

ARTICLE

# The Role of Kant in Sidgwick's Classical Utilitarianism: Two Self-Evident Axioms and the Partial Convergence between Kantianism and Utilitarianism

Annette Dufner

University of Bonn, Bonn, Germany  
Email: [annette.dufner@uni-bonn.de](mailto:annette.dufner@uni-bonn.de)

## Abstract

Among the most surprising claims in *The Methods of Ethics* is Sidgwick's assertion that his key ethical axioms are corroborated by Kant. This article analyses Sidgwick's claim that his axioms of justice and benevolence closely correspond to particular features in Kant. I shall argue that his claim of agreement with Kant was a serious overstatement. In particular, the restrictions which Sidgwick places on his acceptance of Kant's universal law formula of the categorical imperative (FUL) seem to call into question whether the alleged convergence with the axiom of justice has a solid basis. Further, Sidgwick seemed unaware of a crucial aspect of Kant's conception of the humanity formula that constitutes a substantial divide between their views on benevolence. The upshot is that the divide between Kantian and Sidgwickian ethics appears deeper than Sidgwick seemed to realize. This analysis is confirmed by Sidgwick's famous worries regarding freedom and the existence of God in Kant's work.

**Keywords:** Kant; universal law formula; humanity formula; beneficence; Sidgwick; axiom of benevolence; axiom of justice; utilitarianism

In his seminal book *The Methods of Ethics*, Henry Sidgwick argues that the axioms of justice and benevolence not only form the foundation of utilitarianism but are also confirmed by the Kantian system.<sup>1</sup> Given the often-presumed, strong dichotomy between utilitarianism and Kantianism, this is a striking claim. Unsurprisingly, many Kantians and utilitarians have doubts about the alleged convergence. While one might initially suspect that the claim of agreement from Kant must be due to a superficial understanding of Kant's ethics, Sidgwick appears to have read Kant very closely.<sup>2</sup> Despite his often very knowledgeable discourse on Kant,<sup>3</sup> there was surprisingly little literature on Kant's role in Sidgwick's utilitarian axioms until recently.<sup>4</sup>

This article seeks to assess the plausibility of Sidgwick's convergence claim. To intensify its focus, the often-pursued metaethical question of whether his classification of the Kantian system as an example of intuitionism is justified will be left aside.

Instead, the focus will be on the formulation of the axioms of justice and benevolence and on the passages and arguments of Kant, to which Sidgwick refers in order to support his convergence claim. Sidgwick believes that there is a close connection between his axiom of justice and Kant's Formula of Universal Law (FUL). However, the remaining divergence in scope between these two positions – which Sidgwick explicitly admits – is so fundamental that one may wonder whether the convergence claim has a solid basis. Sidgwick also argues that there is a close correspondence between the axiom of benevolence and benevolence as a duty in Kant. This article argues that previous views calling this claim into question do not offer a complete picture. Sidgwick seems to have been unaware of a crucial element of the humanity formula of the categorical imperative that constitutes a further substantial divide between the respective views. Moreover, the view defended here has the virtue of connecting the ethical differences between the authors to Sidgwick's famous worries about the metaphysical basis of Kant's system.

### 1. The axiom of justice and the universal law formula of the categorical imperative

How many self-evident axioms Sidgwick detected is open to debate, but on first sight, there appear to be three: the axiom of justice, the axiom of prudence and the axiom of benevolence.<sup>5</sup> The axiom of prudence will be left aside since Sidgwick does not claim convergence with Kant on this subject. The canonical formulation<sup>6</sup> of the axiom of justice reads as follows:

[I]t cannot be right for A to treat B in a manner in which it would be wrong for B to treat A, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment. (Sidgwick 1981 [1907]: 380)

In short, this axiom claims that it cannot be right to treat two persons differently, unless a relevant difference between them can be identified as the decisive reason for this. Sidgwick claims that there is a literal logical dependency relationship between this principle, or axiom, of justice and the categorical imperative. He argues that Kant's FUL, 'duly restricted', is 'an immediate practical corollary' of the axiom cited above (Sidgwick (1981 [1907]: 386).<sup>7</sup> Kant's FUL requires in Sidgwick's own words: 'acting on a maxim that one can will to be law universal' (ibid.; corresponding to Kant, *G*, 4: 421). As quoted, Sidgwick clearly states that this formulation of the categorical imperative follows logically from the axiom.

Sidgwick explicitly claimed to be impressed by FUL, but he never accepted all of Kant's examples of it and rejected its ultimate rationale. In particular, he did not believe that FUL could support a general prohibition against lying or the duty of beneficence.<sup>8</sup> Even more importantly, he denied that morality can ultimately be grounded in human freedom. In his introduction to the sixth edition of the *Methods*, he writes about Kant's system:

I . . . was impressed with the truth and importance of its fundamental principle: – *Act from a principle or maxim that you can will to be a universal law* . . . Kant's resting of morality on Freedom did not indeed commend itself to me . . . What commended itself to me, in short, was Kant's ethical principle rather than its metaphysical basis. (Sidgwick 1981 [1907]: p. xix)

The restrictions about the examples and about the metaphysical basis of Kant's view which Sidgwick placed on his acceptance of FUL seem to support Kantians who are doubtful about the existence of the close relationship that Sidgwick sees between his axiom and Kant's formula. Nonetheless, some possible attempts to formulate the central difference are less promising than others. In particular, it is often thought that utilitarians tend to hold a problematic, teleological interpretation of FUL, according to which the world would simply end up in a worse state if everyone acted on the basis of non-universalizable maxims.<sup>9</sup> It is possible that Mill is assuming a teleological interpretation of FUL when he says that '[t]o give any meaning to Kant's principle, the sense put upon it must be, that we ought to shape our conduct by a rule which all rational beings might adopt *with benefit to their collective interest*' (Mill 1985 [1833]: 249, emphasis added). However, there is no evidence in the *Methods*, or anywhere else, that Sidgwick held such a teleological view. On the contrary, as he particularly emphasized in the *History of Ethics*, actions based on non-universalizable maxims cannot even be conceived or involve a contradiction of the will for Kant. To quote a central passage in full:

The rule excludes wrong conduct with two degrees of stringency. Some kinds of immorality – such as making promises with the intention of breaking them – *we cannot even conceive universalized*; as soon as every one held himself free to break his promises no one would care to have promises made to them. Other maxims, such as that of leaving persons in distress to shift for themselves, we can easily conceive to be universal laws, but *we cannot without contradiction will them to be such*; for when we are ourselves in distress we cannot help desiring that others should help us. (Sidgwick 1988 [1902]: 273, emphasis added)

Passages such as these make it very unlikely that Sidgwick can actually be accused of having held a teleological interpretation of Kant's FUL.

