tends to pass, when the demand for really self-evident principles is rigorously pressed" (ME 388; also 406-407). But Sidgwick more frequently describes what he gets from (U) and (R) as the axiom or maxim of rational benevolence or first principle of utilitarianism (ME 387, 400, 418, 419, 421). This serves as the "basis" of utilitarianism (ME 387). (U) and (R) do not alone, then, get Sidgwick utilitarianism or the utilitarian method of ethics.

This appears to be confirmed by the fact that Sidgwick does not believe that (U) and (R) supply a method of ethics, that is, a "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" (ME 1). The problem is that the intuitions are "of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case; particular duties have still to be determined by some other method" (ME 379).

It is important to note that Sidgwick thinks that the abstractness point applies to (U), (R) and the principle of rational benevolence. He twice describes the last as abstract (ME 382, 462n1).[19] It would not, then, be plausible to think that he means his comment to apply to (U) and (R) and not the principle of rational benevolence, which could serve as a method of ethics. This fits with the primary focus of the discussion in which Sidgwick lays out his intuitions, which is principles rather than methods. He does not turn to the utilitarian method of ethics until Book IV of ME.

That Sidgwick aims to get utilitarianism's first principle from (U) and (R) and not the utilitarian method of ethics makes it possible to reply to Phillips's claim that there is a puzzle at the heart of Sidgwick's epistemology and that his argument against common-sense morality is unfair. Before articulating these replies (in §§ III and IV) it is important to outline the other features of Sidgwick's argument for utilitarianism.

Sidgwick's argument for utilitarianism relies in addition on showing that the fundamental moral requirements of common-sense morality do not, presumably in his view unlike (U) and (R), possess the "characteristics by which self-evident truths are distinguished from mere opinions" (ME 338). To qualify as self-evident a proposition must be clear and precise, self-evident on reflection, consistent with other self-evident propositions one accepts, and not denied

[19] He also frequently says that the principle of rational benevolence is an intuition (ME 388, 400, 421, and 462n1). This is misleading. Officially it is represented as an inference from (U) and (R) (ME 382).

by a competent judge who one believes is no more likely to be in error than oneself (ME 338-342).

Sidgwick thinks that when we reflect on the rules of common-sense morality, including the requirements of fidelity to promises, of justice and of veracity, we see that they fail to possess one or other of the characteristics of self-evidence (ME 360). In sum, his argument is that if you leave the main requirements of common-sense morality vague – e.g., you ought to keep your promises – there is little or no dissent from them. In this case, the requirements possess the fourth but lack the first characteristic. However, if you try to give them "the definiteness which science requires", it is no longer true that there is little or no dissent from them (ME 342). Competent judges disagree on the possible specifications of the rules of common-sense morality. In this case, the requirements of common-sense morality possess the first but not the fourth characteristic. In either case, the rules of common-sense morality fail to qualify as self-evident (ME 342).

Phillips calls the foregoing argument the criterial argument for utilitarianism (SE 64, 67, 95, 98). He rightly notes that Sidgwick uses a further argument in his attempt to establish utilitarianism. This argument takes the form of a proof directed to the proponent of common-sense morality (ME 420-422), which he describes as a "line of argument which on the one hand allows the validity, to a certain extent, of the maxims already accepted, and on the other hand shows them to be not absolutely valid, but needing to be controlled and completed by some more comprehensive principle" (ME 420).

Relying on this strategy, Sidgwick tries to show that the utilitarian method of ethics supports the main rules of common-sense morality in general and that it injects greater clarity, completeness and system into ethical thinking, thereby remedying the practical "defects" which

Sidgwick allegedly finds in common sense (ME 422). He thinks that this furnishes the proponent of common-sense morality with considerations sufficient to determine her mind to accept utilitarianism. Phillips refers to this argument as Sidgwick's bi-partite argument for utilitarianism (SE 63).

II

To understand and reply to Phillips's claim that there is a puzzle at the heart of Sidgwick's epistemology we must get clear on its nature.

It is relatively clear that Sidgwick's argument for utilitarianism relies on an appeal to an intuitionist epistemology. He thinks that (U) and (R) are, for example, self-evident, that is, non-inferentially warranted propositions or intuitions (ME 382, 383, 387, 421).

Phillips agrees that the appeal to self-evidence plays an important role in Sidgwick's account of the justification of (U) and (R) (SE 60-65). It is not the only thing to which Sidgwick appeals, however. Phillips maintains that in the criterial argument Sidgwick relies in addition on coherence considerations to justify (U) and (R) (SE 61-62, 79). (U) and (R) are both inferentially and non-inferentially justified (SE 81). Sidgwick is therefore a proponent of what Phillips calls "moderate foundationalism" (SE 80).

Phillips's case for this interpretation depends in part on a particular understanding of Sidgwick's characteristics of self-evidence. Sidgwick describes these as "four conditions, the complete fulfillment of which would establish a significant proposition, apparently self-evident, in the highest degree of certainty attainable" (ME 338). The most important of these for our purposes is the fourth condition, which, on Phillips's reading, states that for a proposition to be self-evident it "must be universally accepted" (SE 60). In passing this test, a proposition gains

epistemic credibility. In satisfying the test a proposition acquires "consistency with ordinary moral opinions", which themselves have "probative" value (SE 79, 84).

The difficulty with this suggestion is that it is in tension with a number of other things that Phillips says. At one point he claims that Sidgwick offers "a conception of criterial argument which gives no evidential role to common-sense morality" (SE 76). This might just be an oversight on Phillips's part, but actually this seems to be the view that he is forced to accept given his interpretation of Sidgwick's bi-partite argument for utilitarianism. The received view of the bi-partite argument is that Sidgwick grants probative value to common-sense morality and that he thinks that utilitarianism's ability to capture and explain key elements of it provides one with justification for believing it. Phillips rejects this reading of the argument in favor of the view that the bi-partite argument is an ad hominen argument directed to the proponent of common-sense morality in which Sidgwick grants no evidential role to common-sense morality (SE 74-76). This suggests that Phillips's view is that Sidgwick grants probative status to common-sense morality in the context of his criterial argument but not in the context of the bi-partite argument. This is puzzling, however. Why grant probative status or evidential value to common-sense morality in the case of the criterial argument but not in the case of the bi-partite argument? The puzzle is deepened by the fact that Phillips grants that common-sense morality might, in the end, possess "imperfect certitude" (SE 84).[20]

Phillips has two possible replies.

