

## ***Ordinary Moral Knowledge and Philosophical Ethics in Sidgwick and Kant***

Massimo Reichlin

Facoltà di Filosofia

Università Vita-Salute San Raffaele

reichlin.massimo@hsr.it

### **ABSTRACT**

Sidgwick considered Kant as one of his masters. However, he never devoted any systematic attention to Kant's ethical theory; moreover, in *The Methods of Ethics* he concluded that Kantian ethics is inadequate to guide moral life. I review Sidgwick's references to Kant in order to show that – along with basic differences – there are significant similarities in the main project of the two philosophers; and I suggest that, should Sidgwick have deepened his understanding of Kant, he might have realised that Kantian ethics offered a somewhat different way to accomplish the philosophical project he was interested in, that is, the systematisation of the morality of common sense through the establishment of certain moral axioms. I also suggest that Sidgwick's misunderstanding of the "formula of humanity" is at the heart of his final dismissal of Kant's ethics and that deepening his understanding of Kant might have led Sidgwick to revise his views on the rationality of egoism, thereby opening the possibility to solve the dualism of practical reason. Finally, I offer some speculations on the reasons why Sidgwick never attempted a thorough confrontation with Kant, suggesting that both his distaste for Kant's metaphysics and his Millian utilitarian bias deterred him from it.

### ***1. A Puzzling Relationship***

In the famous autobiographical note added to the sixth edition of *The Methods of Ethics*, Sidgwick declares Kant one of «my masters» (ME 7, p. xviii)<sup>1</sup> alongside with Mill; he describes his ethical project as a struggle «to assimilate Mill and Kant» (Ibid.), and says that his final reconciliation of utilitari-

---

<sup>1</sup> I will use the abbreviations ME 1 and ME 7 to refer to *The Methods of Ethics*, 1<sup>st</sup> edition (1874) and 7<sup>th</sup> edition (1907), both quoted from the Thoemmes reprint (Bristol 1996); OHE to refer to the *Outlines of the History of Ethics* (1886), quoted from the Hackett reprint (Indianapolis 1988); HSM to refer to *Henry Sidgwick: A Memoir*, by A. Sidgwick, E. M. Sidgwick (Macmillan, London 1906); and G to refer to Kant's *Grundlegung*, quoted from the English translation *Fundamental Principles of the Metaphysic of Morals* in I. Kant, *The Critique of Pure Reason, The Critique of Practical Reason and Other Ethical Treatise and The Critique of Judgment*, Encyclopaedia Britannica, Chicago 1952.

anism and intuitionism was reached in part through the realisation of the «perfect harmony» (ME 7, p. xx) between the Kantian principle and the utilitarian one. This late reconstruction is confirmed by a short paper written three years after the first publication of *The Methods*, where he already pointed out the centrality of the Kantian element in his ethical viewpoint: «I identify a modification of Kantism with the missing rational basis of the ethical utilitarianism of Bentham, as expounded by J. S. Mill»<sup>2</sup>. Given the emphasis with which he includes Kant among the main inspirers of his project, it comes as a surprise that Sidgwick never set out, in his long career as a philosopher and a university teacher, to address Kant's ethics with any detailed attention. Indeed, as noted by M. G. Singer, «his failure to come to terms adequately with Kant's ethics may be the most difficult thing to understand about his approach to ethics and the most serious deficiency in it»<sup>3</sup>. Such failure is particularly puzzling since: i) Sidgwick taught ethics constantly from the '60s to his death; ii) he devoted considerable attention to other influential moral philosophers, such as Bentham, Martineau, Grote, Green, Spencer and Stephen; iii) he published a number of essays on Kant's metaphysics and epistemology<sup>4</sup>, and taught an entire course on the Critique of Pure Reason and the Prolegomena<sup>5</sup>. Why then did Sidgwick — apart from some passages in *The Methods* — never devote to Kant's ethics more than the few pages contained in the *Outlines of the History of Ethics*?<sup>6</sup>

The pages in the *Outlines* — it must be added — are indeed deeply inadequate, considering that Sidgwick could read German, that he was a very remarkable historian of philosophy, and that his Lectures on the Philosophy of Kant are profound and extensive. The brief summary contained in the *Out-*

---

<sup>2</sup> H. Sidgwick, *Mr Barratt on 'The Suppression of Egoism'* (1877), in *Essays on Ethics and Method*, ed. by M. G. Singer, Oxford University Press, Oxford 2000, pp. 27-28, here at p. 27. One earlier testimony is a 1866 letter in which he declares Kant's phraseology «quite a revelation to me», and, after having censured German Idealism as «a monstrous mistake», he concludes that «we must go back to Kant and begin again from him. Not that I feel prepared to call myself a Kantian, but I shall always look on him as one of my teachers» (HSM, p. 151).

<sup>3</sup> M. G. Singer, *A Note on the Content*, *ibid.*, p. xlii.

<sup>4</sup> These are: *The So-Called Idealism of Kant*, «Mind» 1879; *Kant's Refutation of Idealism*, «Mind» 1880; *A Criticism of the Critical Philosophy*, part I and II, «Mind» 1883; and *Kant's View of Mathematical Premises and Reasonings*, parts I and II, «Mind» 1883.

<sup>5</sup> *Lectures on the Philosophy of Kant and Other Philosophical Lectures* (1905), Thoemmes Press, Bristol 1996.

<sup>6</sup> The 1888 essay on *The Kantian Conception of Free Will* might here be added, though it in fact discusses the metaphysical underpinnings of Kant's conception, rather than his ethical theory *qua talis*.

*lines*, on the contrary, lends itself to criticism on several grounds: i) it never mentions the *Metaphysics of Morals* (though it implicitly refers to it in various passages)<sup>7</sup>; ii) it does not recall the central doctrine of the moral law as a fact of reason, stated in the second *Critique*; iii) it clearly misunderstands (OHE, pp. 274-5) the significance of the second formula of the categorical imperative (the “formula of humanity”, more on which will be said later); iv) it never refers to such central ideas, in the Kantian perspective, as those of the autonomy of the will and of a universal kingdom of ends; v) it attributes to Kant the view that the belief in a moral government of the world is necessary to motivate moral action — a view Kant holds in the first *Critique*, from which Sidgwick quotes (OHE, p. 276), but repudiates in all his ethical treatises<sup>8</sup>.

The lack of a direct confrontation with Kant’s ethical thought clearly has to do with Sidgwick’s classification of Kant as an intuitionist, as well as with his failure to acknowledge Kant’s as a distinctive method of ethics<sup>9</sup>. This is again very surprising, since Kant’s moral philosophy is doubtless very different from those of the British moralists, from Cudworth to Whewell, that are the paradigmatic exponents of the polemic target constructed by Sidgwick under the heading of “intuitionism” and discussed in Book III of *The Methods*. True, it could be argued that Sidgwick did show a certain awareness of the fact that Kant is not simply a member of the intuitional school; in fact, he writes that we can find «distinct traces of Kantian influence in Whewell and other writers of the intuitional school» (OHE, p. 271), and cautiously speaks of a particular affinity of Kant with Price (*Ibid.*; both emphases are added): these expressions may suggest that perhaps Sidgwick was not willing to rank Kant among the members of the intuitional school tout court. However he does seem to conflate Kantian ethics and intuitionism throughout

---

<sup>7</sup> E.g. OHE, pp. 274-5. *The Doctrine of Virtue* is instead quoted repeatedly in *The Methods*: see for example ME 7, III, 9, note 1; ME 7, III, 13, concluding Note and note 15.

<sup>8</sup> The same passage concerning the «glorious ideas of morality» as «objects of applause and admiration, but not springs of purpose and action» is quoted as representing the definitive Kantian position in the paper read to the Synthetic Society on February 25, 1898 (*On the Nature of the Evidence for Theism*, in HSM, pp. 600-608, at p. 605). The critical judgment on the treatment of Kant in OHE may be partly qualified by noting that Sidgwick’s work is intended for English readers; for the modern period, in fact, it is «mainly confined to English ethics, and only deals with foreign ethical systems in a subordinate way, as sources of influence on English thought» (OHE, p. v). Not by chance, the last paragraph of the work, where the pages on Kant appear, bears the title “German influence on English ethics”.

<sup>9</sup> As lamented by J. Rawls, *Kantian Constructivism in Moral Theory*, «Journal of Philosophy», 77, 1980, pp. 515-572, at p. 556.