Another issue which could be raised is whether Sidgwick's axiom of justice could be a version of the Golden Rule, which Kant discredited. If this were the case, then Sidgwick's claim would actually mean that FUL is a logical corollary to a version of the discredited Golden Rule. Even though this would be an interesting position,<sup>10</sup> it is very unlikely that it can be attributed to Sidgwick. Immediately before his central characterization of the axiom, Sidgwick discusses various versions of the Golden Rule, agrees explicitly with Kant that the Rule is problematic and discusses some of the well-known reasons for this: the rule 'Do to others as you would have them do to you' could lead to absurd consequences if a person wishes that others would cooperate in his sins; he also rejects the negative version, according to which 'you should not do to others what you would take to be wrong if done to yourself', because there could be relevant differences between people which would render this version incorrect (Sidgwick 1981 [1907]: 379–80). In fact, it appears that Sidgwick's declared goal in

constructing the exact formulation of the axiom of justice was to provide a more precise and better principle than the various versions of the Golden Rule.

One difference lies in the fact that the exact formulation of the axiom of justice contains no reference to any subject's wishes, will or opinions about what is right, as is the case in the discredited versions of the Golden Rule. Instead, it claims that in cases in which it is simply wrong for B to treat A in a certain way, irrespective of anyone's beliefs or wishes about the matter, the reverse must also be false. A similarity which remains between the axiom of justice and these versions of the Golden Rule is the mirror-like comparison of particular kinds of scenarios, but the claim that what is right can be deduced from a particular individual's wishes, will or opinions is abandoned in favour of an assumption of objective rightness.

A further, related objection to the allegedly close relationship between Sidgwick's axiom of justice and FUL is that FUL was intended as a way of determining the moral rightness of a much larger variety of relevant maxims. In an earlier part of the *Methods*, Sidgwick explicitly states that an action that is acceptable according to FUL could in fact turn out to be wrong for other reasons:

Kant seems to have held that all particular rules of duty can be deduced from the one fundamental rule 'Act as if the maxim of thy action were to become by thy will a universal law of nature.' But this appears to me an error analogous to that of supposing that Formal Logic supplies a complete criterion of truth. I should agree that a volition which does not stand this test is to be condemned; but I hold that a volition which does stand it may after all be wrong. (Sidgwick 1981 [1907]: 209–10)<sup>11</sup>

As this passage demonstrates, Sidgwick did not accept the general law formula as a final, or a complete, test of the rightness or wrongness of all sorts of morally relevant maxims. In a similar way, Sidgwick only grants limited powers to the axiom of justice – the axiom is unable to settle all morally relevant questions on its own. Had he thought that the axiom could do so, then it could have remained the sole axiom – but he found others which are not identical in scope and meaning.

What remains somewhat puzzling is the reason Sidgwick gives for his assessment that FUL has only limited scope. Immediately after the passage quoted above, he explains that assuming that the categorical imperative can give complete guidance would lead to the obliteration of subjective and objective rightness: it would imply that anyone acting conscientiously could sincerely will his maxim to become a universal law, even if it were wrong. He writes:

For I conceive that all (or almost all) persons who act conscientiously could sincerely will the maxims on which they act to be universally adopted: while at the same time we continually find such persons in thoroughly conscientious disagreement as to what each ought to do in a given set of circumstances. Under these circumstances, to say that all such persons act rightly – in the objective sense – because their maxims all conform to Kant's fundamental rule, would obliterate altogether the distinction between subjective and objective rightness; it would amount to affirming that whatever anyone thinks right

is so, unless he is in error as to the facts of the case to which his judgment applies. (Sidgwick 1981 [1907]: 210)

It would indeed be strange if a conscientiously formed subjective belief – even a false one – according to which a certain maxim is universalizable, would in and of itself indicate that the respective act is permissible. If this were possible, there would no longer be a difference between subjective views about the right and objective rightness. Moreover, in cases in which there is moral disagreement, this would also lead to the absurd consequence that a particular act could turn out to be permissible and impermissible at the same time, depending on whose conscientiously formed belief we are looking at.

Kantians would have a number of possible responses to this very brief consideration. For example, in contrast to Sidgwick, a Kantian might be willing to simply bite the bullet and insist that the good will is decisive even if it is based on false beliefs and beliefs that others disagree about. The Kantian could also claim that merely *believing* that your maxim is universalizable does not amount to it *being* universalizable. Unrestricted, first-personal access to the rightness of a maxim is not a necessary feature of Kant's moral metaphysics. Sidgwick may literally not have seen some of the strong argumentative resources that Kant could have utilized to rebut this charge: a charge that seems to have led Sidgwick to grant only limited scope to FUL.

Despite Sidgwick's attempt to arrive at a contrary conclusion, the FUL probably achieves more than the axiom and thus has a broader scope.<sup>12</sup> In particular, it is supposed to explain what makes an action right or wrong in the first place. The canonical formulation of the axiom, on the other hand, is a meta-rule which already presupposes that certain actions are right or wrong and merely demonstrates that mirror-like extensions to similar cases are obligatory. While some of the restrictions that Sidgwick places on his acceptance of FUL, especially regarding the lying and FUL's support of beneficence, might be a matter of debate, one of the central reasons for his restrictions – the one about the obliteration of subjective and objective rightness – could rather easily be rejected by Kant. A stipulation of unrestricted subjective access to whether a maxim can rightly be willed is certainly not a feature of Kant's view.

Overall, claiming convergence with Kant on the grounds that Kant's view could be different in ways that he clearly rejects can be viewed as a strange argumentative approach. The question of whether the 'duly restricted' version of FUL that Sidgwick imagines and the axiom of justice are compatible has comparatively little to do with the relationship between the axiom and FUL as presented by Kant.<sup>13</sup>

## 2. The axiom of benevolence as supported by Kant's duty of beneficence

Sidgwick also believes that Kant would agree with the axiom of benevolence. The most famous formulation of this axiom<sup>14</sup> states:

[E]ach 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. (Sidgwick 1981 [1907]: 382)

According to this axiom, the welfare of all individuals is of equal value and one should promote general welfare rather than the welfare of any particular individual. Preference is to be given to individuals for whom one can promote more good than for others. Sidgwick argues that there is a close relationship between this axiom and Kant's views. More precisely, he argues that 'Kant's conclusion appears to agree to a great extent with the view of the duty of Rational Benevolence that I have given' (Sidgwick 1981 [1907]: 386).

