Utilitas (2025), 1–18
doi:10.1017/S0953820825100113

CAMBRIDGE UNIVERSITY PRESS

ARTICLES

Sidgwick's Critique of Deontology: Scrupulous Fairness or Serpent-Windings?

Tyler Paytas

Australian Catholic University, North Sydney, Australia
Email: tyler.paytas@acu.edu.au

Abstract

David Phillips (2011) and Thomas Hurka (2014a, 2014b) argue that Sidgwick's critique of deontology contains three serious flaws. First, it has no force against moderate deontologies composed of prima facie duties rather than unconditional duties. Second, Sidgwick's preferred principles fail to meet the very criteria by which he rejects deontological principles. Third, Sidgwick's employment of his key maxim of Rational Benevolence equivocates between all-things-considered and other-things-equal formulations. I defend Sidgwick against all three criticisms. (1) While some of Sidgwick's arguments apply only to absolute deontology, others apply to moderate versions as well. (2) Although Sidgwick's preferred principles do not fare perfectly against his criteria, they still fare better than the deontological principles. (3) The suggestion that Sidgwick relies on an all-things-considered formulation of Rational Benevolence is based on a misunderstanding of the structure of his argument. The upshot is that Sidgwick's overarching line of argument is stronger than recent critics suggest.

Keywords: Sidgwick; consequentialism; deontology

1. Introduction

Sidgwick has long been admired for his impartial and objective approach to philosophical investigation. According to Brand Blanshard, Sidgwick's contemporaries viewed him as the paradigm of impartiality and fairmindedness: "He stood in their view as the exemplar of objectivity in thought, of clear and passionless understanding. The light he threw on his subject was uniquely uncolored by feeling, prejudice, or desire" (Sidgwick 2000: xiv). This view continued into the early part of the twentieth century, as C. D. Broad (1930: 309) called *The Methods of Ethics*[1] a "conspicuously honest" book, and John McTaggart (1906: 412) praised Sidgwick for his "scrupulous fairness to his opponents."

Despite this precedent of reverence, recent decades have brought a wave of challenges to Sidgwick's reputation as the gold standard of objectivity and fairness. Critics have

[1]I shall follow the convention of using the abbreviation *ME* for parenthetical references to *The Methods of Ethics* (Sidgwick 1981).

argued that, contrary to his own account of the development of the *Methods*, Sidgwick's investigation was colored by prejudice in favor of consequentialism (Donagan 1992, Brink 2003, Irwin 2009, Korsgaard 2009). An influential charge along these lines, made by David Phillips (2011) and Thomas Hurka (2014a, 2014b), is that Sidgwick applied a double standard in his arguments for the superiority of consequentialism over deontology. The allegation is that while Sidgwick rejected deontology because it fails to meet his criteria for justification, he did not test his preferred principles by the same standards. Further, while one of Sidgwick's main objections to deontological principles hinges on their being equivocal between other-things-equal and all-things-considered formulations, Phillips and Hurka both claim that Sidgwick's principle of Rational Benevolence, which serves as the rational basis for utilitarianism, equivocates in the same way. Phillips and Hurka also agree that Sidgwick's critique of deontology has no force against moderate versions that utilize the concept of prima facie duty. The upshot of these objections is that, contrary to what we might expect given his reputation, Sidgwick's arguments were not only unpersuasive but also unfair.[2]

This paper defends Sidgwick against all three charges. I begin by outlining the key elements of his critique of deontology and explaining why they are not avoided by moderate deontology. Next, I provide an interpretation of Rational Benevolence that allows the principle to do the work Sidgwick needs it to without any equivocation. Finally, I argue that the allegation of unfairness is based on the mistaken assumption that Sidgwick's criteria for justification are all-or-nothing. The upshot is that (1) the *Methods* is not beset by equivocation or unfairness, and (2) Sidgwick's overarching line of argument against deontology is stronger than recent critics maintain.

2. The Critique of Deontology

2.1. Sidgwick's Criteria for Trustworthiness

In order to understand Sidgwick's critique of deontology, we must first examine the epistemic criteria he sets out for proper justification in ethics. His preferred form of intuitionism, which he labels "philosophical," is predicated on the need for genuinely self-evident principles to serve as a foundation for an adequately justified ethical theory. For Sidgwick, "self-evident" does not mean obvious or easy to grasp—sometimes the self-evidence only becomes clear after careful reflection. The idea is, rather, that the proposition is non-inferential, seems plausible once the terms are understood, and does not appear in need of further justification.[3]

Of course, there are many instances in which a proposition that initially seems self-evident turns out to be false. Partly to mitigate this problem, Sidgwick posits four criteria

[2]While many commentators find Sidgwick's argument to have some force when applied to absolute deontology, Alan Donagan (1974, 1977, 1992) claimed that the absolute deontology advanced by Whewell (and further developed by Donagan himself) survives Sidgwick's critique. Although a comprehensive study of Sidgwick and deontology would require analysis of Donagan's arguments, here I limit my focus to the influential line of objection posed by Hurka and Phillips, who draw primarily on Ross's moderate deontology. For an overview of Donagan's response to Sidgwick, see Phillips (2022: 110-11). Since the critics I'm responding to are sympathetic to Sidgwick's metaethics, I shall also set aside important objections from constructivist deontologists such as Korsgaard (2009).

[3]One clear indicator that an ethical principle is not self-evident is that it posits arbitrary definitions or boundaries (*ME* 293, n. 1). For further discussion of Sidgwick's notion of self-evidence, see Crisp (2015: 107-109).

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

that a proposition must approximately realize if it is to possess the "highest degree of certainty attainable" (*ME* 338). Although propositions that fare well against these criteria are not thereby guaranteed to be true, they are substantially more trustworthy than those that do not. The four criteria are as follows:

(1) The terms of the proposition must be clear and precise.
(2) The self-evidence of the proposition must be ascertained by careful reflection.
(3) The proposition must not conflict with other self-evident propositions.
(4) The proposition must not be denied by an epistemic peer. (*ME* 338–41)[4]

When Sidgwick applies these criteria to the deontological principles of common-sense morality, he finds them lacking. One major problem is that while some of the principles seem to meet the consensus criterion, the universal assent lasts only as long as the principle is presented in vague terms. Once we attempt to render the principle clear and precise, the consensus breaks down and the air of self-evidence disappears.

This point is illustrated by the purported duty of promise-keeping. Although this duty is widely accepted, uncertainty arises when we consider particular cases. For instance, common sense seems undecided about whether a promise is binding when material circumstances have changed such that fulfilling the promise will either be harmful to the promisee or inflict a disproportionate sacrifice on the promiser (*ME* 306–7). There is also uncertainty about cases in which the promisee has died or is unable to be communicated with (*ME* 305). Similar uncertainty arises when keeping one's promise would require violating another duty (*ME* 305). The reality of such circumstances suggests that the duty to keep one's promises must be qualified, and there will be disagreement and uncertainty over where to draw the boundaries.