*The Methods*, and explicitly declares Kant an intuitionist in at least one passage (ME 7, p. 366)<sup>10</sup>.

Why Sidgwick never devoted more of his scholarly attention to Kant's moral philosophy is very difficult to investigate; I will be offering my tentative speculations in § 5. What may perhaps be more confidently said is that, should Sidgwick have deepened his understanding of Kant, he might have found that: i) Kant's ethics is not as inadequate to the task of giving «complete guidance» (ME 7, p. xix) to our moral life, as he finally came to believe; ii) Kant's project is much more similar to Sidgwick's than the latter thought, with particular reference to the relationship between ordinary moral knowledge and philosophical ethics. In fact, Kant's system offers a way of elevating the Morality of Common Sense into a system of philosophical ethics that is different both from the attempts of traditional “intuitional” moralists and from Sidgwick's problematic incorporation of that morality within the utilitarian system. I will not venture to say that, should Sidgwick have understood Kant more in depth, he would have become a Kantian; what the following pages are going to suggest is rather that he would have had to choose among two alternative ways in which to accomplish his own main project, that is, to provide a philosophical defence of the morality of common sense. And — for reasons that will emerge in due course — it is not wholly certain that he would have chosen the utilitarian one.

## 2. *The Relationship between Ordinary Moral Knowledge and Philosophical Ethics*

The project of *The Methods* is deliberately Socratic: through «impartial reflection on current opinion» (ME 7, p. xx), Sidgwick tries to bring consistency to the Morality of Common Sense of his era, just as Aristotle had done for the morality of fifth century B.C. Athens. Sidgwick clearly does not accept Common Sense as a definitive authority: he claims that «the aim of a philosopher, as such, [is] to do somewhat more than define and formulate the common opinion of mankind. His function is to tell men what they ought to think, rather than what they do think» (ME 7, p. 373)<sup>11</sup>. The aim of moral

---

<sup>10</sup> Another passage explicitly including Kant among the «intuitive moralists» occurred in ME 1, p. 303. The passage is modified in the following editions.

<sup>11</sup> The point is perhaps most clearly stated in a later essay: «though I have always been anxious to ascertain and disposed to respect the verdict of Common Sense in any ethical dispute, I cannot profess to regard it as final and indisputable: I cannot profess to hold that it is impossible for me ever to be right on an ethical point on which an overwhelming

philosophy is to correct and rationalise the morality of common sense in view of a more systematic construction: this can be effected by confronting it with genuine intuitions such as the Kantian principle of impartiality, the utilitarian principle of universal benevolence and the principle of rational egoism. The upshot of this procedure is well known: the alleged opposition between intuitionism and utilitarianism is in fact due to a misunderstanding, while a deeper opposition lingers between morality and rational egoism, i.e., the famous dualism of practical reason.

I think it important to stress the analogies that this Socratic project bears to the procedure followed by Kant, particularly in the *Grundlegung* — presumably a book very well known to Sidgwick. What Kant is here trying to do is in fact, first, to use the analytic method (see Preface) to extract, from what he calls the «common rational knowledge of morality» (“gemeine sittlichen Vernunftkenntnis”, note that for Kant this basic knowledge is already in itself rational), the very idea of duty, thus moving to a philosophical knowledge of morality (Section I); second, to search the principles of this philosophical morality, passing from «popular moral philosophy» to the «metaphysic of morals» (Section II); third, in a synthetic vein, to show that morality is not a «creation of the brain» but a reality, thus passing from the metaphysic of morals to the «critique of pure practical reason» (Section III). In other words, Kant is in fact assuming that morality exists, and that it is just like ordinary people conceive it; what he tries to do is to elucidate the concept of it that is implicit in ordinary moral knowledge, before trying to vindicate it rationally, by showing how pure reason can be practical.

The method employed by Kant is in fact different from Sidgwick's<sup>12</sup>. Kant does not provide a large review of the morality of common sense, in order to show both its strengths and its difficulties, as done by Sidgwick; he starts with what he considers the implicit understanding of common sense, relative to what is unconditionally good — i.e., the good will — and tries to bring out what is contained in this idea: that is, the idea of being subject to duty, which in turn means being subject to a law of reason that objectively and interpersonally constrains the satisfaction of individual inclinations and the pursuit of individual and collective happiness. This leads him to single out the categorical imperative, in the formula of universal law, as the fundamental principle of morality, not as a principle needed to systematise the plural-

---

majority is clearly opposed to me» (H. Sidgwick, *Some Fundamental Ethical Controversies* [1889], in *Essays on Ethics and Method*, pp. 35-46, here at p. 35).

<sup>12</sup> This contrast is very much emphasized by A. W. Wood, *Kantian Ethics*, Cambridge University Press, Cambridge 2007, pp. 43-65.

ity of moral imperatives acknowledged by ordinary moral consciousness, nor as one generated by a theorist's speculations, but as the principle that is ordinarily — though implicitly — used by common men; these, of course, do not conceive the principle in such an abstract and universal form as presented by Kant,

yet they always have it really before their eyes and use it as the standard of their decision. Here it would be easy to show how, with this compass in hand, men are well able to distinguish, in every case that occurs, what is good, what bad, conformably to duty or inconsistent with it, if, without in the least teaching them anything new, we only, like Socrates, direct their attention to the principle they themselves employ; and that, therefore, we do not need science and philosophy to know what we should do to be honest and good, yea, even wise and virtuous. Indeed we might well have conjectured beforehand that the knowledge of what every man is bound to do, and therefore also to know, would be within the reach of every man, even the commonest (G, pp. 260-261)<sup>13</sup>.

In other words, starting from the idea, supposedly acknowledged by common sense, that the value of an action done from duty stems from its principle of willing and not from the object it pursues, Kant comes to the conclusion that the principle of «the moral knowledge of common human reason» (G, p. 260) — that is, the method used by ordinary men in reaching moral conclusions — is the one that tests moral maxims by asking whether they are the product of any inclination or are apt to become principles of a universal legislation.

On the contrary, Sidgwick embarks on a large review of the morality of common sense, in order to show that it does not provide a systematic construction, since many of its precepts are too vaguely stated and often at odds to one another. He then proceeds to extract from that large discussion three

---

<sup>13</sup> As is well known, Rousseau's influence was decisive for the development of Kant's ethical views on this point. In his *Notes* on his own copy of the *Observation on the Feeling of the Beautiful and the Sublime*, Kant already wrote (in 1765): «I am myself a researcher by inclination. I feel the whole thirst for knowledge and the curious unrest to get further on, or also the satisfaction in every acquisition. There was a time when I believed that this alone could make the honor of humanity and I despised the rabble that knows nothing. Rousseau set me right. This dazzling superiority vanishes, I learn to honor man and I would find myself more useless than the common labourer if I did not believe that this observation would impart to all else a value to restore the rights of mankind» (quoted in J. B. Schneewind, *The Invention of Autonomy. A History of Modern Moral Philosophy*, Cambridge University Press, Cambridge 1998, pp. 488-489).

or four immediately evident and more formal principles<sup>14</sup>, that are genuinely axiomatic — that is, that are self-evident upon reflection for every rational individual — and that he finds partly in the works of past moralists and partly implicit in the ordinary way of dealing with moral questions. Armed with these principles, he goes on to show that they are able to provide the required systematisation of the morality of common sense, within the context of a revised utilitarian theory<sup>15</sup>.

It is not the case that Kant, unlike Sidgwick, meant to withhold initial trust from the main normative principles that ordinary moral knowledge assumes to be true. On the contrary, in the *Grundlegung*, he seems to consider the fact that normative conclusions that we generally trust and assume to be true — e.g. that suicide is morally wrong, that we should not make promises with the intention not to keep them, and so on — can be derived from abstract formulations such as those of the various formulas of the categorical imperative as confirming that these formulas are in fact implied in the ordinary processes of moral thinking. Moreover, Kant's project in the later *Metaphysics of Morals*, which is his explicit attempt to construct a system of moral duties, is precisely to show the capacity of his philosophical system to vindicate most of the particular moral conclusions that were commonsense in his days and for his cultural and religious milieu. Kant's attitude towards the morality of common sense is in fact even more positive than Sidgwick's; he is however at least as sceptical as Sidgwick about the previous philosophers' attempt to provide a philosophical account of such ordinary knowledge. He therefore believes that the first philosophical task is to investigate the formal processes that are embedded in ordinary moral thinking, in order to establish its fundamental principle.