Immediately before this remark about the presumed convergence with Kant, he elaborates on Kant's virtue of beneficence. In particular, he says that rational benevolence, or the happiness of others, is the only ultimate end Kant recognizes, and that Kant takes it to be evident *a priori* that each person as a rational agent is bound to aim at the happiness of others (Sidgwick 1981 [1907]: 386). In an added footnote about the Doctrine of Virtue, Sidgwick admits that one's own perfection is acknowledged by Kant as a second ultimate aim. However, in the realm of duties towards others, for Kant the ultimate purpose, as stated in the Doctrine of Virtue, is indeed their happiness (*Glückseligkeit*).<sup>15</sup>

In a short passage in Sidgwick's *History of Ethics*, there is a further, more explicit explanation of Kant's duty to promote the happiness of others which refers to both the *Groundwork* and the Doctrine of Virtue.<sup>16</sup> In this monograph, Sidgwick correctly characterizes the humanity formula as the imperative to treat humanity, both in one's own person as well as in the person of any other, as an end and never merely as a means (Sidgwick 1988 [1902]: 275, discussing G, 4: 429). He also highlights Kant's claim that human rational nature constitutes an end in itself (p. 274, discussing G, 4: 429). Thereafter, he raises the notorious question of what exactly it might mean to treat someone as an end, and comes to the conclusion that it cannot mean to treat others in such a way that their own moral perfection is promoted, because perfection is something that a person can only ever accomplish on their own. Instead, he further concludes that it can only mean that one ought to promote the happiness of others – more specifically, happiness in the form in which it represents itself to these other people. He finally states that the subjective happiness of others must become my own end if I want to treat them as an end.

The elucidations in the *History of Ethics* are adequately supported by references to the relevant passages in the Doctrine of Virtue and the *Groundwork*.<sup>17</sup> The relevant passage in the Doctrine of Virtue to which Sidgwick refers in the *History of Ethics* is the one in which Kant says that there are two ultimate ends, happiness and virtue, and that my own happiness is a natural inclination rather than a duty, and that others' perfection is not something that I can accomplish for them. In the realm of virtues towards others, Sidgwick continues, the promotion of their happiness is the only ultimate end acknowledged by Kant (Sidgwick 1988 [1902]: 275, discussing MM, 6: 385–6). The *Groundwork* passage to which Sidgwick refers in the *History of Ethics* is part of the discussion on why the humanity formula of the categorical imperative requires beneficence towards others. Kant admits that humanity could exist without anyone ever contributing to another person's happiness but demands a positive contribution to humanity in the form that assisting others in their pursuit of ends has to become one's own end to the largest extent possible (*ibid.*, discussing G, 4: 430).

Overall, that Kant acknowledges a duty to promote the happiness of others is not controversial. Sidgwick's modest formulation, that Kant 'appears to agree to a great

extent' with the axiom of benevolence, lends further initial credibility to his claim. Perhaps Sidgwick never meant to say more than that Kant agrees that benevolence is one's duty. Sidgwick's apparent general desire for finding points of convergence rather than disagreements might support such a modest interpretation of his claim. However, it would also make the alleged convergence with Kant much less interesting.

Many Kantians' primary objection to Sidgwick's partial convergence claim about benevolence will be on other grounds. For example, they will argue that Kant never thought that the size of the benefit that one can bestow on someone could be a reason for this person's preferential treatment and that Kant never thought that one should aim for the general good of all, rather than the individual good of particular others, presumably those with whom one is engaged in a rather immediate way. However, it is interesting that the passages which Sidgwick seems to have had in mind, namely the statement of the humanity formula and a further passage from the *Groundwork* which Sidgwick actually quotes in the *History of Ethics*, are written in a somewhat misleading way in this respect. The passage in question follows the central formulation of the humanity formula according to which you ought to act so that you use humanity, both in your own person and in the person of any other, as an end and never merely as a means.<sup>18</sup> The subsequent elucidatory passage, quoted by Sidgwick in the *History of Ethics*, contains a suggestive superlative:

[T]he ends of any subject which is an end in himself, ought *as far as possible* to be my ends also, if the conception of him as an end in himself is to have its full effect with me. (Sidgwick 1988 [1902]: 275, emphasis added)<sup>19</sup>

In this passage, Kant – if taken literally – seems to say that one ought to promote any other person's happiness to the greatest extent possible.<sup>20</sup> If this is correct, it seems to claim that it is not sufficient to promote the good of others merely to some extent, but rather that the impartial maximization of the good of others is required: a thought that can often be found in general descriptions of the utilitarian doctrine.

There is a chance that Sidgwick fell prey to an overly utilitarian understanding of this passage<sup>21</sup> – an understanding that is incompatible with Kant's actual view. This is confirmed by a passage in the *Doctrine of Virtue* where Kant says that there is discretion regarding the ways and extent to which the duty of practical love should be exercised. The duty of practical love allows for narrowing specification; for example, in form of the rule that one should honour one's parents (*MM*, 6: 390).

By now, several different interpretative attempts seek to make sense of this latent disagreement between the two authors. As Roger Crisp argues, in accordance with Terence Irwin (Crisp 2015: 123, 126; Irwin 2009: 512–13), ethicists in Kant's camp would likely object to the strong impartiality of Sidgwick's formulation of the axiom, which seems to preclude such partiality towards friends and family. Nonetheless, as Anthony Skelton has recently demonstrated (Skelton 2020: 22ff.), it should be remembered that Sidgwick explicitly considers the possibility that partiality towards close others might be beneficial overall.<sup>22</sup> This means that his general support of impartiality might have been weaker than the formulation of the axiom suggests and he might have only thought that there is no *ultimate* reason for partiality. This could imply that the divergence between the authors on this point is not as pronounced as one might think.

There is a further way in which Kant rejects impartiality, but which Sidgwick could, in principle, follow to some extent. Already on the first page of the *Groundwork*, Kant points out that there is no worth in the happiness of a morally undeserving being.<sup>23</sup> Arguably, this view has an impact on Kant's duty of benevolence: if one could choose either to make a morally deserving or undeserving person happy, one might have to be partial towards the morally deserving person. Even if human beings do not have a duty to distribute happiness according to moral desert and only God is in charge of this, the point remains that for Kant the value of happiness seems to depend on moral desert. In this sense, the value of happiness is not an impartial matter for Kant. However, arguably Sidgwick could also agree with this form of partiality. Just like partiality towards friends and family members, partiality towards the deserving might be indirectly conducive to the general good. Since Sidgwick's support for partiality towards the deserving would likely be instrumental as well, while there does seem to be a difference on the level of justification, of course, the practical divergence seems less clear.