Over several chapters of Book III, Sidgwick points out similar problems for a range of common-sense duties. In each case, an initially plausible and widely accepted proposition becomes dubious under closer inspection.[5]

2.2. Does moderate deontology avoid Sidgwick's critique?

Some critics allege that Sidgwick's critique of deontology has force only against absolute versions composed of all-things-considered duties (Broad 1930: 217–23; Phillips 2011: ch. 4; Hurka 2014b; Shaver 2014). According to moderate deontologies, such as that of Ross (1930), duties are prima facie rather than absolute. While it is possible to be in a situation in which two prima facie duties require incompatible actions, this needn't be worrisome because one of the duties can simply outweigh the other. To take one of Ross's examples, your duty to keep a promise to meet a friend can be outweighed by your duty to help the victims of an accident whom you encounter along the way. However, the fact that you should feel regret about breaking the promise, along with the ostensible fact that you may owe your friend compensation, are taken as evidence that the duty was still binding despite being overridden by another duty (Ross 1930: 18, 28; Hurka 2014b: 135).

[4]Interpretation of the criteria is controversial. For discussion, see *inter alia* Crisp (2015: ch. 4), Phillips (2011: ch. 3), and Skelton (2008).

[5]Part of Sidgwick's argument that is less frequently discussed is the purported failure of the duties against the second criterion. Sidgwick finds that these duties call out for a deeper justification and that their initial air of self-evidence may be due to epistemically pernicious influences. I shall return to this point below.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

Hurka suggests that Sidgwick failed to consider moderate deontology because he lacked the concept of prima facie duty (2014b: 135). As evidence, he draws attention to Sidgwick's insistence that the "conflicts" and "collisions" of deontological principles require them to be restated to include exception-clauses. In discussing his third criterion (that propositions accepted as self-evident must be mutually consistent), Sidgwick remarks that "any collision between two intuitions is a proof that there is error in one or the other, or in both" (*ME* 341). However, with prima facie duties, the possibility of collisions neither suggests the need for an exception clause nor that one (or both) of the relevant intuitions must be in error. In the example of breaking a promise in order to help accident victims, one can consistently maintain that the agent has a prima facie duty to help the victims as well as a prima facie duty to keep the promise. There would be a problematic collision only if the agent were purportedly under an all-things-considered obligation to do both.

If Sidgwick's concern about collisions were exclusively about cases of all-things-considered obligations to perform incompatible actions, this would indeed be evidence that he lacked the concept of prima facie duty (since prima facie duties easily avoid the problem). It would also show that a central strand of his critique of deontology has no force against moderate versions. However, there is little reason to believe that Sidgwick's concern about conflicting duties was only about the all-things-considered variety. Consider that in summarizing the overall case for utilitarianism, Sidgwick notes that vindication requires, among other things, demonstration that the different common-sense duties "are liable to conflict with each other, and that we require some higher principle to decide the issue thus raised . . . " (*ME* 421). The need for a higher principle suggests that what must be decided is which of the two purported duties takes precedence in a given context. This point applies to prima facie duties no less than to absolute duties.[6]

One way of resolving conflicts within absolute deontology is to add exception clauses to the principles (e.g., "One ought to keep one's promises unless breaking the promise is the only way to help in an emergency"). However, the addition of exception clauses diminishes the clarity, self-evidence, and consensus. Although moderate deontologies avoid formal inconsistency without having to add exception clauses, prima facie duties still generate practical conflicts that must be resolved. A principle's being binding though outweighed is no less problematic than its needing to be restated to include exception clauses. In either case, we end up without clear guidance, and within the method of common-sense deontology there is no obvious principled means of resolving the problem. Sidgwick is aware that a deontologist can in principle avoid problematic conflicts by explaining how the duties are to be weighed against each other. The problem is that there is no clear, intuitive, and agreed upon explanation in the offing (other than appealing to utilitarian considerations).[7] Moreover, Sidgwick's critique of common-

[6]Crisp (2015: ix, 114) criticizes Sidgwick for not being open to the possibility that agents can reliably exercise judgement in individual cases, and hence that higher principles explaining how duties are to be weighed against each other in any and all cases are unnecessary. However, Sidgwick provides reasons for skepticism about the reliability of individual judgements. Judgements of conscience concerning individual circumstances are (1) especially vulnerable to non-rational influences such as emotions and desires, (2) inconsistent across time, and (3) liable to conflict with the judgements of apparent epistemic peers (*ME* 214, 339-40).

[7]A moderate deontology might hold that the various duties can be weighed precisely such that we don't need a higher principle to adjudicate conflicts. For example, the duty not to kill an innocent person might be exactly 10 times as strong as the duty to save one, which would make it wrong to sacrifice one innocent life to save two but permissible to sacrifice one to save 100. Sidgwick would object that any such weighting would be arbitrary and lack self-evidence.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

sense duties is not limited to the issue of conflicts—it also involves questions about which acts fall under the respective principles (even apart from conflict cases) such as in the example of a promise made to someone who has since become deceased. He also finds the lack of a deep unifying explanation of the purported duties problematic (*ME* 102; see Lazari-Radek and Singer 2014: 67). In light of these concerns, which apply to both moderate and absolute versions, Sidgwick concludes that the method of common-sense deontology is inadequate.

3. The Equivocation Charge

3.1. Rational Benevolence

The aforementioned problems with common-sense duties lead to Sidgwick's presentation of a set of ethical axioms that he believes stand a better chance of providing secure answers to our ethical inquiries. Two of the key axioms are said to yield a principle labeled 'Rational Benevolence.'[8] The first axiom states that "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." The second states 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." Taken together, these self-evident propositions yield the principle of Rational Benevolence: "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). Sidgwick claims that Rational Benevolence (henceforth RB) is required as a rational basis for utilitarianism (*ME* 387).

Hurka argues that Sidgwick's understanding and employment of RB is equivocal (2014b, 150).[9] He suggests that the only way for RB to serve all of Sidgwick's needs is if it is given different readings in different contexts. In order to yield determinate guidance, RB must be given an all-things-considered reading on which the principle essentially states that one ought always to maximize impartial welfare. However, the all-things-considered version seems much less intuitive, and it fares poorly against Sidgwick's non-dissensus criterion. In order to appear intuitive and widely accepted, RB must be given an other-things-equal reading. But read this way, the principle can neither ground utilitarianism nor give rise to the dualism of practical reason that occupies the final chapter of the *Methods*. This is because an other-things-equal imperative to promote the general good need not conflict with a principle of self-interest. The formulations of RB suggested by Hurka's analysis can be stated as follows:

> *RB All-Things-Considered (ATC):* One ought always to do whatever has the best consequences for all.
>
> *RB Other-Things-Equal (OTE):* One ought to do whatever has the best consequences for all whenever a competing consideration does not take precedence.