It might be observed that the difference between the two philosophers lies simply in the order in which the different steps are accomplished: Sidgwick just postpones the search for more formal principles after the review of common sense morality's material principles, but he nonetheless concurs with Kant in stressing the need for such principles in order to accomplish the systematisation that is philosophy's main task. This is not quite true, for it

---

<sup>14</sup> The question is notoriously controversial as to how many really self-evident principles Sidgwick is willing to accept: the figures range from three (H. Rashdall, *A Theory of the Good and the Right*, Clarendon Press, Oxford 1907, vol. I, p. 147) to eight (W. K. Frankena, *Henry Sidgwick*, in *Encyclopedia of Morals*, ed. by V. Ferm, Philosophical Library, New York 1956), with four perhaps being the most reasonable answer.

<sup>15</sup> Of course, he also shows the reasons why someone might not want to accept the principle of universal beneficence, thus confining himself to the narrower view of individual hedonism; this is what triggers the problem of the «dualism of practical reason».

seems to miss one important point: the fact that for Kant the formality of the fundamental principle is strictly connected with the formality of his conception of moral obligation. That is, while both philosophers clearly accept the idea that moral imperatives are dictates of rationality, Sidgwick assumes that they have to do with bringing about some good: in particular, the “ultimate good on the whole” must be identified with «what as a rational being I should desire and seek to realise, assuming myself to have an equal concern for all existence» (ME 7, p. 112). On this view, that choice is practically most reasonable which brings about the greatest good, however defined<sup>16</sup>. Kant, on the other hand, believes that, from the moral point of view, only a good will is unconditionally good, and it is good on account of its principle of willing, not on account of its object. Therefore, a deep difference between the two projects of vindicating philosophically the morality of common sense lies in the different conception of goodness that they assume as implicit in ordinary consciousness: on the one side, the idea of goodness as some state of affairs that can be produced — and that is eventually identified by Sidgwick with some pleasurable state of consciousness; on the other hand, the idea of goodness as good will, that is, the disposition to act only on maxims that may be conceived and willed as universal laws.

This basic difference should not prevent us from stressing the affinities between the two philosophical projects of founding a scientific ethics by giving philosophical systematisation and vindication to the morality of common sense. The Kantian way of proceeding is in fact doubly consonant to Sidgwick’s mind: on the one hand, it shares its Socratic bent, by according serious philosophical relevance to the ordinary processes of moral knowledge; on the other hand, it clearly denies the sufficiency of ordinary moral knowledge for a genuine philosophical system of morality. In fact, while acknowledging that the common intellect may often surpass the philosopher in the practical domain, Kant declares that, lacking a precise philosophical determination of the principle of morality, it is difficult for ordinary wisdom to outdo the inclinations; that is, it is practically difficult to overcome the natural «disposition to argue against these strict laws of duty and to question their validity»,

---

<sup>16</sup> Both the axiom of prudence and that of beneficence are formulated by Sidgwick, by using the formal notion of “good”; they then receive “material” content through the demonstration that «Desirable Consciousness» is the only thing that can be considered as ultimate Good (ME 7, p. 397). But these formal readings would not escape Kant’s objection that, by prioritising the good over the right, impure and heteronomous elements are introduced in the very concept of morality.



with a view, wherever possible, to «make them more accordant with our wishes and inclinations» (G, p. 261)<sup>17</sup>.

On this account, it seems reasonable to say that Sidgwick might have found in Kant one alternative way of developing precisely what he wanted: the recognition of both the importance and insufficiency of ordinary moral knowledge, along with the philosophical effort to find a fundamental principle to systematize it<sup>18</sup>. The importance of the philosophical effort to bring systematic order into the ordinary moral knowledge of humanity is especially strong in both authors. As for Kant, his passion — or rather his obsession — for systematic philosophy, and for the critical foundation of the system of science, is too well known to be worth stressing. Let me just recall his observation that the innocence of practical wisdom is easily seduced; so that, «when practical reason cultivates itself, there insensibly arises in it a dialectic which forces it to seek aid in philosophy, just as happens to it in its theoretic use; and in this case, therefore, as well as in the other, it will find rest nowhere but in a thorough critical examination of our reason» (Ibid.).

As for Sidgwick, it is perhaps enough to quote a very strong passage from the I edition of *The Methods*, in which his quite rationalistic pretensions, as far as the foundation of morals are concerned, are very well voiced: «conduct appears to us irrational, or at least imperfectly rational, not only if the maxims upon which it is professedly based conflict with and contradict one another, but also if they cannot be bound together and firmly concatenated by means of some one fundamental principle. For practical reason does not seem to be thoroughly realised until a perfect order, harmony, and unity of system is introduced into all our actions»<sup>19</sup>. Doubtless, it is this epistemological ideal that renders Sidgwick's acknowledgment of the unsolvable dualism of practical reason so dramatic. He is notoriously emphatic about the uneasiness caused in him by the lack of a final foundation, and even considers a modification of his epistemology in order to close the gap between duty and happiness. After mentioning his previous willingness to accept a provisional postu-

---

<sup>17</sup> Schneewind appropriately stressed Rousseau's influence on this point as well; in fact, it is because human nature has been profoundly corrupted by its historical development, that feelings cannot by themselves reliably guide our action, and we need reason (*The Invention of Autonomy*, p. 504).

<sup>18</sup> Borrowing Schneewind's apt terms, we could say that Kant accepts the "dependence argument", but offers a "systematization argument" different from Sidgwick's (cf. *Sidgwick's Ethics and Victorian Moral Philosophy*, pp. 279-285 and 331-336).

<sup>19</sup> ME 1, p. 26 (the passage no longer appears after the second edition). On the strictly axiomatic character of Sidgwick's epistemic model, see J. Deigh, *Sidgwick's Epistemology*, «Utilitas», 19, 2007, pp. 435-446.

lation of immortality, in Kant's wake, while searching for the empirical evidence of an afterlife, he continues:

If I decide that this search is a failure, shall I finally and decisively make this postulate? Can I consistently with my whole view of truth and the methods of its attainment? And if I answer "no" to each of these questions, have I any ethical system at all? And if not, can I continue to be Professor and absorb myself in the mere erudition of the subject [...]. I am nearly forty-nine, and I do not find a taste for the old clothes of opinions growing on me (HSM, p. 467).

To decisively make the postulate would in fact be to accept the epistemology set out in the very last sentence of *The Methods*, according to which we would be justified in accepting as universally true propositions that are founded on our strong disposition to affirm them, together with their being indispensable to the systematic coherence of our beliefs. It is now a generally accepted view that Sidgwick never brought himself to make this postulate<sup>20</sup>, even though he seems to waver significantly on this issue throughout his life: two years before his death, he in fact still seemed to consider such postulation a very serious possibility. The conclusion of his conference on theism is paradigmatic of his lingering doubts:

It seems to me, then, that if we are led to accept Theism as being, more than any other view of the Universe, consistent with, and calculated to impart a clear consistency to, the whole body of what we commonly agree to take for knowledge — including knowledge of right and wrong — we accept it on grounds analogous to those on which important scientific conclusions have been accepted; and that, even though we are unable to add the increase of certitude derivable from verified predictions, we may still attain a sufficient strength of reasoned conviction to justify us in calling our conclusion a "working philosophy" (HSM, pp. 607-608)<sup>21</sup>.

---

<sup>20</sup> See for example the discussion in J. L. Mackie, *Sidgwick's Pessimism*, «Philosophical Quarterly», 105, 1976, pp. 317-327; repr. in B. Schultz (ed.), *Essays on Henry Sidgwick*, Cambridge University Press, Cambridge 1992, pp. 163-174.

<sup>21</sup> In the paper on *Authority, Scientific and Theological*, presented to the *Synthetic Society on February 24, 1899*, he returns on Kant's practical postulate again with a somewhat more sceptical attitude; in fact, «for most minds a belief recognised as assumed merely for practice is liable to decline to a belief of which there is an intellectual need, but a need that does not carry with it its own satisfaction: the satisfaction of the need has to be obtained, if at all, through some other line of thought» (HSM, pp. 608-615, here at p. 615).