There are further reasons for the claim that reading an individual or a general happiness maximization requirement into the quoted passage from the *Groundwork* – as Sidgwick may have done – must be wrong. For instance, Kant would assume that the rational ends of human beings must, at a minimum, be permitted by the categorical imperative – a requirement that happiness, as an end, obviously does not automatically satisfy.<sup>24</sup> Furthermore, for Kant, benevolence is an imperfect duty, so if there were a conflict with a perfect duty, Kantians typically argue that the duty of beneficence would actually be suspended.<sup>25</sup> While Sidgwick seems to have been to some extent open to such a non-absolute view about benevolence in his own account (Sidgwick (1981 [1907]: 421), he does not address the fact that it is clearly a feature of Kant's account.

As a consequence, Thomas Hurka and Terence Irwin plausibly conclude that Sidgwick did not distinguish clearly enough between benevolence as an absolute principle and as an 'other-things-equal' principle (Hurka 2014: 149, 151–2; Irwin 2009: 518). Sidgwick might not have taken full account of the imperfect character of Kant's duty of beneficence, which can be outweighed by perfect duties. Consequently, he might have assumed Kant would agree with the more broadly construed utilitarian form of benevolence.

However, as shall be demonstrated, the Hurka-Irwin thesis does not offer a complete picture of the divergence between Kant and Sidgwick when it comes to beneficence. As Alan Donagan points out, Sidgwick did not accept the view that humanity, with its capacity for rationality, can be viewed as an end in itself.<sup>26</sup> For Sidgwick, the humanity formula *primarily* says that one should adopt the practical ends of others as one's own. But this is not how Kant saw it. For Kant, an end is not merely something to be produced, a consequence to be brought about; rather an end in itself is an already existing rational individual for the sake of whom one ought to act. Treating someone as an end in itself is important because it means respecting this human individual's inherent freedom. Adopting the aims of others, practising beneficence, is merely a plausible way of respecting what is *really important* to human beings: to be respected as rational free agents, as ends in themselves.

We see clearly Sidgwick's worries about the role of rationality in the humanity formula in a lengthy footnote to the chapter containing the convergence claim.



Here he explains his problems seeing or accepting Kant's derivation of the duty to promote the happiness of others from the humanity formula. As he puts it toward the end of the footnote:

It is hard to see why, if man as a *rational being* is an absolute end to other rational beings, they must therefore adopt his subjective aims as determined by his non-rational impulses. (Sidgwick 1981 [1907]: 390)<sup>27</sup>

In addition to expressing his central concern about how to derive positive duties, in particular the duty of benevolence from an end understood as something already there, he is generally sceptical about Kant's understanding of humanity as an end in itself.

Moreover, he did not see that Kant's meaning of the term 'end' in the humanity formula provides a principled reason for excluding at least the morally irrational, and arguably even the prudentially irrational impulses of others.<sup>28</sup> The passage about having to adopt other people's ends as one's own has to be read in light of the humanity formula according to which the *reason* for taking the ends of others as important is their character as rational human beings who are ends in themselves. Promoting morally (and perhaps also prudentially) irrational ends of others is not a plausible way of showing respect for their capacity as rational beings who are ends in themselves. It can certainly be argued that Sidgwick's puzzlement in his footnote can be explained by an assumption on his part according to which he did not see the non-absolute character of Kant's duty of benevolence, as Hurka and Irwin would argue. However, Sidgwick's error can be even better explained by noting a fundamental oversight regarding the full Kantian meaning of an end and its implications for the duty of beneficence.

Sidgwick failed to take proper account of the decidedly non-absolute nature of beneficence in Kant. For Kant beneficence is merely an imperfect duty that can be overridden by perfect duties. To say this differently, beneficence is overridden whenever the ends of the other person that beneficence is meant to promote are morally irrational. In this case the ends do not express the rationality constitutive of a human being as an end in and of itself.

The analysis suggested here is strongly supported by a passage in the *History of Ethics*. As indicated earlier, in the *History of Ethics* Sidgwick explicitly raises the question of what it might mean to treat someone as an end. His response to this rhetorical question, in the exact context of the discussion of the humanity formula, very clearly indicates that he adopts an understanding of an end as being a goal to be produced, and that he rejects the idea that treating someone as an end might mean treating him as an end in itself, by respecting his rational human capacities:

If we ask what precisely are the ends of reason – meaning by 'end' a result which is sought to be produced by action – Kant's proposition that 'all rational beings as such are ends in themselves for every rational being' hardly gives a clear answer. (Sidgwick 1981 [1907]: 274)

Of course, one might admit that it is not immediately clear what respect for someone's rational capacities entails in terms of actions, but this further and certainly

important Kantian issue must not disguise the fact that Sidgwick did not consider, or even understand, this possible interpretation of the term 'end' in the humanity formula.

In short, while the Crisp-Irwin and Hurka-Irwin analyses are illuminating, they can also be called into question to some extent. Proponents of both of these views must acknowledge that there are brief passages in which Sidgwick considers views in the respective, more Kantian directions. As already mentioned, he explicitly considers the possibility that partiality towards one's peers might, in practice, be beneficial overall. He also briefly indicates a willingness to consider non-absoluteness concerning benevolence in the chapter on the proof of utilitarianism, in which it is stated that the duty to promote the general good might not be the sole or supreme principle of ethics (Sidgwick 1981 [1907]: 421). These passages are very brief and they are by no means fully explicated. Nonetheless, Sidgwick's editing process was extensive and he might have developed them further had there been sufficient time and energy for an eighth edition of the *Methods*. It is an advantage of my alternative analysis that it is not called into question even by any brief remarks in his texts. Moreover, it seems to provide a *reason* for why he did not see the non-absolute character of Kant's duty of benevolence, rather than simply stating the point.

In other words, the tension between Sidgwick and Kant's views on benevolence, about which Sidgwick claimed that there is partial convergence, lies only partly in a divergence about the proper extent of impartiality or the relevant grounds of partiality, or in the fact that beneficence has the form of a non-absolute rather than an absolute principle for Kant. Another important feature of the divergence lies in the fact that Sidgwick did not accept the full Kantian understanding of what constitutes an end.<sup>29</sup>

### 3. How this interpretation is tied to Sidgwick's worries about metaphysics

A further advantage of this interpretative thesis over others consists in the fact it can provide a clear link to Sidgwick's famous and central concerns about the metaphysical basis of Kant's system, in particular in regard to the relationship between rationality and freedom and in regard to the existence of a Supreme Being. First, Sidgwick had well-known worries about Kant's view of the relationship between rationality and freedom. The most well-known treatment of this is found in his paper 'The Kantian Conception of Free Will' (Sidgwick 1888), which first appeared in one of the early issues of the journal *Mind* and was later partly added as an appendix to the *Methods*. One of his worries regarding rationality and freedom concerns the way in which Kant derives beneficence from the humanity formula. As already noted, Sidgwick finds this argument perplexing, because he thinks that it requires adopting other people's subjective ends determined by irrational impulses. However, as Paul Guyer argues (Guyer 2000: 156ff.), Sidgwick's interpretative mistake consists in the fact that he mistook respect for ends as meaning respect for pure rationality, without any regard for desires and aversions. As Guyer insists, the ends of others that we must respect are the ones that are mutually compatible with the freedom of all others. In other words, respect-worthy ends must meet a respect for rationality test: they may indeed be derived from subjective and non-rational impulses, as long as they are not incompatible with the freedom (of others) to set their own ends. This freedom-related

concern is another facet of Kant's view of ends that Sidgwick fails to be fully cognizant of.