(1) He might argue that when Sidgwick appeals to "ordinary moral opinions" he is appealing not to the moral beliefs of the plain man but to the moral beliefs of "(well-informed) [moral] experts" (SE 79; cf. SE 61).

[20] It does this by possessing one of the characteristics of self-evidence (SE 84).

There are two difficulties with this reply. First, some of the so-called moral "experts" or "competent judges" to which Sidgwick appeals are themselves exponents of common-sense moral opinions (e.g., Whewell and Kant). Second, in the discussion of (U) and (R) Sidgwick is keen to note that the "plain man" does not disagree with the principle or maxim of rational benevolence (ME 382; also LE 331-332).

(2) He might avoid this charge by exploiting a distinction that he makes between common-sense morality broadly construed and common-sense morality narrowly construed (SE 78). He describes the first as consisting of "ordinary moral opinions", including "those best systematized by dogmatic intuitionism and those best systematized by [egoism and utilitarianism]" (SE 79). The second is described as a "subset of common moral beliefs that Sidgwick systematizes as dogmatic intuitionism" (SE 78). The idea might be that Sidgwick appeals to the former but not the latter in the criterial argument.

There are two infelicities with this suggestion. First, it is not obvious that Sidgwick cuts the distinction between common-sense morality narrowly construed and common-sense morality broadly construed. Phillips relies on scant evidence to show that this is Sidgwick's view. The passage that he does cite to justify his interpretation seems to list things – concern for the general good, prudence, and so on – which find a firm home in common-sense morality narrowly construed (SE 79; see ME 14, PC 564). The situation is made worse for Phillips by the fact that he appears to give no indication that Sidgwick thinks common-sense morality broadly construed has probative force. (At least in the case of common-sense morality narrowly construed there is some evidence that Sidgwick thinks it has evidential value (see, e.g., ME 373).)

Second, it does not seem possible to think that while common-sense morality narrowly construed has no probative status, common-sense morality broadly construed does. The latter

includes the former (SE 79). In addition, the most obvious arguments (appeals to agreement, for example) provided for thinking that common-sense morality broadly construed has probative status seem to imply that common-sense morality narrowly construed has it, too.

Perhaps Phillips's main concern is simply to show that common-sense morality narrowly construed has no "special" epistemic role in justification (SE 79). But this is consistent with it having some epistemic role, and this is all that is needed to secure the received view of the bi-partite argument.

Phillips's problems can be traced back to a misunderstanding of the fourth of Sidgwick's "characteristics" of self-evidence. In his attempt to separate self-evident truths from mere opinions Sidgwick is not *pace* Phillips concerned to show that the justification of self-evident propositions is amplified when it possesses the characteristic of being agreed to by others. It would after all be odd for Sidgwick to appeal to Kant for justification in this regard when he rejects so many of Kant's other views (see, e.g., ME 209-210, 222-223, 389-390, 394-395).

Rather, Sidgwick's idea seems to be that dissent functions as a potential defeater of an intuition, perhaps explaining why he describes the absence of dissent as an "indispensible negative condition of the certainty of our beliefs" (ME 342). On this view, a proposition qualifies as self-evident only when rational disagreement about its truth is absent or rationally explained away. A proposition lacks self-evidence, not when it lacks universal acceptance, but when there is disagreement about its truth and where "I have no more reason to suspect error in the other mind than in my own" (ME 342). Sidgwick is worried about disagreement that cannot be rationally explained away, because disagreement, when undefeated, impugns a proposition's claim to self-evidence. This need not involve holding the claim that agreement enhances probative status.

Of course, Phillips argues that his account of the tests is supported by Sidgwick's discussion of similar tests in his writings on epistemology. But this provides inconclusive results. In these writings, Sidgwick says only that a proposition counts as self-evident when it "does not conflict with the beliefs of other persons competent to judge" (LK 465; SE 82).

III

Phillips maintains that there is a puzzle at the core of Sidgwick's moral epistemology. He holds that Sidgwick relies on both the criterial argument and the bipartite ad hominem argument to establish ethical first principles (SE 65-69).[21] Sidgwick offers an independent criterial argument in ME III.xiii, only to abandon it in ME IV.ii, where he offers only the ad hominem bipartite argument. The puzzle is this: why does Sidgwick abandon the independent criterial argument in IV.ii where it is "clearly relevant" (SE 71; also 69)?

Phillips considers and rejects a solution according to which the independent criterial argument drops out in IV.ii because the context makes only bipartite arguments relevant (SE 73). This seems to me to be the most persuasive solution, however.

As noted, in the context of the criterial argument Sidgwick aims to establish the first principle of utilitarianism, and that there are no other such competing propositions within common-sense morality. He thinks that the proponent of common-sense morality accepts utilitarianism's first principle (ME 382, 421), and that competent judges accept it, too, e.g., Kant and Clarke (ME 384-385).

The criterial argument does not establish that the correct method of moral reasoning is utilitarian. Indeed, it does not appear by Sidgwick's own lights to directly or immediately establish any method of moral reasoning (ME 379). The criterial argument cannot, then, alone establish a method of ethics. This is where the ad hominem bi-partite argument enters the

[21] Phillips finds support for this at EP 106-107.

picture. In the context of the bi-partite argument, Sidgwick is trying to establish the utilitarian method of ethics. The common-sense moralist disagrees, of course, with utilitarianism (LE 331-332). It is unsurprising, then, that in IV.ii where the utilitarian method of ethics is the focus he attempts to show that there are considerations that determine the mind of the common-sense moralist to accept the utilitarian method.[22]

In light of what the arguments are designed to do, it is not at all puzzling that Sidgwick would deploy two arguments and that he would not rely on an independent criterial argument in Book IV. In Book IV the focus is on establishing a method of moral reasoning, and the criterial argument, Sidgwick admits, does not alone establish this. He thus needs another argument to get from the principles that he thinks are self-evident to the utilitarian method of ethics.

