

UNIVERSALIZING AS A MORAL DEMAND

Oliver SENSEN¹

Kant formulated several moral ideas which remain highly popular. Among these are autonomy, dignity, and respect for humanity,² but also the idea of universal law. It is a powerful idea that laws should be valid equally for all human beings, independently of race, gender, birth, income or age. Kant expresses this idea in his Categorical Imperative. But it is controversial whether the imperative can yield concrete duties, and how exactly it is to be applied. Since Kant's own time, the Categorical Imperative has been criticized as being empty, i.e., devoid of content, and therefore unable to generate any duties on its own. In later times, especially when Kantianism was considered as an alternative to Utilitarianism, the criticism was that the law rules out too much in one respect, too little in another, and what it rules out, it rules out for the wrong reasons. It seems that the Categorical Imperative has been abandoned as a decision procedure in ethics. However, the general idea remains popular, and has received support from empirical philosophy. It has been argued that even very young children make a distinction between moral rules, which are seen to be generalizably and universally valid, and mere conventions (cf. Nichols 2004, 5-7).

My aim in this paper is therefore to revisit the issue. The empirical evidence and continued popularity warrants a fresh look. I shall consider why the Categorical Imperative is a recognizably moral idea, and how it might generate specific duties. In order to do so, I shall first specify the most popular objections that have been raised against the Categorical Imperative (Section 1), before considering the main Kantian reply (Section 2). I shall then argue – with the help from Kant's Formula of Humanity – that the Categorical Imperative expresses a different but genuine moral idea (Section 3). Finally, I shall consider how one can apply the Categorical Imperative, and suggest a way how it can be used reliably to generate concrete duties (Section 4). My claim is that the Categorical Imperative expresses the demand of fairness: One should not make an exception for oneself in the sense that one should not regard oneself as something better. I shall argue that the main objections against the imperative

¹ Oliver Sensen is Associate Professor and Director of Graduate Studies in Philosophy at Tulane University of New Orleans, USA. He is the author of *Kant on Human Dignity* (Walter de Gruyter 2011), the editor of *Kant on Moral Autonomy* (Cambridge University Press 2012), as well as the co-editor of *Kant's Tugendlehre* (Walter de Gruyter 2013) and *Kant's Lectures on Ethics: A Critical Guide* (Cambridge University Press, forthcoming).

² For a discussion of Kant's usage of these concepts see Sensen 2011, and 2012.

can be answered if one asks the question: Do I try to make an exception for myself to a rule that I regard to be *objectively necessary*?

SECTION 1: OBJECTIONS TO UNIVERSALIZING

Kant argues that the Categorical Imperative, “*act only in accordance with that maxim through which you can at the same time will that it become a universal law*” (GMS, AA 04: 421.07f)³, is the supreme principle of morality (cf. GMS, AA 04: 392.04). He also applies a formula of this imperative in four examples to show that what one commonly takes to be duties can be derived from the imperative (cf. GMS, AA 04: 421.18-423.35). But from early on the imperative has been criticized as being empty, i.e., as unable to generate any concrete duties on its own (cf. Hegel 1820, §135). This challenge has been made more specific over time, especially during the 20th century, when philosophers discussed whether Kantianism or Utilitarianism is the more plausible ethical theory. In particular, one can level three challenges at Kant’s principle: It seems that the imperative (a) rules out too much as being morally forbidden, (b) that it rules out too little as immoral, and (c) that it rules out things for the wrong reasons. In detail:

- (a) The standard way the Categorical Imperative has been applied is that one takes any proposed action, e.g., to play tennis at 10am on Sundays⁴, and asks whether everyone could follow that law. If everyone were to play at that time, however, then the court would be too crowded, and so no one could play. If the Categorical Imperative is a criterion for the moral permissibility of actions, then it is forbidden to play tennis at 10am on Sundays. By the same reasoning, almost anything would be ruled out as immoral, although it does not strike us as such. For instance, one could not universalize becoming a doctor, since if everyone became a doctor, no one would do other jobs which are vital. Humanity would die out, and no one could become a doctor. How could Kant have decided not to have children? For if everyone did that, humanity would die out, and no one could make the decision. Finally, to give another example, the Categorical Imperative seems to rule out that one is allowed to pay one’s credit card bill on time. Credit card companies can only exist because several people accumulate debts on their cards, and pay them back with high interest. If everyone were to pay back the credit card on time, the institution of credit cards would vanish.
- (b) The reverse problem is that the procedure seems not to rule out other actions which we would consider to be immoral. Should one refuse to accept a bribe, for instance (cf. Brentano 1889, 11 note 14)? But if everyone were to refuse a bribe, then the institution of bribery would vanish. This seems to lead to the counterintuitive conclusion that not only should one not refuse to accept bribes, it would be immoral to do so.