For Kant, the duty to treat humanity as an end in itself is not just due to the fact that humans tend to possess the capacity for rationality; it is also important because the exercise of this rationality is related to their freedom. For Kant, human beings exercise their moral freedom from sensuous inclinations by acting rationally (setting their own ends) in accordance with the moral law. This Kantian relationship between rationality and freedom appears problematic to Sidgwick, as he lays out in 'The Kantian Conception of Free Will'. In this article, he argues that Kant uses the term 'freedom' in incompatible ways without noticing it. On the one hand, he argues, there is the notion of rational freedom in Kant, which means acting in accordance with the moral law, and on the other hand, there is neutral freedom, which means having a choice, including the choice not to act according to the moral law. The upshot of this criticism is that if Kant wants to stick with his strong tie between rationality, freedom and the moral law, then rational and deliberately wrong actions no longer appear to be possible. This would, of course, make ascriptions of moral responsibility for wrongdoing problematic.<sup>30</sup> Sidgwick is not the only author who raises this concern. Reinhold advances a similar worry,<sup>31</sup> and Kant's understanding of freedom and its various forms is still under discussion today<sup>32</sup> – sometimes even in light of Sidgwick's criticism.<sup>33</sup>

In other words, there are two aspects of humanity as an end that Kant usually highlights: the rationality and the freedom of human beings. Sidgwick saw a tension between these two aspects – a tension which he considers at the heart of the Kantian notion of respect for humanity as an end in itself. Representing the bigger picture of Sidgwick's misunderstanding of Kant's view provides a clearer and more comprehensive interpretation of the argument between them.

Secondly, there are some symmetries between freedom and the existence of a Supreme Being in Kant. Both themes play a central role in the *Critique of Pure Reason*, especially in the discussion of the antinomies, and Kant ultimately decides that they are postulates of practical reason. In the discussion of the third antinomy, Kant tries to establish that it is not contradictory to the causal laws of nature to assume that human volition can be a starting point for a chain of events in the world. There is evidence that Sidgwick found the relationship between freedom of the will and a Supreme Being in the antinomies to be problematic. As he argues in his rarely discussed *Lectures on the Philosophy of Kant* – which deal with Kant's first *Critique* – this speculative argument in favour of transcendental freedom of the will seems to depend on the transcendental argument in favour of a necessary Being in order to be fully convincing. In a lecture addressing Kant's third antinomy (on free will) and the fourth antinomy (about a necessary Being) in the *Transcendental Dialectic*, Sidgwick writes:

We have now to observe a flaw in the symmetry of Kant's system. His interest in the question of human freedom has led him to make the freedom of *man* prominent in the discussion of the third antinomy. But the kind of Transcendental Freedom which the argument for the thesis naturally suggests is not human freedom, an uncaused beginning of the various particular series of effects that we attribute to human volition: but an uncaused beginning of

the whole complex process of cosmical change. *Human freedom is certainly not enough*, the effort to find an unconditioned cause to explain Nature can certainly not be satisfied by finding a free causality for human volition. (Sidgwick 1968 [1905]: 172, second emphasis added)

And a few pages later:

[I]t appears to me that the solution of the third antinomy has to be combined with that of the fourth, in order to afford the Speculative Reason that moderate amount of satisfaction which is all that the critical philosophy professes to afford to it. (Sidgwick 1968 [1905]: 176)

In other words, Sidgwick could not see how the argument in favour of transcendental human freedom could work without also accepting the transcendental argument in favour of the existence of a necessary Being at the very beginning of all further contingent free causes. Even though many Kant commentators are sceptical whether the claim that we have to assume a necessary Being to explain the assumption of further free causes in the world is convincing or was even intended by Kant,<sup>34</sup> Sidgwick clearly saw a strong link there – which must have aggravated his reservations about the metaphysical basis or implications of Kant's moral philosophy. It is well known that Sidgwick had deep-rooted concerns with theology.<sup>35</sup> He could not possibly have been content with a view according to which both a Supreme Being and human freedom are necessary. This point becomes even harder to deny if one considers the fact that he took the whole issue of the freedom of the will to be rather irrelevant to ethics in the first place (Sidgwick 1981 [1907]: 57ff.).

Of course, Sidgwick is aware that the theoretical discussion in the first *Critique* is not the ultimate justification of the postulate of human freedom. It is the moral law – rather than speculative reason – that implies, or even requires, the existence of God and the afterlife in order to enable the reconciliation of happiness and moral merit. As Tyler Paytas (2020) argues, both Kant and Sidgwick worried deeply about the relationship between adhering to one's duty and the possibility of reaching happiness, and both saw the possibility of reconciling the tension with an appeal to God and the afterlife. There is in fact a similarity here, and both authors ultimately reject a divine command theory of moral motivation. However, there is also some reason to believe that, regarding the motivational force of the moral law itself,<sup>36</sup> Sidgwick's Kant exegesis was focused on an overly strong reading of Kant. As he points out in both the *Lectures on the Philosophy of Kant* and in the *History of Ethics*, Kant's first *Critique* says that 'without a god and without a world not visible to us now but hoped for, the glorious ideas of morality would be indeed objects of applause and admiration, but not springs of purpose and action'.<sup>37</sup> According to this passage, the moral law itself does not motivate us to act; it is the existence of a God who reconciles duty with happiness. It is interesting that Sidgwick repeatedly quotes this particularly strong passage from the first *Critique*, and does not mention any more moderate passages, especially from the second *Critique*. It can be maintained that, according to the second *Critique*, the existence of God is not literally necessary for the moral law to be persuasive, but his existence is merely implied by the already-given rational persuasiveness of the moral law.<sup>38</sup> Sidgwick may either have not been aware of this change, or it could

be that a mere implication of the existence of God was just as unpalatable to him as the stronger view about the alleged necessity of his existence that he quotes.