This solution to the puzzle stands even when we agree with Phillips that the criterial argument does not secure the truth of the hedonism or the truth of the maximizing conception of rationality that Sidgwick's utilitarianism includes (SE 68, 97, 102, 111n7, 118). If Sidgwick does establish hedonism and maximization in some way other than by appeal to intuition, this is not sufficient (with (U) and (R)) to establish utilitarianism. This would show only that the maxim of benevolence amounts to the requirement that one maximize surplus pleasure for the aggregate even at the expense of one's own greatest pleasure. Sidgwick would presumably still have to establish the "method of utilitarianism", which is the focus of ME IV.ii (ME 422). In addition, at least in the case of hedonism, a bi-partite argument is unnecessary since an argument is provided for the view before the bi-partite argument is presented (ME 391ff.). Indeed, Sidgwick says that the first principle of utilitarianism that he is working with at the outset of the bi-partite argument is that we ought to aim at "universal happiness" (ME 418).

IV

[22] He concedes that he cannot convince the egoist (ME 420-421).

Phillips holds rightly that Sidgwick's criterial argument involves showing that (U) and (R) do, but that the main rules of common-sense morality do not, possess the "characteristics by which self-evident truths are distinguished from mere opinions" (ME 338; see SE 95). Sidgwick is especially keen to show that the rules of common-sense morality lack the first characteristic of self-evidence, namely, clarity and precision in practical direction (ME 342-343; ME 360-361). This is a serious defect in an ethical method: "we study Ethics…for the sake of practice: and in practice we are concerned with particulars" (ME 215; cf. viii).

Phillips charges that in this argument Sidgwick is ultimately unfair to his opponent. He demands that the main rules of common-sense morality be "made precise enough to give determinate verdicts in every case" (SE 101). He makes no such demand of his own principles, however. The demand appears not to apply to his principles because these are "of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case; particular duties have still to be determined by some other method" (ME 379).[23] Phillips argues that it is unfair of Sidgwick to ask of common-sense morality that it furnish us with practically precise principles but not to ask the same of, for instance, (U) and (R) (SE 101-103).

This is a strong argument. One possible reply for Sidgwick is to admit that his philosophical intuitions are not themselves precise enough to provide determinate ethical verdicts in every case, but deny that this impugns them. His intuitions are of principles, not methods, and there is no requirement that principles deliver determinate verdicts in every case immediately. Phillips himself grants that a proposition can be self-evident without being immediately practical. He claims that Sidgwick's case for egoism depends on the truth of the following claim: "if the distinction between any one individual and any other is real and

[23] He must mean some method rather than some other method.

fundamental, then "I" ought to be concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I ought not to be concerned with the quality of the existence of other individuals" (SE 127; also ME 498, FC 484). This principle, Phillips maintains, is "self-evident (intuitively compelling)" (SE 129), but, like (U) and (R), it is not precise enough to provide determinate ethical injunctions in every case. Perhaps, then, Sidgwick thinks (U) and (R) are self-evident in this sense.

The problem with this reply is that it does not demonstrate the superiority of Sidgwick's principles to the rules of common-sense morality. For the proponent of common-sense morality might attempt to argue that (at least some of) its rules are self-evident in the same way as the principles (U) and (R) are. In this case, the proponent of common-sense morality might suggest that, e.g., the claim that "I ought to keep my promises", is a self-evident in the same way as (U) and (R) and that we should accept it as such even though it cannot "be made precise enough to give determinate verdicts in every case" (SE 101). It does, after all, like (U) and (R), have some practical value.

Sidgwick might, then, try a different reply to Phillips. He might grant again that his principles are not immediately clear and precise. They are in this sense no better than the rules of common-sense morality. However, unlike the rules of common-sense morality, Sidgwick might argue, his principles are mediately clear and precise. His principles satisfy the second condition of self-evidence mediately rather than immediately. This is because (U) and (R), together with other arguments, point to a method of ethics that is on the face of it precise enough to give determinate verdicts in every case, namely, utilitarianism. The principles (U) and (R) together with arguments about the good and with ad hominem arguments point to the method of utilitarianism which is (at least in principle) capable of delivering determinate ethical verdicts in

every case. This makes Sidgwick's principles superior to the rules of common-sense morality; they are mediately applicable.

The problem with the rules of common-sense morality is that unlike (U) and (R) there is no route from them to a method of ethics that is capable of delivering determinate verdicts in every case. Unlike (U) and (R) the main rules of common-sense morality do not seem to point to a method of ethics beyond themselves that is capable of delivering determinate verdicts in every case while at the same time relying on the resources found in common-sense morality itself. The difficulty for common-sense morality is that the main rules that it offers are put forward as though they are to function as a method of ethics. It offers them as immediately self-evident. Sidgwick might be right that these fail to be immediately self-evident. He cannot use this against them. His own principles fail to be immediately self-evident. He can, however, use the fact that these rules also fail to be mediately self-evident. His own principles do not, it seems, have this failing, and so they are a better approximation of clarity and precision.

There are two worries about Sidgwick's reply.

(1) It is not clear that the utilitarian method of ethics is capable of delivering determinate verdicts in every case. Sidgwick himself notes that there are problems with utilitarian moral reasoning (ME xxiii, 87, 140-150, 414, 461-462). The main problem is that we lack the epistemic power and resources to figure out the short and long term outcomes of the institutions, rules and actions that we seek to evaluate.

The only reply Sidgwick has to this (self-inflicted) criticism is to argue that the problems that affect the utilitarian method of reasoning are more tractable than those that affect the common-sense method of moral reasoning. He might claim that we know almost nothing of how to solve the problems that Sidgwick points out for the requirement of (say) justice without appeal

to utilitarianism or other (non-common-sense-based) principle, whereas we seem to be relatively clear at least on how to solve the problems with the acquisition of empirical information and for overcoming biases that plague utilitarian forms of moral reasoning. These problems more closely mirror those that are encountered in other domains of inquiry. In this case, Sidgwick might argue that his principles are closer approximations to clarity and precision than the rules of common-sense morality.

(2) Sidgwick must establish more clearly that common-sense morality does not possess the resources within itself for the purpose of dealing with its own practical defects. He will have to say more about what counts as common-sense morality, such that there is no element of common-sense morality that might help with its own practical defects, and about how he understands the alleged defects in common-sense morality.