³ All translations of Kant’s work are taken from *The Cambridge Edition of the Works of Immanuel Kant*, ed. by Paul Guyer and Allen Wood.

⁴ Cf. Herman 1993, 138; the example is originally from Scanlon.

- (c) Finally, it is not clear that the procedure, even if it were to give the right results, would reach these results for the right reasons. What is *morally* bad about committing a contradiction? We would not, for instance, consider it to be immoral if someone committed a contradiction on a logic or math test.

SECTION 2: THE STANDARD KANTIAN RESPONSE

Kantians have not been without a response to these charges. The main problems in applying the Categorical Imperative arise if a maxim contains very specific references, such as what a particular person aims to do at a specific time and place. Could one universalize that Paul tells a lie in Lisbon on May 27th? If one universalizes this maxim, then no contradiction, and no implication for the institution of truth-telling is to be expected. Often a problem only arises if literally everyone engages in the action, but under a certain threshold no adverse effects occur. If everyone were to walk over the lawn of an Oxford college, the lawn would be ruined. But if only the fellows of the college are permitted to walk on the grass, then the lawn does not get damaged. In tying the maxim to a particular person and place, it does not yield a contradiction if one universalizes it.

A plausible Kantian response is to restrict maxims to the most general statement, the *maxima propositio*, the major premise, in a practical syllogism, or the general way to lead one's life in certain fundamental respects (cf. O'Neill 1989, 129; Höffe 1992, 90f). This way one precludes any references to particular persons or times and places. The moral questions would merely concern the most general statements: Should one help others, and develop one's talents; is false promising and killing under normal circumstances (not considering self-defense, for instance) morally permitted? One could not will a universal maxim of not helping, for instance, because this would prevent oneself from being helped. One could not even state a universal maxim of false promising without contradiction, because the other would know that one is lying, and would not believe one's promise. The textual evidence for this interpretation is Kant's definition of a maxim as a subjective practical principle in the *Critique of Practical Reason*. There he specifies: "Practical *principles* are propositions that contain a general determination of the will, having under it several practical rules." (KpV, AA 05: 19.07f) The maxim to be tested would then not be: "I want to play tennis at 10am on Sundays," but "I want to develop my talents." If this second maxim is morally permissible, there are now several ways of implementing it. One could do sports or play an instrument, for instance. If one does sports, one option is playing tennis, and one can play it at 10am on Sundays or at another time, and so forth.

I believe that this is a very plausible and promising Kantian response, and my own attempt will incorporate elements of it. However, one could object that Kant himself often specifies maxims in the more individualized sense (cf. Bittner 1974, and Timmermann 2000). In addition, it seems that we now do not just take what Kant says and apply it, but first think about the results we would like to get out of applying the imperative, and adjust the procedure accordingly. But then the Categorical Imperative is not our guide. Finally, one might be left with

the worry that even the new procedure does not rule out immoral actions for a recognizably moral reason. I will argue that Kant's text contains a further element to the universalization procedure, and that it is this element that brings out the moral quality as well as a more reliable guide to determining moral permissibility.