Given that both candidate formulations of RB are at odds with lengthy and important discussions located throughout the *Methods,* charity requires searching for an

[8]There is controversy over the precise number and content of Sidgwick's axioms. See Schneewind (1977: 290 n., 294-95); Skelton (2008); Irwin (2009: ch. 83); Phillips (2011 ch. 4); Shaver (2014); Lazari-Radek and Singer (2014: ch. 5); and Crisp (2015: 115-126).

[9]This issue is also raised by Phillips (2011: ch. 4).

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

interpretation that would allow RB to do the work Sidgwick needs it to without running into the difficulties that Hurka articulates. I believe there is a plausible interpretation that does just that. According to this reading, RB is not a distinctively consequentialist principle but rather a principle addressing the issue of whether there are basic (i.e., non-derivative) reasons of partiality with respect to promoting the good. The proposed formulation is as follows:

> *RB Universal Good (UG)*: With respect to the rational aim of promoting the good, one ought to aim at universal good (rather than just one's own).[10]

This formulation better matches Sidgwick's own presentations of the principle than do the ATC and OTE formulations mentioned above. Recall that Sidgwick initially phrases the principle as follows: "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). On the very next page, Sidgwick characterizes the principle as "I ought not to prefer my own lesser good to the greater good of another" (*ME* 383). And in the concluding chapter of the *Methods,* Sidgwick describes RB as a principle "that sets before each man the happiness of all others as an object of pursuit no less worthy than his own . . . " (*ME* 496). Each of these formulations emphasizes the impartialism in promotion of the good expressed by UG.

An additional point in favor of UG is that it reveals a structural parallel between the derivations of Prudence and RB. The principle of Prudence essentially states that with respect to the rational aim of promoting one's own good, one should aim at one's good on the whole rather than prioritizing the present or near future. (*ME* 381).[11] The truth of this principle is seen by considering the fact that no temporal part of one's life is more intrinsically important than any other, combined with the fact that when it comes to the rational aim of promoting the good, one should treat equally valuable portions with equal regard (i.e., rational beings are "bound to aim at good generally" (*ME* 382)). Similar reasoning leads one to recognize the truth of RB (UG). Just as no temporal part of one's life is more valuable than any other, no individual's good is more intrinsically valuable than anyone else's (*ME* 382). Since the rational end of promoting the good requires treating equally valuable portions with equal regard, the fact that everyone's good is of equal inherent value means that one's efforts to promote the good ought to be directed at the total good of everyone rather than just one's own.

Unlike the ATC formulation, UG explains how Sidgwick could coherently claim agreement from the moralists he refers to, including Kant.[12] Although there is disagreement among scholars over how much latitude (if any) is allowed for prioritizing personal projects and private interests within Kant's ethics, one can plausibly read Kant as holding that one's efforts to promote happiness ought to be directed at mankind in

[10]Note that this principle does not suggest that promotion of the good is the sole rational aim. It thus leaves open that there are other rational aims, including adherence to deontological principles. In this respect, UG is similar to the OTE formulation described above. As I shall explain, the key difference is that UG rules out basic reasons of partiality.

[11]Sidgwick's exposition of Prudence does not include a single canonical formulation. However, the discussion strongly suggests the gloss I've given here.

[12]The *Methods* does not provide a reconstruction of Kant's ethics that is robust enough to substantiate a claim of general ethical agreement between Kant and Sidgwick. However, Sidgwick's claim of agreement with Kant in Book III is restricted to his principle of Rational Benevolence. This claim is rendered more plausible on my proposed interpretation.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

general without giving priority to one's own.[13] Relatedly, UG also makes better sense of Sidgwick's claim that RB is intuitive and widely accepted. Though many people vehemently reject the claim that one ought always to maximize the good by any means (as on the ATC reading), it would not be entirely unreasonable for Sidgwick to believe that most people would endorse UG (at least once they reflect on his arguments in Book III).

UG is similar to the OTE formulation in that both are compatible with deontological constraints. However, UG is still a stronger principle in that it overtly rules out basic reasons of partiality in promoting the good. Unlike the broader OTE formulation, UG provides "a rational basis for utilitarianism" (*ME* 387) by setting universal good, rather than one's own good, as an ultimately reasonable end. UG is not identical with utilitarianism because it does not preclude duties that might serve as constraints on promotion of the good. But Sidgwick acknowledges this fact when he claims that to show that RB is "*sole* or *supreme*" it must be demonstrated that ostensible principles that might compete with it are ultimately subordinate to it (*ME* 421).[14]

3.2. Objections

One might object to my interpretation by pointing to passages in which Sidgwick seems to identify RB with utilitarianism, which would require an ATC reading of the principle. Indeed, in the passage just referenced, Sidgwick refers to RB as "the Utilitarian first principle" (*ME* 421). Another example occurs shortly after his presentation of the axioms, where he writes: "I arrive, in my search for really clear and certain ethical intuitions, at the fundamental principle of Utilitarianism" (*ME* 387).[15]

On closer examination, these texts do not show that Sidgwick views RB as a statement of an ATC consequentialist principle. Just before the passage at *ME* 387, Sidgwick says that Rational Benevolence is "required as *a rational basis* for the Utilitarian system" (emphasis added). A clue as to what he means is provided in the subsequent passage in which he discusses a gap in Mill's argument for utilitarianism that can only be filled by a principle such as RB. The purported gap concerns the move from an aggregate of desires

[13]For discussion of the scope and stringency of Kantian duties of benevolence, see Timmerman (2005) and Van Ackeren and Sticker (2015).

[14]In this passage, Sidgwick says that the reasoning that led to RB in III.XIII "as addressed to the Intuitionist" only shows the principle "to be *one* moral axiom" and that other principles must be ruled out in order to incline the intuitionist/deontologist toward utilitarianism (*ME* 421). The phrase "as addressed to the Intuitionist" is a source of controversy. Some scholars (Singer 1974; Hurka 2014a) hold that Sidgwick views the derivation of RB (*ME* 382) as sufficient for ruling out deontological constraints and hence that the discussion at *ME* 419-422 is an ad hominem argument addressed to a deontologist who confusedly fails to see the full force of Sidgwick's derivation of RB as the supreme moral principle. This interpretation suggests a stronger reading of RB that is closer to ATC than UG. A key motivation for this interpretation is that it explains why Sidgwick uses the phrase "as addressed to the Intuitionist" rather than making a more general statement about the need to add a discursive argument against other candidate principles. However, there is an equally viable explanation that is compatible with my reading of RB. Sidgwick is aware that some readers will not find deontological constraints intuitive to begin with (e.g., consequentialists—whether impartialist or partialist). For such readers, the derivation of RB (on my UG interpretation) would likely be enough to convince them to accept that *impartial* pursuit of the good is the supreme moral maxim. However, readers who antecedently find deontological constraints intuitive will still need convincing because RB itself (on the UG reading) does not rule out deontological constraints. That is why he uses the phrase "as addressed to the Intuitionist."