In any case, it must be stressed that his aspirations to a philosophical foundation were such that his incapacity to solve this problem amounts for him to acknowledging a radical failure; as he confesses in 1887 note, «the recognised failure of my efforts to obtain evidence of immortality affects me not as a Man but as a Moralist» (HSM, pp. 471-472). In fact, while he does not feel anxious about the fact that, somehow or other, morality is going to get on, he sees clearly that, as a philosopher, his «special business is not to maintain morality somehow, but to establish it logically as a reasoned system; and I have declared and published that this cannot be done, if we are limited to merely mundane sanctions, owing to the inevitable divergence, in this imperfect world, between the individual's Duty and his Happiness» (HSM, p. 472)<sup>22</sup>.

### *3. Kant's search for the fundamental principle and Sidgwick's misguided critique*

Notwithstanding the affinity of the two authors' philosophical projects and some sparse similarities that will be noted in a while, Kant's moral deontology is doubtless very far from Sidgwick's utilitarianism «on an intuitional basis». I have already mentioned the fact that Kant assumes an anti-teleological notion of goodness at the very start of his philosophical inquiry on morality. Differences increase if we look at the development of Kant's fundamental principle, as spelled out in the second section of the *Grundlegung*. Here he distinguishes between hypothetical and categorical imperatives, and declares that moral imperatives command categorically. This implies that there are true answers to moral questions, that such answers can be found through rational reflection and that they are found by keeping such reflection "pure", that is, by discarding any empirical element, including of course the inclination towards certain objects.

Sidgwick does concur on part of this perspective. For one thing, he accepts a form of moral cognitivism, declaring, against the Humean view of reason shared by most part of the empiricist tradition, that «what ought to be is a possible object of knowledge» (ME 7, p. 33)<sup>23</sup>. And he intends this in the meaning of an objective rationalism, claiming that ethical judgments are

---

<sup>22</sup> This situation even led him to seriously wonder whether he had to resign his position as a teacher of ethics; see the 1888 letter in HSM, pp. 484-486.

<sup>23</sup> In the first edition he explicitly accepts the common view according to which «in saying that Reason apprehends moral distinctions, it would seem that no more is usually meant than that there is such a thing as moral truth and error; that two conflicting judgments as to what ought to be done cannot both be true and sound» (ME 1, p. 23).

«dictates» or «precepts» of reason, so that «what I judge ought to be must, unless I am in error, be similarly judged by all rational beings who judge truly of the matter» (Ibid.). Sidgwick's basic problem with the so called «dualism of practical reason» is in fact that, should we abandon the project of completely rationalising morality, cases of conflict between self-interest and duty would show practical reason as «divided against itself» and unable «to be a motive on either side»: the conflict would thus be adjudicated by the prevalence of one or the other group of non-rational impulses, and we should «lapse to the position which many utilitarians since Hume have avowedly held — that ultimate ends are determined by feeling, not by reason»<sup>24</sup>.

This being so, it is also clear that Sidgwick does not accept only hypothetical imperatives, for he believes that reason also has a role in the determination of the ends, not only of the means. In fact, he distinguishes between “moral” and “prudential” judgments, meaning a distinction between «cognition or judgments of duty» (ME 7, p. 25) and «cognition or judgments as to what “is right” or “ought to be done” in view of the agent's private interest or happiness» (ME 7, p. 26). In the first edition, he even went so far as to refer this distinction to the one between an «authoritative, “categorically imperative” function of the Practical Reason» and «another in which its operation is more subordinate, prescribing not the end of the action but only the means to a given end. In this latter case the end is determined by desire or impulse of some kind, which may or may not be itself rational» (ME 1, p. 24). But the Kantian phraseology appears in some passages also in the last editions, for example where he says he wants to exhibit moral obligation as an «unconditional or categorical imperative» (ME 7, p. 35), and where he contrasts this categorically imperative interpretation of “ought” with the “ought” of the hypothetical imperative (ME 7, p. 37). Moreover, he declares that i) certain kind of actions «are commonly held to be right unconditionally, without regard to ulterior results» (Ibid.) and ii) that the same is true for the adoption of certain ends, such as the common good or general happiness. Lastly, there is also a very Kantian flavour in what today we would call the frank rationalistic internalism of Sidgwick's moral epistemology: to say that ethical judgments are dictates of reason, in fact, is for him also to say that «in rational beings as such this cognition gives an impulse or motive to action» (ME 7, p. 34), even though not always a predominant one.

Sidgwick's acceptance of the very notion of practical reason is indeed very rationalistic, anti-Humean, and generally foreign to the empiricist tradition; and Schneewind rightly suggested that there is a Kantian strain in Sidgwick's

---

<sup>24</sup> H. Sidgwick, *Some Fundamental Ethical Controversies*, p. 44.

notion of intuition «as the understanding a rational being has of the nature of his own activity as reasonable»<sup>25</sup>. This is in fact what seems to be implied in Sidgwick's notion that there are certain absolute practical principles, or axioms, «the truth of which, when they are explicitly stated, is manifest» (ME 7, p. 379): that, although there is no universal code of moral norms that are unconditionally valid for all human beings, there are certain formal principles that no rational being can ever deny, for they are, so to speak, consubstantial to the rational mind. However, it is also clear that Sidgwick does not accept Kant's specific idea of practical reason, that is, the idea that reason can be practical only by being pure, i.e. by putting aside all inclination and every other empirical element<sup>26</sup>. While it is clear that Sidgwick does not think of reason as a mere faculty of means, and of judgments as to what is right or ought to be done as mere instrumental judgments, it is also evident that, for him, the formal requirements of reason do not generate ends independently of any inclination; in fact, of the above mentioned absolute practical principles, he says that «they are of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case» (ME 7, p. 379). And he links the unconditional obligations he admits to the recognition of a universal end at which it is ultimately reasonable to aim, so that the obligation must concern acts mostly conducive to such end: in this way, he observes, «The obligation is not indeed “unconditional”, but it does not depend on the existence of any non-rational desires or aversions» (ME 7, p. 35).

As for Kant, in the second section of the *Grundlegung* he strongly denies that happiness can be the source of the fundamental principle of morality, even though it is the only end that can be said to be pursued by all rational beings; the problem is that the imperatives of happiness command not necessarily, but only assertorically, in that they command something not for itself, but for something else, which we naturally will; moreover, such imperatives are *consilia*, rather than *praecepta*, for it is not possible to determine with certainty what will promote the happiness of a rational being at best. So Kant concludes that there is «but one categorical imperative, namely, this: Act only on that maxim whereby thou canst at the same time will that it should become a universal law» (G, p. 268).

---

<sup>25</sup> J. B. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy*, p. 420.

<sup>26</sup> See O. O'Neill, *Sidgwick on Practical Reason*, «Proceedings of the British Academy», 109, 2001, pp. 83-89. It could be argued, nonetheless, that Sidgwick's way of formulating the principle of universal benevolence in ME comes close to the purely formal indication of the others' happiness as an end that is at the same time a duty in *The Metaphysics of Morals* II, Intr., V. B.

This formula of the universal law is the one that Sidgwick constantly assumes as defining the Kantian position; according to him, the formula expresses the Golden Rule in a philosophically respectable form and gives general formulation to the idea of Justice as Impartiality (i.e. «that whatever is right for me must be right for all persons in similar circumstances», ME 7, p. xvii). Of this formula, however, he also says that it is «inadequate for the construction of a system of duties» (Ibid.), and that it does not «settle finally the subordination of Self-Interest to Duty» (Ibid.). In other passages — though with no direct reference to Kant — he adds that, strictly speaking, the effect of this principle «is merely to throw a definite onus probandi on the man who applies to another a treatment of which he would complain if applied to himself» (ME 7, p. 380); and he repeats the charge of insufficiency within the context of the administration of law, for «[the principle of impartiality] does not help us to decide what kind of rules should be thus impartially applied; though all admit the importance of excluding from government, and human conduct generally, all conscious partiality and ‘respect of persons’» (Ibid.). Finally, he observes that the principle must be qualified by the belief that, in practice, the action whose maxim is being tested will not be widely imitated; otherwise, we should reject maxims such as the one to adopt celibacy, for, were it universally applied, it would determine the greatest of all crimes, i.e. the disappearance of the human kind. In short, the Kantian principle, for Sidgwick, «means no more than that an act, if right for any individual, must be right on general grounds, and therefore right for some class of persons; it therefore cannot prevent us from defining this class by the above-mentioned characteristic of believing that the act will remain an exceptional one» (ME 7, pp. 486-487).