Anthony Skelton

University of Western Ontario

Replies to Crisp, Shaver, and Skelton

It is a great privilege to have one's work critiqued by such a distinguished trio of philosophers and Sidgwick scholars. I owe further debts to Anthony and Rob, who were the OUP referees for my book. As will have been quite evident from the preceding discussion, they would not want to be held responsible for the book's detailed contents, on which they gave me much excellent commentary. But, in thanking them here, I do want to say in particular that it seems to me the published version is much better shaped than the first draft was, and that that improvement is the direct product of their sage advice.

Between them, Roger, Anthony, and Rob raise more issues than I can possibly address in the space available. Roger asks really nice questions about all four of my substantive chapters, on metaethics, moral epistemology, utilitarianism versus dogmatic intuitionism, and utilitarianism versus egoism. Rob rejects my central conclusions about utilitarianism versus egoism. Anthony raises very important issues particularly in moral epistemology.

I will have to be somewhat selective in my responses. But I will try to say at least something about all the Chapters on which my commentators remark, while devoting most of the space to the two Chapters to which Roger, Anthony, and Rob devote the most critical attention: the moral epistemology chapter and the utilitarianism versus egoism chapter.

(1): Metaethics:

Begin with Roger's first question, about whether Sidgwick is really committed to what I call (drawing on language of Michael Smith's) the "realist conceptual thesis".[24] Roger says "a presupposition does not have to be conceptual…what [Sidgwick] is rejecting is not a particular understanding of the concept of morality, but a conception of it."

[24] Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994), esp. pp. 63-66.

I suppose that the crucial passages for adjudicating this interpretive issue are those in which Sidgwick is rejecting a form of moral sense theory, the "suggestion" (as he puts it) that

> The judgments or propositions which we commonly call moral…really affirm no more than the existence of a specific emotion in the mind of the person who utters them. (ME 26)[25]

The question is whether Sidgwick takes this suggestion to express an alternative *conception* of the moral, or rather whether he takes the suggestion to fail to capture the *concept* of the moral at all. I am still inclined to read him in the latter way, inspired by passages like the following:

> The peculiar emotion of moral approbation is, in my experience, inseparably bound up with the conviction, implicit or explicit, that the conduct approved is 'really' right – *i.e.* that it cannot, without error, be disapproved by any other mind. If I give up this conviction…I may no doubt still retain…a sentiment of repugnance to the opposite conduct: but this sentiment will no longer have the special quality of 'moral sentiment' strictly so called. (ME 27)

There is still the further question whether Sidgwick regards the claim that there is such a thing as moral truth and error as a presupposition or as "incorporated in the basic, conventional, meanings of moral terms" (to use Mackie's language).[26] But I don't think Sidgwick regards the claim as *merely* part of something as optional as (one) *conception* of the moral.

A second, very intriguing, question Roger raises is about whether Sidgwick's argument for non-naturalism is really any different from Moore's.[27] Alternately put, it's the question: what is the open question argument? In writing about this in my metaethics Chapter I adopted, probably too uncritically, a contrast Gibbard draws in *Wise Choices, Apt Feelings* between, on the one hand a "master argument" against all naturalistic definitions and, on the other hand,

---

[25] Henry Sidgwick, *The Methods of Ethics*, 7th edition (Indianapolis: Hackett, 1981; original publication: 1st edition 1874, 7th edition 1907). I will refer to it as "ME"; page references will be placed in the text.
[26] J.L. Mackie, *Ethics* (Harmondsworth: Penguin, 1977), p. 35.
[27] G.E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), esp. Chapter 1.

(piecemeal) "attacks on specific examples of naturalistic analysis".[28] Is there really any difference here?

I still think that there is; but I do think the way I put it made the difference look too stark. The difference would be a matter partly of how much argumentative work is done by the initial intuition that the factual is one thing, the normative another, and partly a matter of just how the arguments against specific proposed naturalistic analyses are supposed to go. On the master argument view, the initial intuition would be close to decisive. By contrast, on the piecemeal arguments view, the initial intuition would establish a presumption, but would be much less close to decisive. And the argumentative resources used to supplement the initial intuition would also be different. On the master argument picture, (to do, necessarily roughly, something that takes a good deal of time and care to do precisely) for any proposed name of a natural property, "N" we have to just ask "Is A, which is N, good?" to see that N is not the same as goodness. By contrast, on the piecemeal argument picture, more work is required. We cannot simply appeal to a single, generic, all-purpose thought experiment. We must instead bring out specific, and different, problems with specific different naturalistic proposals. One example of such a specific objection to a specific naturalistic proposal would be the objection Sidgwick makes to moral sense theories a little earlier in the characteristically succinct and nuanced discussion I quoted from above: that at least their simple versions get it wrong about when people disagree:

> It is absurd to say that a mere statement of my approbation of truth-speaking is properly given in the proposition 'Truth ought to be spoken'; otherwise the fact of another man's disapprobation might equally well be expressed by saying 'Truth ought not to be spoken'; and thus we should have two coexistent facts stated in two mutually contradictory propositions. (ME 27)

(2) Moral Epistemology

[28] Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990), esp. p.11.

One central claim in my moral epistemology chapter is that there is a puzzle in Sidgwick's moral epistemology. The categories necessary to frame the puzzle come from Sidgwick's short 1879 paper, "The Establishment of Ethical First Principles".[29] He there distinguishes two general kinds of solution to the problem faced by the proponent of one putative ethical first principle who wants to give others "rational inducement" to accept this first principle. In the first solution:

> I may begin by regarding some limited and qualified statement as self-evident, without seeing the truth of the simpler and wider proposition of which the former affirms a part; and yet, when I have been led to accept the latter, I may reasonably regard this as the real first principle, and not the former, of which the limitations and qualifications may then appear accidental and arbitrary. (EE1P 106)

I call arguments of this first kind "*bipartite ad hominem arguments*," or "*bipartite arguments*" for short.