[15]I thank Tom Hurka for urging me to address these passages.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

for individual happiness to a desire for general happiness. In other words, what is needed is a premise that one ought rationally to desire universal good rather than just one's own good. This premise is supplied by the impartialist principle of UG. That is the sense in which RB is required as a rational basis for (and the fundamental principle of) utilitarianism—it establishes *impartial* pursuit of the good as a rational imperative. But that doesn't mean it rules out deontological constraints—doing so requires further argument.

Even if UG is indeed the correct interpretation of RB, challenges to Sidgwick's use of the principle in the *Methods* remain. Although UG is more widely accepted than an all-things-considered consequentialist principle, there are still some who believe that one's own good is the supreme rational aim. And setting egoism aside, common-sense seems to suggest that there are basic reasons of partiality. Hence, it is not clear how the principle can meet Sidgwick's fourth criterion for trustworthiness (i.e., the non-dissensus criterion).

Although several critics have noted these problems, they are not as significant as initial appearances suggest. Sidgwick addresses the worry about the conflict with common sense. After acknowledging that the common-sense duty of benevolence is less stringent than RB, he provides an explanation for the discrepancy: "But I think it may be fairly urged in explanation of this that *practically* each man, even with a view to universal Good, ought chiefly to concern himself with promoting the good of a limited number of human beings, and that generally in proportion to the closeness of their connexion with him" (*ME* 382; see also 432–34). When Sidgwick introduces his non-dissensus criterion, he includes the qualifier, "if I have no more reason to suspect error in the other mind than in my own" (*ME* 342). The disagreement with the judgment of common-sense concerning RB does not reduce Sidgwick to a state of neutrality because he sees strong reasons for believing that common-sense intuitions about this issue are distorted by the practicalities of human life (not to mention non-rational impulses such as familial affection). Moreover, Sidgwick believes that those who would not initially assent to RB could be persuaded by his axiomatic deduction and his various appeals to the need for system and synthesis that would be hard to square with basic reasons to promote the good of some over the equal good of others.[16]

As for the disagreement with egoists, it is surprising that Sidgwick does not mention this worry in the main text of the chapter in which he presents his axiomatic deduction of RB. He does include a footnote that reads, "On the relation of Rational Egoism to Rational Benevolence—which I regard as the profoundest problem in Ethics—my final view is given in the last chapter of this treatise" (*ME* 386 n. 4). The lack of discussion of this issue in Book III may be explained by Sidgwick's devoting an entire chapter to the dualism of practical reason (which ultimately leads to a reduction of confidence in RB). At this stage, what he takes himself to have established is that the search for clear and genuinely self-evident intuitions (i.e., intuitions that meet the first and second criteria) leads away from deontological duties and towards RB, which doesn't call out for deeper explanation as do maxims like "one ought to keep one's promises" (*ME* 382–83).

3.3. Self-Sacrifice and the Dualism of Practical Reason

Before turning from the equivocation charge to the broader unfairness objection, it will help to consider Sidgwick's explanation for the loss of confidence in RB in the

[16]See Shaver (2020). The strategy of debunking partialist intuitions in order to vindicate Sidgwick's Rational Benevolence is utilized by Lazari-Radek and Singer (2014: ch. 7).

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

concluding chapter of the *Methods*. While the dualism is often understood to be a conflict between egoism and utilitarianism, it is doubtful that full-fledged egoism is what troubled Sidgwick.[17] Full-fledged egoism holds that self-interest is the *only* source of non-derivative practical reasons. This implies that the preventable agony of a billion people is not in itself a source of reasons for an agent who is in a position to help. While this extreme view is deeply implausible, Sidgwick does find a different principle of self-interest intuitive. This principle states that it is always irrational to sacrifice one's own happiness (we can label this 'No Self-Sacrifice,' abbreviated NSS). NSS acknowledges that the good of others can be a source of ultimate reasons, but it forbids promoting that good when doing so comes at a net cost to oneself.[18]

Evidence for this reading comes from several passages where Sidgwick describes the relevant intuition in terms of the irrationality of self-sacrifice. Consider the following passage from the concluding chapter of the *Methods*:

> And further, even if a man admits the self-evidence of the principle of Rational Benevolence, he may still hold that his own happiness is an end which *it is irrational for him to sacrifice to any other;* and that therefore a harmony between the maxim of Prudence and the maxim of Rational Benevolence must be somehow demonstrated, if morality is to be made completely rational. This latter view, indeed (as I have before said), appears to me, on the whole, the view of Common Sense: and it is that *which I myself hold. (ME* 498, emphasis added*)*[19]

Here Sidgwick explicitly describes his own view as admitting the self-evidence of Rational Benevolence while simultaneously holding that self-sacrifice is irrational. This idea is echoed in "Some Fundamental Ethical Controversies," where Sidgwick characterizes the relevant intuition as the idea that "it would be irrational to sacrifice any portion of my own happiness unless the sacrifice is to be somehow at some time compensated by an equivalent addition to my own happiness" (1889: 483; see also, *ME* xviii, 109 n.1, 418).

A key benefit of reading the self-interested half of the dualism as a prohibition against self-sacrifice rather than full-fledged egoism is that it explains why Sidgwick believed

[17]For interpretations of the dualism, see *inter alia* Parfit (2011: 130-4), Crisp (2015: 227-234), Phillips (2022: ch. 11), Skelton (2025).

[18]This is similar to the principle that Shaver (1999: 2) calls "veto egoism," which says that contributing to the agent's well-being is a necessary (but not sufficient) condition for rational action. I've chosen to formulate the principle as a rational prohibition on self-sacrifice because Sidgwick frequently puts it in these terms. For other veto egoism-like formulations, see McLeod (2000: 284), Phillips (2011: 134, 136, 140), Crisp (2015: 228–230).

[19]This passage could be read as telling against my UG interpretation of RB. Here Sidgwick seems to equate the question of how egoism and utilitarianism are related with the question of how RB and Prudence are related. One might think that this could only make sense if RB makes as strong a claim as utilitarianism. But at this stage, the doubts about utilitarianism do not concern its rejection of deontological constraints but rather its demand for impartial promotion of the good. Hence, it makes sense for him to focus on RB (read as UG) in discussing the conflict between utilitarianism and egoism. That said, I do think Sidgwick is being careless when mentioning the maxim of Prudence as conflicting with RB. It's hard to read the key discussion of Prudence at *ME* 381 as advocating a principle that conflicts with RB. Not only do none of the formulations in that discussion suggest such a principle, it would also have been a major blunder for Sidgwick to present two conflicting principles on subsequent pages and conclude that these are among the most secure ethical intuitions. Given his gloss at the start of the *ME* 498 passage, it seems likely that when he says 'Prudence' here what he has in mind is NSS.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

that God could resolve the problem (*ME* 503–509). Sidgwick's basic thought seems to be that the apparent conflicts between promoting one's own happiness and promoting universal happiness would be resolved by God's existence because divine sanctions would ensure that it is always prudent (in the long run) to promote universal happiness. Yet, as noted by Moore (1903: 103) and Broad (1930: 158–59, 244–45, 253), the conflict between Sidgwick's principles (as traditionally interpreted) appears to be not merely practical but also theoretical. If the principle of self-interest that constitutes half of the dualism implies that one's own good is the sole source of practical reasons (i.e., full-fledged egoism), then there is an inherent inconsistency with Rational Benevolence that not even God could alleviate. However, if the principle of self-interest is merely a prohibition against self-sacrifice, it becomes clear why Sidgwick thought that God's existence would solve the problem. To see this, consider the two principles (on my preferred formulations) side by side:

> *RB Universal Good (UG):* With respect to the rational aim of promoting the good, one ought to aim at universal good (rather than just one's own).
>
> *No Self-Sacrifice (NSS):* One ought never to sacrifice one's own good.