While these contentions are fundamentally acceptable, a presentation of Kant’s ethics, such as the one given by Sidgwick, centring only on this formula is in itself highly doubtful. It is true, of course, that Kant does say that «In forming our moral judgement of actions, it is better to proceed always on the strict method and start from the general formula of the categorical imperative: Act according to a maxim which can at the same time make itself a universal law» (G, p. 275). However, it cannot be forgotten that Kant does give at least two more formulas of the imperative; moreover, it is not at all clear that the “general formula” to which he refers in the passage just quoted must be identified with the formula of the universal law. In fact, it is clearly more sensible, and even more true to the letter of Kant’s text, to interpret the sequence of the three formulations of the categorical imperative not just as the repetition of the same principle, but as a development of one central idea, progressively viewed from different perspectives. The ideas of humanity

as an end in itself, and of autonomy as the universally legislative will of every rational being, do in fact add much to the mere non-contradiction of the maxims, that is, to the purely formal condition set by the first formula. In view of this progressive development, there are serious reasons for the view that, when speaking of the “general formula”, Kant is really intending the formula of autonomy, that is, the one that, in synthesizing the formula of the universal law and that of humanity, constitutes the most complete wording of the one fundamental principle<sup>27</sup>.

Sidgwick never mentions the formula of autonomy, nor its variant centring on the kingdom of ends — not in *The Methods* nor in the *Outlines*; and shortly discusses the formula of humanity, both in the *Outlines* and in a long note in *The Methods*.

In the *Outlines*, Sidgwick introduces the formula of humanity after recalling Kant’s thesis that ethics, unlike jurisprudence, is concerned with the realisation of internal freedom through the pursuit of rational ends, as opposed to the ends of natural inclination. Of Kant’s statement that rational beings are ends in themselves, he notes that it is hardly a clear answer to the question asking what are the ends of reason. That statement might be interpreted as meaning that we should pursue the development of rationality, and therefore of morality, in every imperfectly rational being; but Sidgwick rightly dismisses this interpretation, since Kant clearly states that it would be a contradiction to promote the others’ perfection: in fact, it is central to the attainment of intellectual and moral perfection that every man should autonomously pursue it. While we have a moral duty to cultivate ourselves, we cannot be morally bound to bring about others’ perfection<sup>28</sup>. Having thus discarded perfection, Sidgwick sees no other way to interpret the formula of humanity than that according to which it commands to aim at the only other producible end of which Kant in fact speaks, that is, the happiness of others: therefore, everyone «is to help others towards the attainment of those purely subjective ends that are determined for each not by reason but by natural inclination» (OHE, p. 275).

In the *Note on Kant* concluding chapter xiii of Book III<sup>29</sup>, Sidgwick links again the discussion of the formula of humanity to the attempt to establish the principle of beneficence. Here he notes that the derivation of this principle from the formula of universal law is not cogent, since we can clearly con-

---

<sup>27</sup> See A. W. Wood, *Kantian Ethics*, pp. 82-84.

<sup>28</sup> In ME 7, pp. 239-240, Sidgwick tries to show that this Kantian thesis is untenable.

<sup>29</sup> This note replaces the slightly larger discussion that appeared in ME 1, pp. 360-363 as § 4 of the same chapter XIII. The main passages here referred to are substantially identical.

ceive of a man who would prefer not to be aided by others to accepting obligations to aid, or who believed that a maxim of pure egoism is on the whole preferable, prudentially speaking. In this context, he expressly says that, for Kant, the fact that others are ends in themselves means that «we must recognise the duty of making their happiness our end» (ME 7, p. 389). Moreover, he reconstructs a different line of argument for the same principle, the one according to which, for Kant, since no particular object of inclination can be constituted as an absolute dictate of reason, only rational beings in themselves can be one such unconditional object or end. Then he goes on to criticise this argument, by noting that: i) to say that humanity is a self-subsistent end is perplexing, «because by an End we commonly mean something to be realised» (ME 7, p. 390); ii) there is a paralogism in saying that Men are ends in so far as they are rational and then deriving from this the duty to adopt as our ends their subjective, non rational ends: «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» (Ibid.).

I will try to show in the next section that Sidgwick fails to understand the concept of humanity as an end in itself; this is not to be identified with a principle of beneficence according to which we are to make the subjective ends of others our own end, but with that self-subsistent end that grounds both our perfect and our imperfect duties towards ourselves and towards other (one of this last duties being the duty of beneficence). I will also argue that the failure to understand this key concept of Kant's ethics is at the heart of Sidgwick's substantial dismissal of the Kantian project.

#### *4. The interpretation of the formula of humanity as the key to Sidgwick's misunderstandings*

Two points must be stressed in Sidgwick's interpretation of the formula of humanity: the first is that Sidgwick clearly fails to grasp the meaning of the phrase "self-subsistent end". Of course, commonly an end is something to be realised; but Kant explicitly says that

since in the idea of a will that is absolutely good without being limited by any condition (of attaining this or that end) we must abstract wholly from every end to be effected (since this would make every will only relatively good), it follows that in this case the end must be conceived, not as an end to be effected, but as an independently existing end. Consequently it is conceived only negatively, i.e., as that which we must never act against and which,



therefore, must never be regarded merely as means, but must in every volition be esteemed as an end likewise (G, p. 276).

This conception of an end is perhaps uncommon in ordinary speech, but definitely not mysterious, and even standard in philosophical language: a self-subsistent end is simply a being already existing, for the sake of which something must be done, that is, a being that sets constraints on actions designed to produce any other end. It is not, therefore, an object of production but of respect; and humanity is such an end on account of the peculiarity of rational nature, that is, because of its capacity to set ends by herself, to have freedom and therefore moral agency. The idea of a self-subsistent end is simply the idea of an already existing source of the value of all ends, i.e., of a being whose unconditional worth grounds all conditional values. This notion of an end is explicitly worked out in Medieval thought, but has its roots in the Aristotelian idea of a final cause<sup>30</sup>; and it is surprising that Sidgwick, who knew very well Aristotle's work, did not grasp the meaning of this traditional idea.

The second point that must be emphasized is that, in commenting on the formula of humanity, Sidgwick fails to see that Kant is here after something which is central to his own enterprise. The formula, in fact, is not Kant's convoluted and perhaps inconsistent way of establishing the principle of rational benevolence, as Sidgwick seems to think; rather, it is his attempt to establish a much more fundamental principle that can be thought of as the basis of more specific normative principles, and, at the same time, as their limiting condition. Kant is actually giving substance, or matter, to his purely formal wording of the categorical imperative, as presented in the formula of the universal law. He seems indeed to agree with Sidgwick on the insufficiency of the formal principle, on its inadequacy for «complete guidance»; therefore, he complements it with a material principle, expressing the central value implicit in the formula of universality, that is, humanity as the capacity to set ends for oneself. Kant's principle of humanity therefore plays the same structural function as Sidgwick's universal benevolence, but has a much wider scope; it is the underlying principle at the basis of such diverse rules as the perfect duties not to commit suicide and not to make false promises, and

---

<sup>30</sup> The origin of Kant's phrase is traced by A. Donagan in the double notion of *finis* present both in Thomas Aquinas and John Duns Scotus (see *Human Ends and Human Actions: An Exploration in St. Thomas's Treatment*, Aquinas Lecture Series, Marquette University Press, Milwaukee 1985; repr. in Id., *Reflections on Philosophy and Religion*, Oxford University Press, Oxford 1999, pp. 81-97). Aristotle's main passage for the same idea occurs in the very famous chapter 7 of Book XII of the *Metaphysics* (1072a-1073a).

the imperfect duties to cultivate one's perfection and to pursue others' happiness; at the same time, it is the ground of the constraints on the imperfect duties. The road that leads Kant to this principle, as we saw, is different from the one followed by Sidgwick: while Sidgwick reaches the maxim of universal benevolence by showing that it is somehow implied in the ordinary rules and duties that he discusses in detail, Kant starts with the ordinary conception of good will and duty, showing that it implies both the idea of a universal legislation and of the autonomy of rational beings. Both philosophers, however, are looking for deeper principles, fundamental axioms or "intuitions", that can firmly concatenate practical maxims and adjudicate conflicts between them; and both understand such principles as expressing the fundamental nature of reason, and therefore as deeply embedded in our nature as rational beings.