Sidgwick then goes on to characterize a second, "quite different", way out of the moralist's dilemma:

> We may be able to establish some general criteria for distinguishing true first principles (whether ethical or non-ethical) from false ones; and may then construct a strictly logical deduction by which, applying their general criteria to the special case of ethics, we establish the true first principles of this latter subject. (EE1P 107)

I call arguments of this second kind "*criterial arguments*."

With these concepts and labels available, we can articulate the puzzle. It is raised by the conjunction of three claims: (a) In Book III of the *Methods*, Sidgwick develops a criterial argument for utilitarianism; (b) In general Sidgwick thinks of criterial arguments as different from and more powerful than bipartite arguments; (c) In Book IV, in discussing the proof of utilitarianism, Sidgwick considers only bipartite arguments, not independent criterial arguments.

---

[29] Henry Sidgwick, "The Establishment of Ethical First Principles," *Mind* Vol. 4, No. 13 (1879), pp. 106-111. I will refer to it as "EE1P"; page references will be placed in the text.

The puzzle then is: why? Why does the independent criterial argument for utilitarianism disappear in Book IV?

I go on to distinguish three possible solutions to the puzzle: (i) Sidgwick ignores the independent criterial argument for utilitarianism in Book IV because in Book IV, as against Book III, he recognizes that the independent criterial argument for utilitarianism doesn't work; (ii) He ignores the independent criterial argument in Book IV because features of the rhetorical context of IV II, absent in Book III, make only bipartite arguments seem relevant – what are required in the context of Book IV are arguments addressed to specific opponents; (iii) He ignores the independent criterial argument in Book IV because he takes special features of common-sense morality to make the bipartite argument against the dogmatic intuitionist much more compelling than bipartite arguments are in general. I argue tentatively for solution (i).

One thing I am concerned to do is to argue that there *is* a puzzle. To that extent, if my commentators agree that there is a puzzle, but defend a particular kind of solution to it, that is fine with me. But I do also argue tentatively in favor of a particular solution to the puzzle, namely solution (i). Roger argues instead in favor of solution (ii). He says "As I understand the context of IV II, Sidgwick does indeed see the criterial argument as having failed – but only rhetorically, not as an argument in itself."

I am less happy with this solution than Roger is. I am not persuaded that the rhetorical context of IV II makes criterial arguments irrelevant. Let me try to make, or to elaborate on, that point in two ways. First, consider again "The Establishment of Ethical First Principles". I don't see the context set there as different from the context set by *Methods* IV II. In that context Sidgwick treats independent criterial arguments as relevant and available. But in the context of *Methods* IV II he ignores them. Second, something like a criterial argument seems to be *part* of

the complex bipartite argument characterized in IV II: after all, what is supposed to be on offer to the interlocutor is not merely a line of argument which "allows the validity, to a certain extent, of the maxims already accepted"; it is also an argument which "shows them to be not absolutely valid, but needing to be controlled and completed by some more comprehensive principle". If a criterial argument is admissible as *part* of a bipartite argument in the context of IV II, how can an *independent* criterial argument not be admissible in that same context?

Rob defends an interesting and different solution to the puzzle. Since it depends on his distinctive view about Sidgwick and the argument for utilitarianism, I will follow his lead and comment briefly on it after discussing that argument.

A second epistemological matter on which Roger and Anthony both press me is the interpretation of the 4th of the criteria. Sidgwick does not (to my mind, anyway) give in the 7th edition of the *Methods* a short, illuminating formulation of this criterion (unlike the other three, for which the first sentence of his presentation in III XI provides such a formulation). Evidently most unwisely, I supply for him (inspired by EE1P) the following one sentence formulation "The proposition must be universally accepted". Both Roger and Anthony argue that the criterion is really a negative one. Nothing nearly as strong as universal acceptance is required. Instead, what there must not be is rationally intractable disagreement with epistemic peers.

I agree with them about the proper interpretation of the criterion, and regret my formulation of it. It is not that there is *no* textual support in the *Methods* for putting the criterion as I do. The first sentence of the passage introducing the criterion in the 1st edition of the *Methods* includes the following

> Any defect in the universal acceptance of a proposition must *pro tanto* impair our confidence in its validity. (ME1, 320-321)

But the phrasing is different by the 7th edition, and I agree that Roger and Anthony are right on balance to attribute the weaker version they articulate to Sidgwick. I am not yet persuaded, however, that anything I argue depends problematically on the stronger and mistaken interpretation. In arguing to the contrary, I take Anthony to be suggesting that meeting a negative condition cannot contribute to justification. So if the fourth criterion is read as negative rather than positive, the view that Sidgwick regards satisfaction of the conditions as contributing to justification is to that extent weakened.

A third epistemological issue, on which both Anthony and Roger press me, is the conception of criterial argument, and whether criterial arguments involve granting common-sense morality probative status. Anthony in particular suggests that I end up with the puzzling view that "Sidgwick grants probative status to common-sense morality in the context of his criterial argument but not in the context of the bipartite argument." I agree entirely that this would be puzzling, and it is certainly not what I want to say. Let me try to argue that I can avoid saying it.

First, on (the conception I label) "criterial argument": I take the general idea of criterial argument to have very broad scope. Consider again in this connection the passage from EE1P which introduces that idea:

> We may be able to establish some general criteria for distinguishing true first principles (*whether ethical or non-ethical*) from false ones; and may then construct a strictly logical deduction by which, *applying their general criteria to the special case of ethics*, we establish the true first principles of this latter subject. (EE1P 107; italics added)

In this passage (which, it is worth emphasizing, is from EE1P, not from the independent epistemological writings) Sidgwick clearly articulates a broad conception of criterial argument. The criteria invoked here give no special role to common-sense *morality* because they give no special role to morality at all. They can, after all, be applied to putatively self-evident non-moral

as well as moral claims; and when they are applied to non-moral claims there is no reason to expect moral convictions to play any special evidential role. Criterial argument is a general strategy; its application to *ethical* first principles is a special case. Thus the general conception of criterial argument gives no special probative role to common-sense morality.