There is no formal contradiction between these principles. However, they do lead to practical conflicts in the world as we experience it. There are cases in which the act most conducive to promoting universal good would involve a sacrifice of one's own good. But if a just God exists, then this would not be the case because the quality of one's afterlife would be determined by how well one adhered to the aim of promoting universal good. Hence, by interpreting the self-interested element of the dualism as prohibiting self-sacrifice, we arrive at an interpretation that renders Sidgwick's views about the relationship between God and the dualism coherent.

One might object that Sidgwick should have rejected NSS because it would be denied by most moralists and hence fail the fourth epistemic test. But this is far from clear. While most of Sidgwick's contemporaries advocated for acts of self-sacrifice, the vast majority were theists who believed that conforming to duty was ultimately in one's long-term self-interest. It's hard to know what their intuitions would be on the supposition that we live in a godless universe. Further, most ancient Greek ethicists were eudaimonists who denied that acting justly could diminish one's happiness. Perhaps the most obvious source of potential disagreement over NSS is Kant, although even he argued that practical reason provides a type of warrant for belief in a just God who will ensure that moral conduct is justly rewarded in the afterlife (see Paytas 2020). Hence, it was not altogether unreasonable for Sidgwick to assume general agreement that a true sacrifice of one's overall well-being is always contrary to reason.

The *Methods* famously concludes with Sidgwick lamenting the apparent lack of harmony between the ends of self-interest and universal benevolence. Even taking natural sympathy and social sanctions into account, there are some cases in which promoting general happiness seems to require a sacrifice of one's own. In light of such conflicts, it looks rather unfair for Sidgwick to have taken the conflicts among deontological principles as a reason for favoring consequentialism. Further, since RB (UG) is not an all-things-considered principle, it cannot yield determinate guidance by itself, which is a standard that Sidgwick criticizes deontological principles for failing to meet. Thus, even if worries about equivocation are avoided on my proposed reading of RB, the charge of unfairness still needs to be addressed.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

4. The Unfairness Objection

4.1. Application of the Four Criteria

Commentators who press the unfairness objection overemphasize certain features of Sidgwick's argument while downplaying others. If Sidgwick's preference for consequentialism were based on the deontological principles failing to perfectly meet all four criteria, then he may indeed be guilty of unfairness. But while the four criteria play an important role in Sidgwick's rejection of deontology, their application is not all-or-nothing. This is made evident in his introduction of the criteria where he states that what we need are propositions that "approximately realize" these standards so that our reasoning can lead us to "trustworthy conclusions" (*ME* 338). Even if a proposition fails to perform perfectly against all four tests, we may still have grounds to privilege it over a proposition that fares worse. And Sidgwick provides substantial evidence that the deontological propositions of common-sense morality fare worse than RB on these measures.

As we have seen, RB runs into difficulty with the third criterion because it seemingly issues verdicts that conflict with NSS. However, we have also seen that Sidgwick thinks there is a potential solution in the existence of God. While Sidgwick is not confident that God exists, it is at least a live possibility, and he even suggests that the fact that God's existence would resolve the dualism might provide a coherentist warrant for theism (*ME* 508–509).[20] This is important because it means that there is a disparity between consequentialism and deontology regarding the issue of conflicts. While both methods contain principles that appear to yield practical conflicts, in the case of consequentialism, it's at least possible that the appearance of such conflicts is illusory (because God might exist). This gives consequentialism an advantage over deontology.[21] Of course, this could not be sufficient for high confidence in consequentialism, given that the divine solution remains speculative. But this is precisely why the *Methods* ends on a somber note of skepticism about philosophical ethics with the mere flicker of hope that theistic belief might somehow be vindicated.

A major discrepancy between RB and the common-sense duties involves the second criterion, which is a test to see whether the initial appearance of self-evidence might have arisen through epistemically pernicious causes. Sidgwick finds it plausible that the initial intuitiveness of common-sense duties may be due to the influence of custom, positive law, or non-rational impulses (*ME* 339–341). Regarding custom, Sidgwick notes that when we consider the moral codes of other cultures, we easily recognize elements of mere convention, and so we may reasonably suspect that the same is true of our own moral code (*ME* 341). Indeed, when he reviews common-sense morality in III.XI, he finds that many of the purported duties are not really self-evident but merely appear so due to conventions that arose from general expediency (e.g., the special duty of parents to children (*ME* 347) and the prohibition against suicide (*ME* 356)). This point is reiterated near the end of III.XIII, where he describes the common-sense rules as "precepts to which custom and general consent have given a merely illusory air of self-evidence . . . " (*ME* 383).

These worries do not arise for RB, which is an abstract principle that is (1) derivable from more basic intuitions (*ME* 382), and (2) does not readily admit of non-rational

[20]There are some interesting parallels between this suggestion and Kant's practical/moral justification for faith in God. I examine these parallels in Paytas (2020).

[21]Deontology can avoid conflicts via exception clauses or prima facie formulations. But as explained in section 2, Sidgwick has objections to these strategies.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

origins such as inclination or convention.[22] Sidgwick reports that when he reflects on propositions such as "I ought to speak the truth" and "I ought to keep my promises," they present themselves as requiring some more fundamental justification for their acceptance (*ME* 383). In contrast, he reports experiencing an "immediate and certain cognition" of RB, and it does not seem to him to require any more basic justification (2000: 25; *ME* 383).

One of Hurka's chief criticisms is that while Sidgwick's stated aim was to find genuinely self-evident ethical principles, utilitarianism is not self-evident, and Sidgwick's case for it depends on a discursive argument against deontological principles (2014b: 150). While this is true, it does not support the notion that Sidgwick was guilty of unfairness.