It is interesting that Sidgwick primarily discusses the stronger reading, according to which God is literally necessary to motivate moral behaviour, in the last section of the *Methods*, after acknowledging a fundamental dualism of practical reason. According to this unfortunate dualism, it is always rational to act morally, but at the same time also always rational to act egoistically. As he sees it, a Supreme Being who rewards us for obeying the rules of duty would resolve the dualism of practical reason, but he cannot bring himself to confirm the existence of such a Being. In fact, in a famous footnote to this point he goes as far as saying that ‘feeling bound to believe for purposes of practise what I see no ground for holding as a speculative truth’ appears to him as a ‘half-wilful irrationality, committed in a violent access [*sic*] of philosophic despair’ (Sidgwick 1981 [1907]: 507, n. 3). Given the immediate context of the footnote, according to which the idea is that God will punish and reward, this exclamation is probably directed at the stronger and more controversial reading of Kant that goes back to the first *Critique* only. Sidgwick’s desperation is at least partly due to an incomplete and one-sided interpretation of the argument in Kant.

At this point, followers of Sidgwick can nonetheless point out how close he comes to adopting a transcendental argument of his own in the very last paragraph of the *Methods*. Despite his worries about a transcendental argument involving a necessary Supreme Being, he concludes the *Methods* by pointing towards a transcendental argument according to which the ‘necessity to avoid a fundamental contradiction in one chief department of our thought’ might be a sufficient reason to accept the reconciliation of duty and self-interest (Sidgwick 1981 [1907]: 508–9). In the hope of finding a more scientific foundation for moral philosophy, Sidgwick pushes this option away. At the same time, though, he explicitly leaves open the possibility that a better foundation might not be available – perhaps not even in the world of the natural sciences.

As Sidgwick must have seen it, for Kant God and the afterlife provide the possibility of reaching the supreme good of merited happiness, and thereby ensure the much-needed reconciliation of duty and inclination. Arguably, the disagreement over whether this kind of reasoning is a satisfying basis for moral philosophy is a key difference between the two authors that lies behind their more specific disagreements. While it is central to the understanding of Kant’s practical philosophy as a whole, to Sidgwick it appeared to be only a second-best solution.

Kant’s understanding of humanity as an end is strongly tied to his views about the freedom of human beings and his related views about theology. A thesis about the most central difference between Kant’s categorical imperative and Sidgwick’s axioms that points in the direction of these views that Sidgwick certainly did not want to adopt and, in the case of theology, even severely struggled with, should be a competitive candidate for the right view.

#### 4. Conclusions

To summarize, Sidgwick’s partial convergence claim regarding utilitarianism and Kant’s position is based on a restricted version of FUL, as well as on the duty of beneficence and Sidgwick’s beneficence-related reading of the humanity formula. He takes these aspects of Kant’s position as corresponding closely to the axioms of justice and

benevolence respectively – the two axioms are at the same time the foundation of utilitarianism, as understood by Sidgwick.

With regard to the axiom of justice, it can be argued that he was either wrong about or simply did not accept the very large scope that Kant had intended for FUL. For Kant FUL is supposed to explain or even determine whether a maxim is ethically right or wrong, while the canonical formulation of the axiom of justice takes the rightness or wrongness of certain actions as given and simply states that similar cases must be judged in the same way. Sidgwick never accepted this broad import of FUL. Since Kant did not hold Sidgwick's restricted version of FUL, the convergence claim on this matter appears to be without a solid foundation.

With regard to the axiom of benevolence, it is possible that Sidgwick read a duty to promote general happiness by adopting other people's ends into Kant's appeal to humanity. Kant's position was probably not intended in this way, even though a suggestive superlative about adopting other people's ends can certainly be found. An important part of the remaining divergence between Kant and Sidgwick on this point lies in the fact that Sidgwick did not accept the full meaning and relevance of an 'end' in Kant's humanity formula. The reason why other people's ends are relevant for Kant lies in the fact that these persons are ends in themselves in the sense that they possess a capacity for rationality, which is tied to their metaphysical freedom and exercised most fully when adhering to the moral law. For this reason, Kant's duty of beneficence does not require helping others to achieve irrational ends in the sense of violating the moral law, and he also has the resources to exclude prudentially irrational ends: restrictions that Sidgwick apparently did not see.

The claim that Sidgwick's broad reading of beneficence in Kant is related to a failure to understand the meaning of 'humanity as an end' has a number of advantages over other interpretative views. In particular, it shows the source rather than merely the result of his broad understanding. Moreover, it is closely linked to Sidgwick's well-known worries about some of Kant's postulates of practical reason, especially human freedom and the existence of a supreme Being.<sup>39</sup>

## Notes

1. Unless indicated otherwise, English translations of Kant are taken from the Cambridge edition. In some cases (that are indicated), translations are taken from Sidgwick or a translation he used.
2. This is especially true of Kant's theoretical philosophy as demonstrated in Sidgwick's *Lectures on the Philosophy of Kant*.
3. Emphasized e.g. by Parfit (2011: esp. preface and part 17).
4. Apart from Crisp, Hurka and Irwin, whose views will be discussed explicitly later on, exceptions were Bernays (1910), Nakano-Okuno (2007) and Skorupski (2011). In 2020 Routledge brought out a substantial collection on *Kantian and Sidgwickian Ethics* (Paytas and Henning 2020).
5. Langenfus (2000: 103). For the disagreement about the exact number and the most central formulations of Sidgwick's axioms, see Schneewind (1977: 290) and Skelton (2008: 186–202).
6. See Schneewind (1977: 293).
7. 'I have already noticed that his fundamental principle of duty is the "formal" rule of "acting on a maxim that one can will to be a law universal"; which, duly restricted, is an immediate practical corollary from the principle that I first noted in the preceding section' (Sidgwick 1981 [1907]: 385–6). The principle first noted in the preceding section is the principle, or axiom, of justice. The expression 'duly restricted' in this quotation comes with a footnote referring to parts of the *Methods* in which Sidgwick argues against a general prohibition of lying (pp. 317ff.) and against the claim that FUL has to result in such a general prohibition (pp. 486–7).