Now, when criterial arguments *are* employed to try to establish the first principles of *ethics*, then appeal to the criteria – especially the fourth, discussed above – seems to give an epistemic role to *common-sense convictions about ethics*. For suppose epistemic peers have convictions about ethics which conflict with some ethical claim we initially take to be self-evident. Then, the 4th criterion, read as Anthony and Roger suggest, means that the self-evidence or justification of that claim is reduced or eliminated. But I persist in wanting to distinguish the set of epistemically important common-sense convictions about ethics in general from the specific subset of such convictions which are associated with one particular method – dogmatic intuitionism. Anthony urges that this distinction I want to make – between common-sense morality in the narrow and common-sense morality in the broad sense – has insufficient grounding in the text of the *Methods*. I persist in thinking, to the contrary, that it is adequately grounded in the very first chapter of the book, and that it is a distinction you really have to make once you note that Sidgwick thinks that each of his three methods is "implicit in our common moral reasoning" (ME 14). Putting together these two points– that criterial argument is a general strategy of argument whose application to *ethical* first principles is only one special case, and that, even when it *is* applied to ethics, the common-sense moral convictions that get a probative role are those of common-sense morality only in the broad sense – I still think I can reasonably deny that Sidgwick's conception of criterial argument gives any special evidential role to common-sense morality in the narrow sense.

Second, on bipartite argument: My view here is that *the general conception* of bipartite argument no more gives a special role to common-sense morality in the narrow sense than does *the general conception* of criterial argument. Of course a bipartite argument addressed to the dogmatic intuitionist (the proponent of common-sense morality in the narrow sense) will give a special role to the convictions of the common-sense moralist. But in just the same way a bipartite argument addressed to the egoist will give a special role to the convictions of the egoist.

So I do not think that either the general conception of criterial argument or the general conception of bipartite argument gives any special epistemic role to common-sense morality in the narrow sense. As I see it, if there is a remaining problem here the problem is not that I am committed to a puzzling view about the different roles of common-sense morality in the narrow sense in criterial as against bipartite arguments. It is rather a problem that Anthony and I both face – that, as Roger urges, there are passages in the *Methods* which seem to give an epistemic role to common-sense morality in the narrow sense over and above anything either Anthony or I wants to allow. I claim that the idea that Sidgwick gives common-sense morality an evidential role is "largely mistaken," and I have basically three tactics for (largely) explaining away these problematic passages: (a) Some such passages (including a number cited by Brink)[30] show only that common-sense morality has an evidential role in the context of bipartite arguments addressed to the common-sense moralist. But that does not show that common-sense morality is *independently* evidential; (b) Remaining passages may reflect Sidgwick's gradual and incomplete separation of the epistemic from the moral-theoretic components of intuitionism, about which C.D. Broad complained when he introduced the term "deontology" in its standard

[30] David Brink, "Common Sense and First Principles in Sidgwick's *Methods*," *Social Philosophy and Policy* Vol. 11, No. 1 (1994), pp. 179-201.

contemporary sense;[31] (c) As above, we need to distinguish a narrow from a broad sense of "common-sense morality". Common-sense morality in the narrow sense is just that part of ordinary moral thinking systematized as dogmatic intuitionism. Common-sense morality in the broad sense includes also those parts of ordinary moral thinking systematized as utilitarianism and as egoism. The fourth criterion gives an epistemic role to common-sense morality only in the broad sense.

Roger is not convinced in particular by my second tactic, and suggests that, given Sidgwick's formidable intellect and multiple careful revisions of the *Methods*, it is "more plausible that [Sidgwick's] epistemology includes both elements of foundationalism and a commitment to some kind of Aristotelian dialectic". By contrast Anthony (I take it) is committed to the claim that Sidgwick does not assign an evidential role to common-sense morality, but, as I noted above, is skeptical in particular of my third tactic of distinguishing narrow and broad senses of "common sense morality".

To be really honest, I am not sure I am entirely convinced myself; I do think some commentators, including Brink, have exaggerated the textual evidence for the claim that Sidgwick takes common-sense morality to be independently evidential and that one has, for reasons including those my tactics bring out, to be quite careful with this textual evidence; and I think one could argue that Sidgwick *ought* not to assign an evidential role to common-sense morality. But I agree that it is tough to make a fully convincing textual case (a case which, again, I take it Anthony is committed to making) that Sidgwick *does not* assign *any* privileged evidential role to common-sense morality. In saying that Brink's idea that Sidgwick assigns an evidential role to common-sense morality is only "largely" (as opposed to "wholly") mistaken," I

[31] C.D. Broad, *Five Types of Ethical Theory* (London: Kegan Paul, 1930), p. 206.

take myself to have a bit more wiggle room than Anthony does. But, again, I take both of us to have plenty of work to do to be fully convincing on this issue.

(3) Utilitarianism versus Dogmatic Intuitionism

I want to offer just two observations about the issues covered in this Chapter. First, there is the question: how many self-evident principles does Sidgwick think there are? Is it just the (probably) four articulated in III.XIII? Is there also some further self-evident principle connected to egoism articulated outside III.XIII? And is Hedonism, defended by Sidgwick in III.XIV, taken by him to be self-evident in the same way as the principles he articulates in III XIII.? Rob thinks the only self-evident principles are those in III.XIII. I think there is an additional apparently self-evident principle connected to egoism articulated outside III.XIII. Roger urges here, and Anthony has argued elsewhere,[32] that Sidgwick thinks hedonism too is self-evident. I (briefly) argued on the other side in the book, but for the record I am increasingly persuaded that Roger and Anthony may be right about this. And certainly if, as I do, one thinks that there is a self-evident principle connected to egoism articulated outside III.XIII, one cannot go on to object to Roger and Anthony's view about Hedonism on the grounds that all the self-evident principles to which Sidgwick is committed are articulated in III.XIII.

Second, I think Roger is right that Sidgwick fails to recognize the important role played by the capacity for practical judgment in any plausible moral theory. Ross's view features both the key concept of *prima facie* duty and a key role for practical judgment in balancing competing *prima facie* duties in specific cases. I emphasized the former; but I agree that I could profitably have said considerably more about the latter.

(4) Utilitarianism versus Egoism

[32] Anthony Skelton, "Sidgwick's Philosophical Intuitions," *Ethics and Politics* Vol. 10, No. 2 (2008), pp. 185-209.