What Sidgwick completely fails to see is therefore that the formula of humanity is the second step in the working out of the fundamental principle of Kant's ethics: this principle, in fact, is not — pace Sidgwick and most other commentators — the mere formal requirement that maxims should be universalisable<sup>31</sup>. Kant himself declares that maxims must have a matter or end; however, all material ends are relative and give rise only to hypothetical imperatives. Therefore, the source of a practical law, that is, of a categorical imperative, cannot be but in something whose existence has in itself an absolute worth, for «if all worth were conditioned and therefore contingent, then there would be no supreme practical principle of reason whatever» (G, p. 272). This something, which is rational nature in humans and other rational beings, is therefore an end, not in the sense of something to be produced, but in that of something to be respected that gives matter or content to the pure formality of a universal law: by rendering them moral beings, that is, capable of acting on the basis of the representation of a law, such matter or content in fact grounds the dignity of human beings, which, for Kant, cannot be replaced by something equivalent, for it is the source of all relative values, both of market values and of fancy values.

It is therefore correct to say that the formula of the universal law is insufficient for the construction of a whole system of duties; as Sidgwick notes, «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» (ME 7, p. 210). The

---

<sup>31</sup> This point is very well developed by A. Wood, *Kant's Ethical Thought*, Cambridge University Press, Cambridge 1999, pp. 182-190. Cf. Id., *Kantian Ethics*, pp. 85-95.

criterion of universalisability is a necessary but insufficient condition of morality, just because that formula is not the whole of Kant's fundamental principle. The principle is completely spelled out only when it is shown that all moral maxims, that is, all the maxims that can be justified from the moral point of view, have i) a form, i.e. universality; ii) a matter or end, i.e. rational nature as «the condition limiting all merely relative and arbitrary ends» (G, p. 275); and iii) «A complete characterization of all maxims by means of that formula, namely, that all maxims ought by their own legislation to harmonize with a possible kingdom of ends as with a kingdom of nature» (Ibid.). The third condition, which sums up the first two formulas through the idea of autonomy and of a kingdom of ends, is that every maxim should aim at a systematic connection of rational beings in a sort of kingdom; such a kingdom is defined by the fact that everyone is at the same time a legislator and subject to laws (i.e. universal objective principles) devised to treat everyone not as a means only but always at the same time as an end; this is the final and really complete formulation of Kant's fundamental principle — a formulation to which Sidgwick never makes reference throughout his work.

Understanding the role of the formula of humanity in the context of Kant's ethical system also helps to solve the specific problem posed by Sidgwick with reference to the rule of benevolence. In fact, the formula of humanity is in the first place a limitative condition on the acceptability of maxims; its main practical effect is to reject maxims that would allow treating rational beings as mere means for others' ends. Moreover, the formula has positive implications in suggesting the adoption of the two ends that, according to *The Metaphysics of Morals*, are at the same time duties: one's perfection and others' happiness, that is, the duties that in the *Grundlegung* appeared as examples of the imperfect duties towards oneself and towards others<sup>32</sup>. When we come to the positive part, however, we must not forget that rational beings are considered as ends in themselves qua rational. Therefore, while limiting oneself to withholding disrespectful actions would be to value rational nature too poorly, only in a negative fashion, adopting others' subjective ends is always constrained by the moral non-rejectability of such ends: when Kant says that «the ends of any subject which is an end in himself ought as far as possible to be my ends also» (G, p. 273), the possibility of

---

<sup>32</sup> It may plausibly be contended that the notion of the two ends that are at the same time duties, put forth in the *Metaphysics of Morals*, constitutes a real development with respect to the *Grundlegung*, where in fact this notion does not appear. However, the development may be interpreted as the working out in detail of the practical impact of the more abstract principle of the *Grundlegung*. On this issue see D. Tafani, *Il fine della volontà buona in Kant*, in L. Fonnesu (ed.), *Etica e mondo in Kant*, il Mulino, Bologna 2008, pp. 145-163.

which he is speaking is moral, not merely physical. In fact, in the later *Metaphysics of Morals*, he says that love is a maxim of benevolence, resulting in beneficence, this consisting in «the duty to make others' ends my own (provided only that these are not immoral)»<sup>33</sup>. In other words, respecting rational nature in any individual does include promoting those contingent ends that make up her life project and from which she can expect her happiness; but i) this is not the main meaning of such respect, and ii) this promotion is in any case constrained by the prior acceptability of those ends<sup>34</sup>.

Sidgwick is therefore wrong when he points to the almost complete coincidence (ME 7, p. 385) between Kantian ethics and utilitarianism, based on the fact that «the only really ultimate end which he [i.e. Kant] lays down is the object of Rational Benevolence as commonly conceived—the happiness of other men» (ME 7, p. 386)<sup>35</sup>. Actually, not only does Kant add the other imperfect duty to cultivate one's intellectual and moral perfection; but Sidgwick also forgets that the duty of beneficence, as well as that of perfection, is limited by the negative part of the respect for rational nature as an end in itself. Kantian beneficence is therefore significantly unlike utilitarian one, for it is not the unlimited pursuit of others' subjective aims, as determined by their natural inclinations, but the pursuit of those subjective aims that pass the scrutiny of rational reflection, that is, that are not to be rejected on the basis of the two tests of universalisability and non-exploitation. The difference between Kantian and utilitarian beneficence was not grasped by Sidgwick, because he did not really understand the notion of a self-subsistent end, nor the role of the formula of humanity as a fundamental axiom or principle grounding those of perfection and benevolence, and not reducible to the last

---

<sup>33</sup> I. Kant, *The Metaphysics of Morals*, Cambridge University Press, Cambridge 1996, *Doctrine of the Elements of Ethics*, § 25, p. 199.

<sup>34</sup> It must be noted that Sidgwick's interpretation of the formula of humanity has shaped its understanding by many contemporary commentators: paradigmatically, it is adopted by R. M. Hare, *Could Kant Have Been A Utilitarian?*, in *Sorting Out Ethics*, Oxford University Press, Oxford 1997, pp. 147-165. In this same perspective, M. Nakano-Okuno recently went so far as to affirm not only that «this formula shares essentially the same claim as the Principle of Rational Benevolence», but also that it «encompasses the essential claim of Sidgwick's Principle of Rational Prudence», since it imposes to treat one's future ends as they were present (*Sidgwick and Kant: On the So-Called "Discrepancies" Between Utilitarian and Kantian Ethics*, in P. Bucolo, R. Crisp, B. Schultz [eds.], *Henry Sidgwick: Happiness and Religion*, Dipartimento di Scienze Umane, Università degli Studi di Catania, Catania 2007, pp. 260-333, here at p. 292 and p. 294).

<sup>35</sup> In a passing note of the first edition, he even committed himself to such an absurdity as to say that «in fact, as we have seen, [the utilitarian first principle] is the first principle of Kantism» (ME 1, p. 440).

one. Moreover, he could not realise the resources of the formula of humanity in adjudicating between conflicting grounds of duty: in fact, the formula clearly seems to justify a relative priority of perfect duties over imperfect ones, while leaving the last word to the exercise of judgment in the circumstances<sup>36</sup>. Finally, it is Sidgwick's failure to understand and to accept such notions as humanity as an end in itself, human dignity and moral autonomy that explains one of the most critical points of his ethical views, probably the one most often quoted in recent debate: his refusal to link moral reasonableness to the demands of publicity, and his consequent acceptance of utilitarianism as an esoteric morality reserved to the enlightened few<sup>37</sup>. While both philosophers emphasize the role of ordinary moral knowledge, Kant is in fact much more open-minded and 'progressive' in his recognition of the intellectual and moral competence of ordinary people; though he does not bring out all the consequences of his notion of morality as self-governance, his ideas of humanity and human dignity made him stand well over Sidgwick's still elitist morality. In the end, as noted by Schultz, the strongest difference between the two thinkers perhaps lies in the fact that it is hard «to find in Sidgwick's idea of a method of ethics an effectively Kantian endorsement of the plain person's capacity for moral self-direction»<sup>38</sup>. The reason of this difference probably lies in the different historical and cultural contexts of the two philosophers: while Kant, who wrote in the age of Enlightenment and in the context of the hopes generated by the American and French revolutions, had moderately optimistic views on history and the potentialities of human development, Sidgwick can be considered a sort of critic of the Enlightenment, and his elitist conclusion is the effect of the disbelief in any optimistic philosophy of history and in the reality of moral human progress.