First, although utilitarianism is not self-evident, it does have an apparently self-evident basis in RB. Sidgwick finds more reasons to be skeptical about the intuitive appeal of the duties of common-sense morality than RB. Second, Sidgwick believes consequentialism enjoys an advantage in that the deontological principles are ultimately subordinate to the principle of utility. When particular maxims of duty come into conflict with one another, there does not appear to be a principled way of adjudicating the conflict other than appealing to considerations of utility (*ME* 316, 348). Likewise, when we attempt to locate the boundaries of a common-sense duty, the most plausible strategy is to appeal to expediency (*ME* 348–9, 352, 354–6, 359). Moreover, common-sense morality needs to appeal to utilitarianism in order to provide a deep unifying explanation for the various purported duties. The explanation is that those are the rules that are most conducive to general utility. Our ethical inquiries are not just an attempt to determine what we ought to do, we are also in search of a deeper explanation for *why* we ought to do it. Because we seek this deeper explanation, Sidgwick believes that a system containing the various duties of common-sense morality without a unifying explanatory principle is unsatisfactory, even if all of the contradictions, vagueness, and indeterminacies were removed. He puts the point thus: "Even granting that these rules can be so defined as perfectly to fit together and cover the whole field of human conduct, without coming into conflict and without leaving any practical questions unanswered,—still the resulting code seems to be an accidental aggregate of precepts, which stands in need of rational synthesis" (*ME* 102; see also 421–22 and 2000: 173).[23]

Sidgwick reports that he did not always recognize the distinction between the common-sense duties and the genuine ethical axioms. But he believes that careful reflection reveals a deep difference that favors consequentialism over deontology. The fact that there appear to be contexts in which RB and NSS cannot be simultaneously adhered to explains why Sidgwick does not take himself to have presented a vindication of utilitarianism (or even consequentialism) in the *Methods*. However, the fact that RB fares better against the first two criteria suggests an advantage over the deontological principles. This is not to say that the advantage should be seen as decisive. The point is just that Sidgwick's application of his criteria yields an argument for consequentialism that cannot be easily dismissed on grounds of unfairness.

[22]Lazari-Radek and Singer (2014: ch. 7) emphasize this point as a means of dissolving Sharon Street's (2006) "Darwinian dilemma" for value realism and as part of an argument for utilitarianism.

[23]It's open to a deontologist to insist that Sidgwick is mistaken about the importance of unifying explanations. My aim here is not to prove that Sidgwick was right about this. My point is that Sidgwick's philosophical preference for unifying explanation is at least not unreasonable and that recognizing this feature of his argument helps to mitigate the charge of unfairness.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

One might object that Sidgwick frequently makes much stronger claims about the epistemic status of RB than my interpretation suggests. For instance, he writes: "I find that I arrive, in my search for *really clear and certain ethical intuitions*, at the fundamental principle of Utilitarianism" (*ME* 387, emphasis added). This is indeed a strong statement. But note that this is not a claim that utilitarianism or RB has been conclusively proven (he obviously doesn't believe that given what he says in note 4 on the previous page about the "profoundest problem of Ethics" (*ME* 386). The remark in question is merely a report of where he has arrived up to this point. The search for clear and certain intuitions has led him to RB because, unlike the deontological principles, RB (1) seems not to require further justification, (2) is not subordinate to other duties, (3) synthesizes common-sense duties, (4) is not likely the result of convention or other non-rational influences, and (5) can be derived from more basic self-evident axioms. All these advantages have led him to RB at this stage, although that is not the end of the story because he still must address conflicts with the intuition against self-sacrifice.

A similar remark is made a few pages later: "Utilitarianism is thus presented as the final form into which Intuitionism tends to pass, when the demand for *really self-evident first principles* is rigorously pressed" (*ME* 388, emphasis added). Again, this should not be read as a claim that either utilitarianism or RB has been proven. What Sidgwick is saying here is that the method of intuitionism, when taken from dogmatic to philosophical, leads away from common-sense duties and towards a principle that systematizes and explains those duties without itself calling out for further justification. When he says "really self-evident," he is not suggesting that RB has been conclusively vindicated. He is referring specifically to the second epistemic criterion, which determines whether the air of self-evidence regarding a given principle is illusory. This is precisely the point emphasized a few pages prior in explaining the advantage of his principles over common-sense duties: "No doubt these principles [Justice, Prudence, and RB] are often placed side by side with other precepts to which custom and general consent have given a merely illusory air of self-evidence: but the [epistemic] distinction between the two kinds of maxims appears to me to become manifest by merely reflecting upon them" (*ME* 383).

Another passage that may seem problematic for my interpretation is the following: "I find that I undoubtedly seem to perceive, as clearly and certainly as I see any axiom in Arithmetic or Geometry, that it is 'right' and 'reasonable' for me . . . to do what I believe to be ultimately conducive to universal Good or Happiness" (*ME* 507). If RB is as certain as the axioms of geometry, this would suggest that Sidgwick's view of the epistemic status of RB is much stronger than my reading would allow. But notice that he does not say that RB (or utilitarianism) *is* as certain as the axioms of geometry. He's reporting his intuition—he *seems to perceive* RB as clearly as he perceives the axioms of geometry. Here again the second criterion is key. Like the geometric principles, RB is ascertained by careful reflection—it does not appear debunkable and it does not call out for further justification. The problem is that although he seems to perceive RB just as clearly as the geometric principles, it cannot ultimately be reliable if it yields practical conflicts with another equally intuitive principle (i.e., NSS). Hence, Sidgwick despondently concludes on the next page that, in the absence of a reliable connection between duty and self-interest, the rationality of ethics is "after all illusory" (*ME* 508). Still, the fact that RB does well by the second criterion (and has the other advantages noted above) explains why he views it more favorably than the deontological principles.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

4.2. Determinate Guidance

Another allegation of unfairness concerns action-guidance. Several passages in the *Methods* suggest that a major source of Sidgwick's concern about dogmatic intuitionism is that the duties fail to yield determinate guidance. For instance, in discussing common principles of justice and veracity, he complains that when we ask pertinent questions about real-life contexts, the principles fail to provide clear verdicts (*ME* 215). But this complaint seems to raise trouble for his own axioms. Sidgwick prefaces the presentation of his axioms by noting that they are too abstract and universal to yield determinate judgements about particular cases (*ME* 379). Given that his own axioms cannot provide clear decisions, how can he be justified in rejecting the deontological principles on these grounds? (Phillips 2011: 101; Phillips 2022: 117; Crisp 2016: 123–24.)

The first step in meeting this objection is to note that Sidgwick's concern about action-guidance within common-sense deontology does not imply that such guidance is required for an ethical proposition to be useful and properly justified. The four criteria do not include a requirement that the proposition in question issues determinate verdicts about particular cases. The first criterion requires clarity and precision, but that is not the same as action-guidance. It's true that Sidgwick often points to a lack of practical guidance in his criticisms of common-sense duties, which come shortly after his presentation of the criteria. However, the lack of action-guidance is problematic because the precepts in question are structured so as to be directly action-guiding (e.g., "Keep your promises"), and yet attempting to comply with them immediately raises a variety of questions that lack clear answers. The resulting uncertainties cast doubt on the trustworthiness of the intuitions. When we realize, for instance, that in real-life contexts it is often unclear whether a particular promise ought to be kept, we should have less confidence that the general duty to keep one's promises is a fundamental intuition rather than a rule of thumb that is justified and explained by the fact that norms of promise-keeping are conducive to promotion of universal good.