Rob wants to reject my overall verdict on the relative success of Sidgwick's central normative claims, about utilitarianism versus dogmatic intuitionism and about utilitarianism versus egoism. I give two versions of the verdict. One is that Sidgwick's treatment of the conflict between intuitionism and utilitarianism is less successful than he supposes, and his treatment of the conflict between utilitarianism and egoism is more successful. Rob, I take it, agrees with the first half of that version: he agrees that Sidgwick's critique of dogmatic intuitionism is subject to an unfairness objection, and hence is less successful than Sidgwick thinks it is. The other version of my verdict focusses specifically on the arguments for utilitarianism and egoism respectively discussed in my final chapter. Here, I claim that the argument for utilitarianism is a failure while the argument for egoism is more successful. Rob disagrees with both halves of that version: he thinks there is more to the axiomatic argument for utilitarianism than I allow, and nothing important and correct in the argument for egoism. Let me focus on those two disagreements in turn.

The Axiomatic Argument for Utilitarianism:

I claim that the axiomatic argument for utilitarianism is a failure. In summarizing my take on the argument, let me distinguish provisionally between my overall picture of the argument's character and success, and my detailed verdict on just which claims in the argument are non-tautological self-evident principles:

In my view the argument is supposed to invoke principles implied in the notion of "universal good". This notion is supposed to yield two self-evident principles:

(U) The good of any one individual is of no more importance from the point of view…of the universe than the good of any other.

And

(R) As a rational being I am bound to aim at good generally…not merely at a particular part of it.

(U) and (R) together are supposed to entail

(B) Each one is morally bound to regard the good of any one individual as much as his own.

I take (B) to express a form of impartial consequentialism, such that by adding hedonism to it we get a form of utilitarianism.

I take there to be two central problems with the argument, one of which Sidgwick acknowledges, the other of which he does not. The problem Sidgwick acknowledges is that the argument begins with the supposition that there is such a thing as universal goodness or goodness period: that there are (to use modern terminology) agent-neutral reasons. As Sidgwick famously says, egoists may deny that there are any such reasons, or, alternately put, that there is any such thing as universal goodness. The problem that Sidgwick does not acknowledge is that the existence of agent-neutral reasons does not entail that that is all the reasons there are. There may be both agent-neutral and agent-relative reasons.

Rob dislikes, or anyway finds odd, the way I put the first of these problems. He says

> Surely there *is* something I ought to desire, assuming myself to have an equal concern for all existence…It seems preferable to say that the egoist is uninterested in this thing – or, as Sidgwick says, refuses to take up the point of view of the universe.

I still want to defend my way of putting it, and, for my part, I find something missing in Rob's suggested alternative phrasing. What is missing is the *authority* of genuine oughts or reasons. If there is such a thing as genuine goodness, if there are agent-neutral reasons, the egoist cannot opt out of them because he is uninterested in them. They bind even those who are uninterested in them. That is why it is a crucial issue whether there is any such thing as genuine goodness, and an issue on which the clear-sighted egoist has to take a stand.

My central claim is that Sidgwick cannot properly get from his starting point to his conclusion, in crucial part because it cannot be established conceptually that there are authoritative or genuine agent-neutral reasons This central claim leaves it open just which, if either, of (U) and (R) are really non-tautological self-evident principles. Let me try to lay out as clearly as I can my view about the status of (U) and (R): does Sidgwick think they are tautological? Should he? Is he right?

As I see it *both* (U) and (R) *look* problematically tautological. But, given the setup in III XIII, Sidgwick is committed to thinking that at least one of them is non-tautological. Given the choice, put this way, I argue it is (U) that Sidgwick should see as non-tautological. That is because, I claim, Sidgwick's views about the concept of goodness make (R) come out tautological: because, for him, to think something good is simply to think that there is a reason to aim at it. But I think also that Sidgwick would still be wrong: that, properly conceived, both (U) and (R) are tautological. To argue for impartial consequentialism you need to show that there are genuine agent-neutral reasons, and that these are the only genuine reasons that there are. I think that in deriving (U), Sidgwick equivocates between a self-evident but tautological claim (roughly: if there is genuine goodness, each person has *some reason* to promote the good) and a non-tautological but also non-self-evident claim (roughly: if there is genuine goodness, *the only* reason each person has is to promote the good). So the axiomatic argument for utilitarianism is a failure. Of the two things that Sidgwick needs to establish (that there are genuine agent-neutral reasons, and that these are the only reasons there are), he recognizes in the end that he does not establish the first, and takes himself to establish the second only by sometimes illicitly construing (U) in a way that makes it not self-evident.

Rob has his own very interesting take on the axiomatic argument for utilitarianism.[33] I didn't have the opportunity to consider that take in the book. If I had to try quickly to compare his take to mine, I would say that his key moves are (a) to make the conclusion, (B) come out as a weaker claim than I make it; (b) to do some very ingenious fancy footwork to make (U) and (R) come out non-tautological and distinct, and (c) to deny that the 1st edition material characterizing the argument as "only [evolving] the suppression of Egoism…implied in the mere form of the objective judgment that 'an end is good'" (ME 1, 364) accurately characterizes the 7th edition version of the argument.

Once Rob's interpretation is available, should we go with his or mine? On balance I still go for my reading; but I see the attractions of his. It is a positive feature of his reading that it makes (U) and (R) come out non-tautological and distinct. It is less crucial to my reading, I think, just which of (U) and (R) is, or is thought by Sidgwick to be, non-tautological. You would get something very like my overall picture of the argument's character and success if you said that Sidgwick ought to think it is (R) that is non-tautological, rather than (U); and even if (like Broad) you thought (R) really was non-tautological. Rob's reading strikes me as more charitable, though I still think mine emerges more straightforwardly from the text. If (as I think and Rob does not) it is a key issue where and how Sidgwick takes himself to have established that there is such a thing as genuine goodness, it is an advantage of my reading that it gives prominence to this issue. On the other hand, as Rob points out in the last paragraph of his comments, it is an advantage of his reading that it helps provide an attractive solution to the puzzle in Sidgwick's moral epistemology. Specifically, as Rob sees it, the arguments of Book III and Book IV do different and important work. The argument of Book III establishes the axioms; but more is

[33] "Sidgwick's Axioms and Consequentialism," *The Philosophical Review* (forthcoming).

needed to secure utilitarianism, and that more (he can then say) is done by the argument of Book IV.