---

<sup>36</sup> This seems to be implicit in the very tentative casuistry sketched by Kant in *The Metaphysics of Morals*. Also the *Lectures on Ethics* often testify to a somewhat more flexible attitude (even as far as truth-telling is concerned) on immediately practical matters than is generally thought.

<sup>37</sup> B. Schultz appropriately notes that «it is perhaps at this juncture that one can best appreciate how Sidgwick parted from the Kantian project» (*Henry Sidgwick: Eye of the Universe. An Intellectual Biography*, Cambridge University Press, New York 2004, p. 264). The centrality of the idea of publicity is particularly stressed, in a Kantian and anti-utilitarian vein, by J. Rawls, *Political Liberalism*, Columbia University Press, New York 1993, pp. 66-71. The charge that Sidgwick's esoterism amounts to "Government House" utilitarianism is notoriously put forward by B. Williams, *The Point of View of the Universe: Sidgwick and the Ambitions of Ethics*, in *Making Sense of Humanity and Other Philosophical Papers*, Cambridge University Press, Cambridge 1995, pp. 153-171.

<sup>38</sup> *Henry Sidgwick: Eye of the Universe*, pp. 267-268.

## 5. *Two speculations*

Sidgwick's avowed incomprehension of both the formula of humanity and the third formula (autonomy/kingdom of ends) is the key to understanding his final rejection of Kant's system, and his interpretation of Kant as an intuitionist<sup>39</sup>: this incomprehension explains his purely formalistic reading and his strategy to complement the formal principle of universalisability (justice, or impartiality) with the substantive principle of rational benevolence.

A further problem remains, as to why Sidgwick never tried to deepen his understanding of the whole Kantian project in ethics. On this, no more than speculations may be offered.

One tentative answer might be that Sidgwick was deeply convinced of the untenability of Kant's theoretical philosophy, and that this conviction deterred him from embarking in a serious study of Kant's ethics. There is in fact at least one passage in the Lectures on the first Critique in which Sidgwick stresses the uncomfortable consequences of Kant's epistemology for ethics. Speaking of the ideality of time in the transcendental aesthetics, he notes that this doctrine has the effect of rendering intellectual and moral progress mere appearances:

Hence the conception of moral progress, on which the practical postulate of immortality — as we saw — is based, is a conception that represents no real fact of any soul's existence, but merely an appearance due to the imperfection of its faculty of cognition. But if moral progress is thus reduced to mere appearance, what becomes of the belief in the immortality of the soul which Kant (in the Critique of Practical Reason) bases on it? Indeed, in any case, if Time is merely a form of human sensibility, — due to an imperfection of man's nature which prevents him from knowing things as they are, — the postulate of immortality seems to become a postulate for the endless continuance of an imperfection. It does not seem that this can afford an inspiring hope for a truth-loving mind<sup>40</sup>.

---

<sup>39</sup> It is curious that Sidgwick never puts any emphasis on the only passage lending true credibility to an intuitionist reading of Kant: the passage in the second *Critique* speaking of the moral law as a "fact of reason". To my knowledge, Sidgwick never mentions that passage in his works; most of his references to Kant are to the *Grundlegung*, and certainly the second *Critique* is never mentioned in *The Methods*, but for a note in passing at the beginning of the last chapter of ME 1, p. 439 — a note that was subsequently removed.

<sup>40</sup> *Lectures on the Philosophy of Kant*, p. 36.

And although he adds that, in his practical philosophy, Kant seems to defend a sort of noumenal freedom of the soul, according to which «the momentous choice between good and evil which every human soul makes is in reality not subject to the condition of time, so that any change that may appear in a man's character is illusory»<sup>41</sup>, it is clear that in either way the metaphysical underpinnings of Kant's position were deeply uninviting for Sidgwick.

This explanation, however, is far from satisfying. For one thing, it is unclear why Sidgwick, as an historian of philosophy, should not have wished to deepen his understanding of Kant's moral doctrine, as he had done with the epistemological ones, even knowing that it was based on metaphysical grounds utterly unpalatable for him. Moreover, Sidgwick might well have done with the notions of humanity, autonomy and the kingdom of ends what he had done with the notion of universalisability: that is, to accept what he found useful in the normative principle, while discarding its theoretical underpinnings. This is in fact how he describes his attitude in the autobiographical note: «What commended itself to me, in short, was Kant's ethical principle rather than its metaphysical basis» (ME 7, p. xvii).

Another possible explanation is that Sidgwick approached Kant's ethics with a serious bias deriving from his previous acceptance of Mill's utilitarianism, and that, although he subsequently tried to revise his first impression by rereading Kant, he never fully succeeded in developing an unbiased analysis. Owing to the prejudices he had inherited from Mill, he never got convinced of the necessity of a deeper understanding of Kant's ethical project, and this may explain the lack of a direct confrontation with it. In other words, Sidgwick started as a utilitarian and never ceased to be one; the difficulties he found in the theoretical frameworks of the masters of his school led him to reappraise the importance of common sense morality, as well as of authors such as Kant, Clarke and Butler: but he never ceased to think of utilitarianism, however revised, as the most satisfactory moral theory (at least, *faute de mieux*). So, he wanted to be a utilitarian — perhaps just as he wanted to be a Christian, though he carried out the latter endeavour less successfully,— and tried what he could in order to rescue utilitarianism from its defects; perhaps this attitude was also suggested by the Millian conviction that utilitarianism was the theory associated with moral and social progress, while all other theories were, in some way or other, conservative.

This explanation might be not particularly respectful of Sidgwick intellectual honesty; however, it has some textual evidence in its favour. Not only Sidgwick does start his auto-biographical note declaring that he adhered to

---

<sup>41</sup> *Ibid.*, p. 37.

Mill's utilitarianism from the start; he also adds that, in his subsequent search for a deeper foundation of ethics, he retained a general «attitude of discipleship to Mill» (ME 7, p. xvi); he says that «through all this search for principles I still adhered for practical purposes to the doctrine I had learnt from Mill» (ME 7, p. xviii); and he explicitly declares that the first time he read Kant, he read it «somewhat unintelligently, under the influence of Mill's view as to its grotesque failure» (ME 7, p. xvii). True, he also adds that he reread it «more receptively» (Ibid.), discovering the importance of its fundamental principle. However, it is clear that even this second reading was in fact influenced by Mill: the idea of Kant's ethics as a merely formalistic system, and the neglect of the second and third formulas on which it is based, though commonsense in most philosophical literature, are central in Mill's reading; and the idea that the only end laid down by Kant through his imperative is rational benevolence is also near to the main tenet of Mill's interpretation. In the passage on the «grotesque failure» of Kant's ethics quoted by Sidgwick, Mill says that the only contradiction that the test of universalisability is able to detect is the one between certain immoral rules of conduct and the general desires of humanity; that is, the Kantian principle is meaningful, and able to justify duties of morality, only if interpreted in consequentialist terms, as rejecting the maxims on account of their consequences. Moreover, in the other passage of Utilitarianism in which Kant is mentioned, Mill says that the only meaningful sense of Kant's fundamental principle is that «we ought to shape our conduct by a rule which all rational beings might adopt with benefit to their collective interest»<sup>42</sup>; here again Mill interprets what Kant intends as an a priori constraint on moral maxims as a concern for the consequences of moral rules. Sidgwick's misunderstanding of Kant is similar: he likewise fails to appreciate the fruitfulness of the categorical imperative, in its three progressive formulations, and insists that it does not justify any moral rule per se, though it is the first step in the process of justifying the utilitarian principle of rational benevolence. In short, notwithstanding his testimony that he reread Kant's ethics «more receptively», Sidgwick seems to have been receptive only to the possibility of incorporating the Kantian idea of the universality of morality into the utilitarian system; that is, he complemented Mill's attempt to show the compatibility of Kant's

---

<sup>42</sup> J. S. Mill, *Utilitarianism* (1861), in *The Collected Works of John Stuart Mill*, Volume X - *Essays on Ethics, Religion, and Society*, edited by J. M. Robson, University of Toronto Press, Toronto 1985, pp. 205-260, here at p. 249. The passage on Kant's failing «almost grotesquely» is at p. 207.



ethics and utilitarianism, a compatibility that embodies a misunderstanding of Kant very similar to Mill's.