This problem would be less significant if the method of common-sense deontology included a principled means of resolving the various indeterminacies. But Sidgwick finds no means of doing so other than appealing to expediency, which he takes as evidence that the maxims are ultimately subordinate to utilitarianism (*ME* 422, 497). Now, it's true that Sidgwick's preferred axioms do not issue clear verdicts about what an agent ought to do in particular contexts. However, the difference is that the axioms do not purport to be immediately action guiding—they are abstract principles that are meant to serve as foundational elements of a comprehensive theory that does issue determinate guidance.[24]

Still, there are additional worries for Sidgwick concerning action-guidance. Just as common maxims of duty are often indeterminate and imprecise, utilitarianism contains its own indeterminacies. As Sidgwick recognizes, utilitarianism can require maximization of the total good of the population or the average good of each member, and it is not clear which option is superior (*ME* 415–16). He also acknowledges that utilitarianism appears silent on the issue of how to appropriately distribute a given quantity of happiness over a given population (*ME* 416–17). Given that resolution of these issues is necessary for comprehensive guidance, and none of the candidate answers to these questions is self-evident, it would appear that utilitarianism is in no better position with respect to action-guidance.

[24]Lazari-Radek and Singer (2014: 144-48) emphasize this point in their response to the unfairness objection.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

As with the epistemic criteria, we should not assume that the criterion of action-guidance is all-or-nothing. Some of Sidgwick's remarks suggest that a method's failing to yield precise verdicts in all contexts does not warrant its abandonment. For instance, at the conclusion of a lengthy discussion of the various indeterminacies inhering in empirical hedonism, Sidgwick writes: "I do not conclude that we should reject it altogether: I am conscious that, in spite of all the difficulties that I have urged, I continue to make comparisons between the pleasures and pains with practical reliance on their results" (*ME* 150).

This point is relevant to the unfairness objection because while there does appear to be ethical uncertainty within utilitarianism, there is substantially less than in pluralistic deontologies. Given that all plausible deontological views include a duty of benevolence, these views will inherit the uncertainties of consequentialism anyway. Even if the duty to promote happiness is just one among many, we still need to know whether we should be concerned with total or average and whether equality of distribution matters. Hence, although consequentialism appears incapable of yielding determinate guidance in all contexts based solely on self-evident principles, a reasonable case can be made that it fares better than deontology on this score. Thus, Sidgwick can be justified in appealing to this issue as a reason for favoring consequentialism over deontology without being guilty of unfairness.

Here one might reply that utilitarianism's emphasis on pleasure maximization yields especially pervasive uncertainty. For instance, when deciding which dessert to order, numerous factors are relevant for determining which option will maximize one's own pleasure as well as overall utility. It is simply not feasible to grasp all of the relevant facts and calculate them accurately. Hence, the only means of proceeding is to use one's judgment to make a reasonable guess (Crisp 2016: 122).

While this is true, it does not undermine the claim that consequentialism fares better than deontology in this respect. Sidgwick's concern about action-guidance is not solely about the desire for assurance that we are acting correctly, which is harder to come by when we have to rely on judgment rather than principles. The issue also pertains to our confidence that the purported principle we are trying to follow is true. In the dessert case, I must use my judgment because I do not have ready access to all the empirical facts. Hence, the uncertainty is empirical rather than ethical. In contrast, within pluralistic deontology, individual judgment is frequently necessary for determining whether a given act falls under the scope of a duty (recall the problematic instances of promise-keeping), and which of two principles takes precedence in cases of conflict. This is uncertainty about basic ethical facts.

Such rampant *ethical* uncertainty may be grounds for having less confidence that the principles we are intuiting are ultimately valid. The suggestion is that if we have a capacity to intuit moral truths, we should expect to be able to intuit principles about their relations and applications, at least in most cases. If a moralist says, "I'm certain that promises should be kept," and when pressed about the myriad complicating factors says, "Well, I don't think we can know the truth about what to do in real-life cases where things get complicated," it doesn't seem unreasonable to become less confident in the moralist's initial assertion about the general duty of promise-keeping. On the other hand, if the moralist can point to a plausible underlying principle that (1) explains what to do in the complicated cases, (2) resolves conflicts between apparent duties, and (3) provides a unifying explanation of them, it would not be unreasonable to find the moralist's claims more credible. I take it that this thought is central to Sidgwick's preference for consequentialism over deontology. There is of course room for deontologists to push back here. Perhaps Sidgwick is wrong about the significance of ethical uncertainty. But being wrong is not the same as being unfair.

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

5. Conclusion

In summary, Sidgwick's case for the superiority of consequentialism over deontology involves four key issues:

(1) Practical conflicts: The duties of common-sense morality often yield contradictory verdicts. This needn't be troubling if there is a principled means of adjudicating the conflicts. But it appears that the only viable option is an appeal to utility. This suggests an advantage for consequentialism. Still, the principle which is said to provide the foundation for utilitarianism—Rational Benevolence—seems to yield practical conflicts with an equally plausible principle prohibiting self-sacrifice. This prevents RB from having high trustworthiness. However, the possibility of God's existence constitutes a potential solution to this problem. Given this possibility, RB is in a superior position regarding the issue of practical conflicts (because it is possible that there are none), even though its truth remains far from certain.
(2) Reliability: It is plausible that duties of common-sense morality gain their initial intuitiveness from the influence of custom, positive law, and innate biases and inclinations. The fact that there is variance in moral codes across different cultures attests to the plausibility that these epistemically pernicious influences play a role in shaping common-sense morality (*ME* 341). There is less reason to believe that the initial credibility of RB is a product of culture or non-rational impulses. Not only do human beings innately tend to prioritize their own interests, but societies are generally structured such that partiality towards oneself and one's kin is the default. Further, RB is derivable from two abstract axioms (that one's own good is no more valuable than anyone else's, and that rational beings are bound to aim at good generally). Hence, RB seems to fare better than deontology on epistemic grounds.
(3) Subordination: One of the chief problems with common-sense deontology is that the various duties appear to be an "accidental aggregate" that calls out for some deeper explanation and "rational synthesis" (*ME* 102). A plausible synthesizing explanation is that the rules of common-sense morality are those that tend to be conducive to the promotion of general welfare. Hence, a fundamental principle of benevolence can provide the needed deeper explanation for the various duties and resolve conflicts between them. These points suggest that deontology is ultimately subordinate to consequentialism.
(4) Action-guidance: One desideratum of an ethical theory is that it provides determinate guidance. Deontology struggles on this score because the boundaries of the various duties are unclear, and they often generate conflicts with no principled (non-utilitarian) means of adjudication. While utilitarianism does not yield comprehensive guidance, it appears to fare better than deontology in this respect for two reasons. First, much of the indeterminacy in utilitarianism is empirical rather than ethical. Second, most (if not all) of the indeterminacy found in utilitarianism, including ethical indeterminacy (e.g., total vs. average good), will be inherited by deontology since any plausible version will include duties of beneficence.