8. Sidgwick (1981 [1907]: 486–7, 317ff., note on 389). See also Korsgaard's comparison of Kant's and Sidgwick's positions on lying (1988: 17ff.).
9. Ripstein went as far as to write that Sidgwick was 'motivated' by such a teleological interpretation (Ripstein 2009: 13).
10. For a defence of a revised version of the Golden Rule and an analysis of its relation to FUL see Parfit (2011: ch. 14).
11. The exact formulation of FUL in this quotation stems from the 1879 translation by Revd Abbott.
12. However, Sidgwick might not have been entirely wrong. One may argue that FUL, as presented in the *Groundwork*, is not a sufficient or even adequate tool for solving problems with a social or political component. Handling these latter issues adequately might require looking at the Doctrine of Right, which introduces the universal principle of right as a tool for such contexts (Ripstein 2009: 1–29 and 325–88; Varden 2015: 215–21).
13. This conclusion might correspond in some ways to Shaver's interesting argument that Sidgwick might not have to consider Kant as his 'peer' in the sense of the disagreement test for ethical axioms (Shaver 2020). Shaver's idea would alleviate the problem that disagreements among experts are supposed to be a reason for scepticism according to Sidgwick – which might give rise to doubt about some of his very own axioms.
14. Roger Crisp holds the view that the core of the axiom of benevolence is the simple claim 'One ought to aim at the universal good', which he takes to be reflected in Sidgwick's formulation 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'. See Crisp 2015: 120–1.
15. Sidgwick uses the term 'happiness' when referring to passages in which Kant uses the term *Glückseligkeit*.
16. Sidgwick (1988 [1902]: 271–2). The argument given in the *History of Ethics* corresponds in some ways to a two-page footnote in the *Methods* (Sidgwick 1981 [1907]: 389–90).
17. In particular, he refers to Kant, *MM*, 6: 385–6 and *G*, 4: 430.
18. Kant, *G*, 4: 429. Revised Cambridge translation by Gregor and Timmermann (Kant 2012 [1786]): '*So act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means.*'
19. Kant, *G*, 4: 430. The revised Cambridge translation by Gregor and Timmermann (Kant 2012 [1786]) reads as follows: 'For if that representation is to have its full effect in me, the ends of a subject who is an end in himself must, as much as possible, also be my ends.'
20. The German text passage in the Academy Edition also contains the superlative: 'Denn das Subject, welches Zweck an sich selbst ist, dessen Zwecke müssen, wenn jene Vorstellung bei mir alle Wirkung thun soll, auch, so viel möglich, meine Zwecke sein' (*G*, 4: 330, emphasis added).
21. For a further utilitarian interpretation of this passage, see Hare (1993: 3–4).
22. Briefly after his statements of the axiom of benevolence, he reminds the reader of his elaborate analysis of benevolence in common-sense judgements in book III, chapter IV, which seem to fall short of the strongly impartial character of the axiom. Immediately thereafter, he states: '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' (Sidgwick 1981 [1907]: 382).
23. '... a rational impartial spectator can nevermore take any delight in the sight of the uninterrupted prosperity of a being adorned with no feature of a pure and good will, and ... a good will thus appears to constitute the indispensable condition even of the worthiness to be happy' (*G*, 4: 382).
24. This point is related to a lengthy footnote, in which Sidgwick expresses puzzlement about Kant, which will be discussed explicitly further below.
25. See also Timmermann (2005: 250–2). There is an indication that Sidgwick was aware of this in Sidgwick (1988 [1902]: 273).
26. Donagan (1977a: 63, sections 1.5, 7.4). Donagan's more general criticism of Sidgwick is that he did not give philosophical intuitionism a truly complete trial. As Donagan argues, for example, Whewell's intuitionism is much better at handling difficult cases of promise-keeping than Sidgwick acknowledges (Donagan 1977b: 456–9).
27. Shortly before, on p. 390, he writes: 'by an End we commonly mean something to be realised'.

28. See also Reichlin (2008: 124–9), Timmermann (2005: 246, 263).
29. In particular, he could not let go of the idea that an end is something to be produced: ‘...by an End we commonly mean something to be realised...’ (Sidgwick (1981 [1907]): 390).
30. A similar criticism has also been raised by Reinhold, and some Kantians believe that it points to an interesting issue (Allison 1990: 133–6; Timmermann 2007: 164–7). See also Kant’s discussion of radical evil in the *Religion within the Limits of Reason Alone* (esp. 6: 43.12–17).
31. To be found in Bittner and Cramer (1975: 310–24, 252–74).
32. See e.g. Bojanowski (2006), Guyer (2000), Mieth and Rosenthal (2006), Schönecker (2005).
33. Nakano-Okuno (2020: 171–5), Timmermann (2007).
34. Timmermann 2003: 98. Sidgwick must have had the following parts of Kant’s remarks on the thesis of free will in the third antinomy in mind:

We have really established this necessity of a first beginning of a series of appearances from freedom only to the extent that this is required to make comprehensible an origin of the world, since one can take all the subsequent states to be a result of mere natural laws. But because the faculty of beginning a series in time entirely on its own is thereby proved (though no insight into it is achieved), now we are permitted also to allow that in the course of the world different series may begin on their own as far as their causality is concerned, and to ascribe to the substances in those series the faculty of acting from freedom. . . . The confirmation of the need of reason to appeal to a first beginning from freedom in the series of natural causes is clearly and visibly evident from the fact that (with the exception of the Epicurean school) all the philosophers of antiquity saw themselves as obliged to assume a *first mover* . . . (CPR, A448–9/B476–7, Cambridge translation).

In the resolution to the third antinomy this issue receives no further mention.

35. Sidgwick 1870; cf. Schultz (2004: chs 1–2).
36. For an extensive treatment of moral motivation in Kant and Sidgwick see Phillips (2020).
37. CPR, A812–13/B840–1; quoted in Sidgwick’s *Lectures on the Philosophy of Kant* (Sidgwick 1968 [1905]: 179) and partly also in his *History of Ethics* (Sidgwick 1988 [1902]: 276).
38. Förster (1998: 344–9). In particular, the passage on the existence of god as a postulate of pure practical reason contains some of these more modest claims (CPR, 5: 254ff.).
39. This article would never have been written without the work of Tom Hurka. It benefited from exchanges with Christoph Horn, Jens Timmermann and Helga Varden, as well as from participants of the ISUS conference in 2018, including Roger Crisp and Paul Hurley. Special thanks also go to Rachel Bryant, Shelley Weinberg, Fridolin Neumann and the anonymous reviewers of this journal.

## References

- Allison, Henry E. (1990) *Kant’s Theory of Freedom*. Cambridge: Cambridge University Press.
- Bernays, Paul (1910) *Das Moralprinzip bei Sidgwick und Kant*. Göttingen: Abhandlungen der Fries’schen Schule.
- Bittner, Rüdiger and Konrad Cramer (eds) (1975) *Materialien zu Kants ‘Kritik der praktischen Vernunft’*. Frankfurt/Main: Suhrkamp.
- Bojanowski, Jochen (2006) *Kants Theorie der Freiheit*. Berlin: De Gruyter.
- Crisp, Roger (2015) *The Cosmos of Duty. Henry Sidgwick’s Methods of Ethics*. Oxford: Oxford University Press.
- Donagan, Alan (1977a) *The Theory of Morality*. Chicago: University of Chicago Press.
- (1977b) ‘Sidgwick and Whewellian Intuitionism: Some Enigmas’. *Canadian Journal of Philosophy*, 7, 447–65.
- Förster, Eckart (1998) ‘Die Wandlungen in Kants Gotteslehre’, *Zeitschrift für philosophische Forschung*, 52, 341–62.
- Guyer, Paul (2000) *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press.
- Hare, Richard (1993) ‘Could Kant have been a Utilitarian?’ *Utilitas*, 5, 1–16.
- Hurka, Thomas (2014) ‘Sidgwick on Consequentialism and Deontology: A Critique’. *Utilitas*, 26, 129–52.