Anthony, I take it, wants also to press the idea that the argument of III XIII is supposed to do less work than I think it is supposed to do. But Anthony wants to emphasize in effect not only the distinction between other-things-equal and all-things-considered versions of the axioms, but also the distinction between first principles and methods. The argument of Book III, especially III XIII is concerned with the utilitarian *first principle*; new and different work in Book IV, especially IV II, is needed to vindicate the utilitarian *method*.

This again is a quite intriguing idea, which of course Anthony can only sketch here. I have two initial reservations about it. First, it seems to me the language Sidgwick uses in IV II often suggests he is (still) there concerned with establishing first principles, not with establishing methods. Thus the chapter begins

> In book ii, where we discussed the method of Egoistic Hedonism, we did not take occasion to examine any proof of its first principle. (ME 418)

And the language of "principles" and "first principles" is prominent throughout the chapter (occurring again on every page of IV II, and not in sentences suggesting that the current concern is with establishing methods rather than principles). So, I suggest, it is hard to read IV II as concerned with establishing methods rather than with establishing principles.

Second, I have some more general reservations about the very idea of establishing a method, as distinct from establishing a first principle. I think I know what it would be to establish the first principle of utilitarianism: it would be (roughly) to establish that the only valid ultimate reasons there are are agent-neutral reasons to maximize pleasure and minimize pain. To determine what would be involved in establishing a method, I take it, one would need to decide what the relation was between methods and first principles, and to decide whether methods could

be established independently of first principles, or only via the establishment of associated first principles. The part of the *Methods* most relevant to these questions are, I take it, in I VI. The following two passages seem to me particularly important:

> What then do we commonly regard as valid ultimate reasons for acting or for abstaining? This, as was said, is the starting-point for the discussions of the present treatise: which is not primarily concerned with proving or disproving the validity of any such reasons, but rather with the critical exposition of the different 'methods' – or rational procedures for determining right conduct in any particular case – which are logically connected with the different ultimate reasons widely accepted. (ME 78)

> Not all the different views that are taken of the ultimate reason for doing what is concluded to be right lead to practically different methods of arriving at this conclusion. Indeed we find that almost any method may be connected with almost any ultimate reason by means of some – often plausible – assumption. Hence arises difficulty in the classification and comparison of ethical systems; since they often appear to have different affinities according as we consider Method or Ultimate Reason. (ME 83)

These passages suggest to me first that, tricky as it is to establish first principles (or "valid ultimate reasons for action"), establishing methods would be a whole lot trickier. So I am reluctant to portray IV II (or Book IV in general) as concerned with the establishment of methods rather than with the establishment of principles. And, second, these passages incline me to suggest, at least tentatively, that the only things Sidgwick is ever concerned with "proving" or "establishing" are first principles or claims about valid ultimate reasons; while Sidgwick really does devote much space to methods, none of that space is concerned with "proving" or "establishing" *methods* at all.

Egoism

I think Rob and I disagree more fundamentally about the egoistic side of Sidgwick than we do about the utilitarian side. On the utilitarian side we largely agree about which arguments are successful; our disagreements are about which to attribute to Sidgwick. By contrast on the egoistic side we disagree pretty fundamentally about whether an argument for something related

to egoism is any good; and that philosophical difference is one key driver of our interpretive disagreements.

Let me try informally to characterize what I take to be the key insight in that argument. The insight is that the separateness of persons means that each of us has special reason to care about good and bad things that happen to him or her, reasons that we do not have to care about good and bad things that happen to other people. The fact that I directly experience pains and pleasures I have gives me reason to care about my pains and pleasures that others do not have. The way of putting this key insight I draw from Sidgwick is this:

> (2) If the distinction between any one individual and any other is real and fundamental, then I ought to be concerned with the quality of my existence…in a sense, fundamentally important, in which I ought not to be concerned with the quality of the existence of other individuals.

There are two important qualifications here: (a) that I have special reason to care about good and bad things that happen to me does not mean I have no reason to care about good and bad things that happen to others. So the insight does not really support egoism ; (b) Rob is right that the insight is most compelling when what we think about are hedonic goods and bads; it is at least less compelling for non-hedonic goods and bads.

Rob discusses one of Roger's handy 2-doors cases, which I appeal to and run a variation or two on at the end of my utilitarianism and egoism chapter.[34] But I think the way Rob treats the case illustrates how deeply we disagree about this. He writes as if we would only come to contemplate accepting (2) to explain specific intuitions about some 2 doors case (or cases). I think rather that (2) is itself intuitively compelling prior to any application of it to a specific 2-doors case.

[34] Roger Crisp, *Reasons and The Good* (Oxford: Clarendon Press, 2006), pp. 131-145.

There is one part of Rob's discussion here I find puzzling. He objects that endorsing (2) as a normative premise "violates Sidgwick's rule that normative conclusions cannot be deduced from wholly non-normative premises. For (2) says that a wholly non-normative claim entails a normative claim." I think this cannot be right. Non-naturalists are typically committed to the possibility of true claims connecting the natural to the normative. As W.D. Ross observes, "[the non-naturalist Hedonist holds] that rightness is something indefinable, and merely [claims] that that which *makes* acts right is their tendency to promote pleasure. [He holds]…that a non-ethical characteristic, a psychological characteristic, is the ground of rightness but not its essence."[35] Whatever status the non-naturalist hedonist takes hedonism to have, (2) can have just as well. It cannot be that the rule that normative conclusions cannot be deduced from wholly non-normative premises rules out all premises connecting the non-normative with the normative. And I can see no good reason not to regard (2), like hedonism, as just one possible such premise.

Roger, Anthony, and Rob ask many more good questions, and raise many more fascinating issues and troubling objections, than I have been able to address here. Let me just close by thanking them all once again for their comments, and thanking Anthony in particular for organizing this symposium, and the ISUS conference session from which it derives.

David Phillips

University of Houston

[35] W.D. Ross, *The Foundations of Ethics* (Oxford: Clarendon Press, 1939), pp. 8-9. Compare Derek Parfit, *On What Matters* Vol. 2 (Oxford: Oxford University Press, 2011), pp. 303-305; see also ME 79.