SECTION 3: UNIVERSALIZING AS FAIRNESS

If one reads the examples Kant gives right after introducing the main formulation of the Categorical Imperative, and the first sub-formula: "*act as if the maxim of your action were to become by your will a universal law of nature*" (GMS, AA 04: 421.18-20), one gets the impression that one can just universalize one's proposed maxim and see whether a contradiction occurs in order to derive concrete duties from the Categorical Imperative. This method faces serious objections, as I have outlined above. But Kant indicates that there might be something else to this procedure, and it is worth exploring whether this additional thought can avoid the objections. What is this additional step? Consider how Kant describes what is wrong with non-universalizable maxims:

If we now attend to ourselves in any transgression of a duty, we find that we do not really will that our maxim should become a universal law, since that is impossible for us, but that the opposite of our maxim should instead remain a universal law, only we take the liberty of making an *exception* to it for ourselves (or just for this once) to the advantage of our inclination. (GMS, AA 04: 424.15-20)

This passage introduces a further thought. What Kant wants to rule out with the Categorical Imperative is that one aims to make an exception. In the case of duties towards others, one wants them to adhere to a law, but wants to make an exception for oneself. In the case of duties towards self, one wants to make an exception to a law that one regards as valid. What Kant regards as immoral is therefore not a contradiction as such, e.g., a contradiction on a logic test. Rather Kant conceives of the contradiction as the means by which one can *find out* whether one aims to make an exception: "Consequently, if we weighed all cases from one and the same point of view, namely that of reason, we would find a contradiction in our own will, namely that a certain principle be objectively necessary as a universal law and yet subjectively not hold universally but allow exceptions." (GMS, AA 04: 424.20-5) The real moral question then becomes whether one wants to make an exception for oneself (duties towards others) or for this time (duties towards self).

How much is gained by the switch from contradictions to exceptions? It cannot be that every exception is ruled out as being morally wrong. For otherwise one does not seem to gain anything from putting Kant's procedure in terms of detecting exceptions. The same objections that can be raised against the contradiction tests can also be raised against the demand not to make an exception for oneself (if one just focuses on duties towards others): The requirement not to make an exception for oneself seems to rule out (a) too much, (b) too little, and (c) maxims for the wrong reasons. The procedure would still (a) rule out too much, since it would, for instance, give the same result regarding the maxim to play tennis at 10am on Sundays. The

reason for playing at 10am is that others would be at Church. If that is one's reasoning, then one would aim to make an exception for oneself by playing at 10am, and the action would be ruled out by the procedure. In addition, the requirement not to make an exception seems to rule out any form of competition in sports or at the workplace. If an athlete trains to be the best, would he not aim to be an exception? The demand also seems (b) to rule out too little. If it is true that everyone tells at least two lies per day, the person who contemplates telling a lie would not be making an exception for him- or herself. Finally, (c), since there are cases in which one makes an exception for oneself that do not seem to have a moral quality (the athlete, the spelling bee competition), the demand not to make an exception for oneself might also be said to rule out things for the wrong reason. So what is gained by interpreting Kant's universalization procedure in terms of exceptions?

The fact that looking for any exception as such does not provide a better procedure, does not rule out that Kant has a *particular* form of exception in mind, and that this special form of exception has a clearly recognizable moral quality, as well as that is more reliable in picking out what is morally wrong. What is this special form of exception? I shall later argue that this special form is already indicated in the quote above, when Kant says that one should not make an exception to a rule that one regards as "objectively necessary." (GMS, AA 04: 424.23f) But the thought will become clearer if one first looks at the second formulation of the Categorical Imperative, the Formula of Humanity, which commands to treat others never merely as a means, but always at the same time as an end in itself. Whether one can look at the second formula in order to shed light on the first, is a hotly debated issue in the literature,⁵ but Kant himself would not object to doing so. For he believed that the Formula of Humanity is "at bottom the same as the basic principle" and "tantamount" in its message (GMS, AA 04: 437.36-438.04). Furthermore, Kant recommends the second formula as bringing the main message "closer to intuition" (GMS, AA 04: 436.12f). Therefore I shall argue that looking at the second formula indeed sheds light on how the first is supposed to work. What is the requirement of the Formula of Humanity?