It has not been my intention to demonstrate that Sidgwick's argument is entirely successful—there are numerous points at which reasonable objections can be raised. The

https://doi.org/10.1017/S0953820825100113 Published online by Cambridge University Press

Methods certainly would have benefited from his being more attentive to certain theoretical difficulties for consequentialist views. And there is merit to the criticism that Sidgwick was not always as clear a writer as his reputation suggests. The complexities involved in interpreting Rational Benevolence and the dualism of practical reason illustrate the point that Sidgwick was often least clear on the most important points (Hurka 2014b: 151). Still, I hope to have shown that Sidgwick's case for the superiority of consequentialism over deontology has more to be said in its favor than recent critics suggest. And most importantly, I hope to have shown that his overall argument was not plagued by unfairness.

Acknowledgements. Versions of this paper were presented at meetings of the American Philosophical Association, the Australasian Association of Philosophy, the Rocky Mountain Ethics Congress, and the International Society for Utilitarian Studies. I thank the audiences at these events for helpful discussions. I am grateful to Peter Zuk and David Phillips for their excellent commentaries following two of these presentations. David Phillips was especially generous in providing additional feedback through multiple conversations and email exchanges. Special thanks are owed to Rob Shaver for reading three different drafts and providing immensely helpful feedback on each. I am also grateful to Tom Hurka for a rigorous referee report on an earlier version, which led to several important additions and refinements. I thank Julia Driver for supervising the PhD thesis in which I began developing the interpretations presented here, and Roger Crisp for lending his expertise and support as the external committee member for that project. Finally, for many enlightening and encouraging conversations about Sidgwick, I thank Katarzyna de Lazari-Radek, Christian Seidel, Peter Singer, and Anthony Skelton.

References

Ackeren, Marcel Van, and Martin Sticker, 'Kant and Moral Demandingness', *Ethical Theory and Moral Practice*, **18** (2015), 75–89.

Brink, David, *Perfectionism and the Common Good* (Oxford: Clarendon Press, 2003).

Broad, C. D, *Five Types of Ethical Theory* (London: Kegan Paul, Trench, Trubner, 1930).

Crisp, Roger, 'Sidgwick and the Boundaries of Intuitionism', in *Ethical Intuitionism: Re-evaluations*, ed. by Philip Stratton-Lake (New York: Oxford University Press, 2002), pp. 56–75.

Crisp, Roger, *The Cosmos of Duty: Henry Sidgwick's Methods of Ethics* (Oxford: Oxford University Press, 2015).

Crisp, Roger, 'Rossian Pluralism, Egoism, and Pleasure', *Etica & Politica/Ethics & Politics*, **18** (2016), 119–33.

Lazari-Radek, Katarzyna, and Peter Singer, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford: Oxford University Press, 2014).

Donagan, Alan, 'Whewell's Elements of Morality', *Journal of Philosophy*, **71** (1974), 724–36.

Donagan, Alan, *The Theory of Morality* (Chicago: University of Chicago Press, 1977).

Donagan, Alan, 'Sidgwick and Whewellian Intuitionism: Some Enigmas', in *Essays on Henry Sidgwick*, ed. by Bart Schultz (Cambridge: Cambridge University Press, 1992), pp. 123–42.

Hurka, Thomas, *British Ethical Theorists from Sidgwick to Ewing* (Oxford: Oxford University Press, 2014a).

Hurka, Thomas, 'Sidgwick on Consequentialism and Deontology: A Critique', *Utilitas*, **26** (2014b), 129–52.

Irwin, Terence, *The Development of Ethics, vol. 3: From Kant to Rawls* (Oxford: Oxford University Press, 2009).

Korsgaard, Christine, *Self-Constitution* (New York: Oxford University Press, 2009).

McLeod, Owen, 'What is Sidgwick's Dualism of Practical Reason?', *Pacific Philosophical Quarterly*, **81** (2000), 273–90.

McTaggart, John, 'The Ethics of Henry Sidgwick', *Quarterly Review*, **205** (1906), pp. 398–419.

Moore, G. E, *Principia Ethica* (Cambridge: Cambridge University Press, 1903).

Parfit, Derek, *On What Matters*, Volume **1** (Oxford: Oxford University Press, 2011).

Paytas, Tyler, 'Beneficent Governor of the Cosmos: Kant and Sidgwick on the Moral Necessity of God', in *Kantian and Sidgwickian Ethics: The Cosmos of Duty Above and the Moral Law Within*, ed. by Tyler Paytas and Tim Henning (New York: Routledge, 2020), pp. 210–44.

Phillips, David, *Sidgwickian Ethics* (Oxford: Oxford University Press, 2011).
Phillips, David, *Sidgwick's Methods of Ethics: A Guide* (New York: Oxford University Press, 2022).
Ross, W. D, *The Right and the Good* (Oxford: Oxford University Press, 1930).
Schneewind, Jerome, *Sidgwick's Ethics and Victorian Moral Philosophy* (Oxford: Clarendon Press, 1977).
Schultz, Bart, *Henry Sidgwick: Eye of the Universe* (Cambridge: Cambridge University Press, 2004).
Shaver, Robert, *Rational Egoism: A Selective and Critical History* (New York: Cambridge University Press, 1999).
Shaver, Robert, 'Sidgwick's Axioms and Consequentialism', *Philosophical Review*, **123** (2014), 173–204.
Shaver, Robert, 'Peerless: Sidgwick, Kant, and Peer Disagreement', in *Kantian and Sidgwickian Ethics: The Cosmos of Duty Above and the Moral Law Within*, ed. by Tyler Paytas and Tim Henning (New York: Routledge, 2020), pp. 118–37.
Sidgwick, Henry, *The Methods of Ethics*, 7th edn (Indianapolis: Hackett, 1981).
Sidgwick, Henry, *Essays on Ethics and Method*, ed. by Marcus G. Singer (Oxford: Clarendon Press, 2000).
Singer, Peter, 'Sidgwick and Reflective Equilibrium', *The Monist*, **58** (1974), 490–517.
Skelton, Anthony, 'Sidgwick's Philosophical Intuitions', *Etica & Politica/Ethics & Politics*, **10** (2008), 185–209.
Skelton, Anthony, *Sidgwick's Ethics* (Cambridge: Cambridge University Press, 2025).
Street, Sharon, 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies*, **127** (2006), 109–66.
Timmermann, Jens, 'Good but Not Required? — Assessing the Demands of Kantian Ethics', *Journal of Moral Philosophy*, **2** (2005), 9–27.

Cite this article: Paytas, Tyler. 2025. Sidgwick's Critique of Deontology: Scrupulous Fairness or Serpent-Windings?. *Utilitas* 1–18. https://doi.org/10.1017/S0953820825100113