- Irwin, Terence (2009) *The Development of Ethics*, vol. 3: *From Kant to Rawls*. Oxford: Oxford University Press.
- Kant, Immanuel (2012 [1786]) *Groundwork to the Metaphysics of Morals*. Ed. and trans. Mary Gregor and Jens Timmermann. Rev. ed. Cambridge: Cambridge University Press.
- Korsgaard, Christine (1988) 'Two Arguments Against Lying [I. Kant and H. Sidgwick]'. *Argumentation*, 2, 27–49.
- (2009) 'Kant's Formula of Humanity'. *Kant-Studien*, 77, 183–202.
- Langenbusch, William (2000) 'The Impossibility of Sidgwick's Proof'. *History of Philosophy Quarterly*, 17, 99–120.
- Mieth, Corinna and Jacob Rosenthal (2006) 'Freedom must be Presupposed as a Property of the Will of all Rational Beings'. In Christoph Horn and Dieter Schönecker (eds), *Groundwork for the Metaphysics of Morals. New Essays* (Berlin: De Gruyter), 247–84.
- Mill, John Stuart (1985 [1833]) *Utilitarianism*. In John M. Robson (ed.), *Collected Works*, vol. 10. Toronto: University of Toronto Press.
- Nakano-Okuno, Mariko (2007) 'Sidgwick and Kant: On the So-Called "Discrepancies" between Utilitarianism and Kantian Ethics'. In Placido Bucolo, Roger Crisp and Bart Schultz (eds), *Proceedings of the World Congress on Henry Sidgwick: Happiness and Religion* (Catania: Università degli Studi di Catania), 260–334.
- (2020) 'Kant and Sidgwick on the Freedom of Will, Morality, and Responsibility'. In Tyler Paytas and Tim Henning (eds), *Kantian and Sidgwickian Ethics: The Cosmos of Duty above and the Moral Law within* (New York: Routledge), 163–83.
- Parfit, Derek (2011) *On What Matters*, vol. 1. Oxford: Oxford University Press.
- Paytas, Tyler (2020) 'Beneficent Governor of the Cosmos: Kant and Sidgwick on the Moral Necessity of God'. In Tyler Paytas and Tim Henning (eds), *Kantian and Sidgwickian Ethics: The Cosmos of Duty above and the Moral Law within* (New York: Routledge), 210–44.
- Paytas, Tyler and Tim Henning (eds) (2020) *Kantian and Sidgwickian Ethics: The Cosmos of Duty above and the Moral Law within*. New York: Routledge.
- Phillips, David (2020) 'Sidgwick's Kantian Account of Moral Motivation'. In Tyler Paytas and Tim Henning (eds), *Kantian and Sidgwickian Ethics: The Cosmos of Duty above and the Moral Law within* (New York: Routledge), 84–104.
- Reichlin, Massimo (2008) 'Ordinary Moral Knowledge and Philosophical Ethics in Sidgwick und Kant'. *Etica e Politica. Revista di Filosofia On-line*, 10, 109–36.
- Ripstein, Arthur (2009) *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press.
- Schneewind, Jerome (1977) *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford: Oxford University Press.
- Schönecker, Dieter (2005) *Kants Begriff transzendentaler und praktischer Freiheit. Eine entwicklungsgeschichtliche Studie*. Berlin: De Gruyter.
- Schultz, Bart (2004) *Henry Sidgwick – Eye of the Universe. An Intellectual Biography*. Cambridge: Cambridge University Press.
- Shaver, Robert (2020) 'Peerless. Sidgwick, Kant, and Peer Disagreement'. In Tyler Paytas and Tim Henning (eds.), *Kantian and Sidgwickian Ethics: The Cosmos of Duty Above and the Moral Law Within* (New York: Routledge), 118–37.
- Sidgwick, Henry (1870) *The Ethics of Conformity and Subscription*. London: Williams & Norgate.
- (1888) 'The Kantian Conception of Free-Will'. *Mind*, 13, 405–12.
- (1968 [1905]) *Lectures on the Philosophy of Kant*. London: MacMillan.
- (1981 [1907]) *The Methods of Ethics*. 7th edn. Indianapolis: Hackett.
- (1988 [1902]) *Outlines of the History of Ethics for English Readers*. 5th edn. Indianapolis: Hackett.
- Skelton, Anthony (2008) 'Sidgwick's Philosophical Intuitions'. *Etica e Politica. Revista di Filosofia On-line*, 10, 185–209.
- (2020) 'Practical Ethics in Sidgwick and Kant'. In Tyler Paytas and Tim Henning (eds), *Kantian and Sidgwickian Ethics: The Cosmos of Duty above and the Moral Law within* (New York: Routledge), 13–39.
- Skorupski, John (2011) 'Impartiality: Kant, Mill, Sidgwick'. In Placido Bucolo, Roger Crisp and Bart Schultz (eds), *Henry Sidgwick: Ethics, Psychics, Politics* (Catania: Università degli Studi di Catania), 634–79.

- Timmermann, Jens (2003) *Sittengesetz und Freiheit: Untersuchungen zu Immanuel Kants Theorie des freien Willens*. Berlin: De Gruyter.
- (2005) 'Why Kant Could Not have Been a Utilitarian'. *Utilitas*, 17, 243–64.
- (2007) 'Freedom and Moral Failure: Reinhold and Sidgwick'. In Jens Timmermann (ed.), *Kant's 'Groundwork to the Metaphysics of Morals'* (Cambridge: Cambridge University Press), appendix E, 164–7.
- Varden, Helga (2015) 'Justice as Freedom'. In Guttorm Fløistad (ed.), *Philosophie de la justice/Philosophy of Justice*, vol. 12 (Amsterdam: Springer), 157–76.

---

**Cite this article:** Dufner, A. (2022). The Role of Kant in Sidgwick's Classical Utilitarianism: Two Self-Evident Axioms and the Partial Convergence between Kantianism and Utilitarianism. *Kantian Review* 27, 345–362. <https://doi.org/10.1017/S1369415422000267>