## 6. *Conclusive remarks*

My main contention has been that Sidgwick did not fully grasp the heart of Kant's moral project. This means that he did not have the opportunity to discover that Kant's ethics was much more congenial to his mature thought on ethics than he assumed, and also that, for certain aspects, it was much less conservative than most "dogmatic intuitionism"; in short, he did not have the opportunity to consider an alternative and plausible way to accomplish the philosophical task that he considered decisive. This far, we have not suggested that, should Sidgwick have understood Kant more deeply, he would have taken up the Kantian way, or that he could have been (some sort of) a Kantian. It is of course possible, and indeed likely, that Sidgwick's prior acceptance of a strictly teleological conception of ethics, according to which for an action to be morally appropriate is to maximally promote some good, would have prevented him from accepting both Kant's general theory and more specific ideas such as the conception of beneficence.

However, we can also consider some aspects that may have recommended to him the Kantian solution. The first point to consider is that the deontological conception prioritising the right over the good seems to be embedded in the morality of common sense in a way that the one-sided utilitarian insistence on consequences seems not. Of course, there are cases in which utilitarianism can be easily accorded with our considered judgments; but the central cases discussed by Sidgwick in Book III, those that are the traditional object of the utilitarian polemic, are not of this sort. Let us take the classical example of promises. Throughout his discussion, Sidgwick basically aims at showing that common sense is in many cases uncertain as to the boundaries of the duty to keep promises. However, he clearly, though implicitly, acknowledges that the traditional casuistry — such as it had been revived and systematised by Whewell<sup>43</sup> — had defined several precise conditions for the treatment of hard cases, such as that a promise is binding «if the promiser has a clear belief as to the sense in which it was understood by the promisee, and if the latter is still in a position to grant release from it, but unwilling to do so, if it was not obtained by force or fraud, if it does not conflict with definite a priori obligations, if we do not believe that its fulfilment will be harmful to the

---

<sup>43</sup> W. Whewell, *Elements of Morality, Including Polity*, III edition, Parker, London 1854.

promisee, or will inflict a disproportionate sacrifice on the promiser, and if circumstances have materially changed since it was made» (ME 7, p. 311). Nowhere does this traditional treatment refer to the mere balance of good versus bad consequences in order to solve particular problems; on the contrary, it always keeps to the deontological intuition according to which principled, a priori solutions, not mere cost-benefit analyses, are needed also for hard cases. Here, as in other cases, Sidgwick's thesis that common sense lacks clear answers is perhaps right, but the same cannot be said of the systems of intuitionist philosophers such as Whewell, whose solutions Sidgwick simply omits to discuss<sup>44</sup>; his conclusion that precise duties can be defined, and conflicts of duties resolved, only by reference to the principle of utility, is therefore unwarranted. Moreover, to expect that the application of the principle of utility would bring to the treatment of promises much more definition and "scientific" precision than afforded by traditional casuistry is both to require from moral philosophy much more than it is legitimate (witness the same Aristotle that Sidgwick is imitating) and to overstate the utilitarian ability to predict specific consequences in particular cases. Finally, on the basis of his alleged aim to provide a philosophical foundation for the morality of common sense, Sidgwick himself ought have been sympathetic to the efforts of ethical theories that tried to treat hard cases and alleged conflicts of duties without giving up the deontological intuitions at the heart of ordinarily acknowledged duties: if not Kant's, at least the more philosophically refined formulations of the so-called "intuitionistic theory", such as Whewell's. But Kant's theory, correctly understood, offered precisely what Sidgwick was looking for: a philosophical principle, developed in the three stages spelled out in the *Grundlegung*, that can both systematise the rules of common sense morality and provide principled ways to adjudicate conflicts between them.

The second point to consider is the conflict between happiness and duty, that is, what Sidgwick styled the dualism of practical reason, and what he considered the most serious problem of ethics. Sidgwick's dissatisfaction with Kant's solution to this problem is well known; in a passage of the *Memoir* already mentioned, he recalls that, when writing *The Methods*, he was «inclined to hold with Kant that we must postulate the continued existence of the soul, in order to effect that harmony of Duty with Happiness which seemed to me

---

<sup>44</sup> On Sidgwick's unfairness in pointing to the difficulties of the morality of common sense in treating with hard cases without seriously discussing the philosophical efforts developed to bring consistency and systematisation to it, see A. Donagan, *Sidgwick and Whewellian Intuitionism: Some Enigmas*, «Canadian Journal of Philosophy», 7, 1977, pp. 447-465; repr. in B. Schultz (ed.), *Essays on Henry Sidgwick*, cit., pp. 123-142, and S. Cremaschi, *Nothing to Invite or Reward a Separate Examination. Sidgwick and Whewell*, in this issue.

indispensable to rational moral life. At any rate I thought I might provisionally postulate it, while setting out on the serious search for empirical evidence» (HSM, p. 467). Such empirical evidence should have come from the parapsychological investigations to which Sidgwick devoted much efforts throughout his life. In 1874 his hopes had probably already weakened enough to make the Kantian postulation unacceptable: in fact, in the first edition of *The Methods*, he already declares what we also find in all other editions, that is, that he could not

fall back on the Kantian resource of thinking myself under a moral necessity to regard all my duties as if they were commandments of God, although not entitled to hold speculatively that any such Supreme Being really exists “as Real”. I am so far from feeling bound to believe for purposes of practice what I see no ground for holding as a speculative truth, that I cannot even conceive the state of mind which these words seem to describe, except as a momentary half-wilful irrationality, committed in a violent access of philosophic despair (ME 1, p. 471; cf. ME 7, p. 507).

The process of disillusion had been (almost) completed by 1887, when he writes: «I have been facing the fact that I am drifting steadily to the conclusion — I have by no means arrived at it, but I am certainly drifting towards it — that we have not, and are never likely to have, empirical evidence of the existence of the individual after death» (HSM, p. 466). Lacking any such evidence, Sidgwick seems to be totally bereft of reasons to accept Kant’s postulation.

Sidgwick might have solved the dualism of practical reason only by questioning the rationality of egoism, which he clearly was not quite prepared to do. As he makes clear in a 1889 paper, such rationality is based on the reality and fundamentality of the distinction between any one individual and any other, so that «I’ am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals»<sup>45</sup>; it is in fact based on the very idea of the «separateness of persons» urged by Rawls against utilitarianism. For Sidgwick, this preference of private happiness to virtue, or general happiness, is just as much a dictate of reason as the proposition that my own good is no more important than the good of any other; for Kant, the demands of reason are in no way conditioned to the effective reconciliation of happiness and virtue, which must be postulated and hoped for, but cannot be the motive of action: morality teaches us how to become deserving of happi-

---

<sup>45</sup> H. Sidgwick, *Some Fundamental Ethical Controversies*, p. 44.

ness, without assuring that we will actually be happy. In the end, persons are for Sidgwick much more separate than they are for Kant. In fact, thanks to his identification of reason with the very capacity for universality, Kant can question the rationality of egoism; to be rational is in fact to acknowledge that maxims by which we are making exceptions for ourselves are not justifiable<sup>46</sup>. Persons are thus separate, but practical reason is a point of view with which all human beings can identify themselves, a common identity rooted in their nature as rational beings. For Sidgwick, instead, the rationality of overcoming egoism can only be seen by viewing things from the point of view of the universe; this, however, is actually no one's point of view: it is not a perspective rooted in our nature as the first person perspective that grounds rational egoism. This account seems to emphasize the separateness of persons more than the Kantian one, for here the viewpoint of universality seems not one that is shared by all, but one to be constructed by summing the perspectives of all; in this perspective, the concern with «the quality of my existence» cannot but trump any interest for universal benevolence.

In conclusion, there are some reasons that may suggest that a deeper understanding of Kant and a more direct confrontation with Kant's ethical treatises might have led Sidgwick to second thoughts on ethics, the rationality of egoism, and his final rejection of the deontological stance of the morality of common sense. It is a fact, however, that Sidgwick never embarked in such a confrontation: and my speculations on the possible reasons for this circumstance lead to single out both his distaste for Kant's metaphysics and his Millian utilitarian bias<sup>47</sup>.

---

<sup>46</sup> Kant's dualism between instrumental and moral rationality is therefore very different from Sidgwick's dualism (and from other dualisms recalled by Sidgwick), as was shown by W. K. Frankena, *Sidgwick and the History of Ethical Dualism*, in B. Schultz (ed.), *Essays on Henry Sidgwick*, pp. 175-198.

<sup>47</sup> I wish to thank Sergio Cremaschi and Gianfranco Pellegrino for very helpful comments on an earlier draft.