3.1 THE REQUIREMENT TO RESPECT OTHERS

In the *Groundwork* Kant presents at least five different formulations of the Categorical Imperative. The main Categorical Imperative (I), the Formula of Law of Nature (Ia), the Formula of Humanity (II), the Formula of Autonomy (III), and the Formula of the Kingdom of Ends (IIIa) (cf. Paton 1947, 129). Kant himself says that the Formula of Humanity, which commands: "*So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means*" (GMS, AA 04: 429.10-3) is "tantamount" to the Categorical Imperative:

to say that in the use of means to any end I am to limit my maxim to the condition of its universal validity as a law for every subject [Formula of Universal Law] is tantamount to saying that the subject of ends, i.e. the rational being itself, must be made the foundation of all maxims of actions,

⁵ For different reactions see O'Neill 1989, ch. 7; Wood 2009, 80-2; Allison 2011, 249-60.

never merely as a means, but as the supreme limiting condition in the use of all means, i.e. always at the same time as an end [Formula of Humanity]. (GMS, AA 04: 438.01-7)

In this passage Kant clearly expresses that he holds the commands of the Categorical Imperative and the Formula of Humanity to amount to the same thing. He explains why they are the same in the following passage from the *Critique of Practical Reason*:

every will ... is restricted to the condition of agreement with the *autonomy* of the rational being, that is to say, such a being is not to be subjected to any purpose that is not possible in accordance with a law that could arise from the will of the affected subject himself; hence this subject is to be used never merely as a means but as at the same time an end. (KpV, AA 05: 87.21-7)

The two formulas are tantamount because the Categorical Imperative demands that one could will one's law to be universal. But this means that the other has to be able to will the maxim as well. In restricting one's maxim by the will of another, one thereby respects the other as an end in itself. And, conversely, if one acts in accordance with the Formula of Humanity one would thereby also follow the Categorical Imperative. One can use the second formula to elucidate the first. But what is the clearer message of Formula of Humanity? The formula has two parts: On the one hand, one should never treat others as mere means, and, on the other, one should also treat them as ends in themselves. The first requirement is about the respect one owes others (and oneself). The respect one owes to others is not a feeling one may have for the merit of another – what Stephen Darwall has called “appraisal respect” (cf. Darwall 1977, 39) – but Kant specifies it as a maxim one should have (cf. MS, AA 06: 449.28). More concretely, it is the maxim “of not exalting oneself above others” (MdS, AA 06: 449.32). This is a popular moral idea: One should not regard oneself as something better, or as deserving more simply because one is oneself. Sure, one person might be better at a task at work, or at sports. A person might be able to type faster than another or run faster, but these are not ways in which one becomes more important as a person: One does not get two votes in an election, for instance. What the requirement to respect others amounts to is therefore that one regards oneself as one among equals, and treats others accordingly.

But this is only the first part of the Formula of Humanity. It is a negative requirement that expresses what one is not allowed to do, “and is thus analogous to the duty of right not to encroach upon what belongs to anyone.” (MS, AA 06: 449.32f) It concerns the respect for the life, freedom and property of the other, and restrictions on lying and false promises (cf. GMS, AA 04:430.3f). However, this requirement is not violated if one fails to help another. Take, for instance, the case of the bad Samaritan (cf. Parfit 2011, 226). If you pass a person in need, but fail to help him, you are not using him as a mere means. But it violates the second part of the Formula of Humanity. The bad Samaritan does not treat the needy person as an end in itself. This second requirement is about positive duties, such as helping. More specifically, it demands that I “make others’ *ends* my own (provided only that these are not immoral)” (MS, AA 06: 450.04f; cf. GMS, AA 04: 430.10-3). The complete demand of Kant’s Formula of Humanity is therefore that one should not exalt oneself above others, but treat everyone as being of equal, high importance.

3.2 UNIVERSALIZING AS A MORAL DEMAND

If Kant intends the Categorical Imperative and the Formula of Humanity to be tantamount, and if he recommends that the second formula can bring the message of the first closer to intuition, then it is the command not to exalt oneself above other, but to give everyone an equal, high importance, that he wants to express with the main formulation as well. Therefore one should construe the command of the Categorical Imperative, not to make an exception for oneself (in case of duties towards others), as the idea that one should not exalt oneself above others. How can the two demands be tantamount? One can elucidate it in the following way: If one regards oneself as something better, one is likely to make exceptions for oneself. Laws and regulations are for the lower people, oneself does not need to follow them. Conversely, if one makes exceptions for oneself, one is likely regarding oneself as something better. Others need to follow the rules, not oneself. The two formulas are then “tantamount” and “at bottom the same” (GMS, AA 04: 437.36-438.04).

This does not mean that there is no difference between the main formulation of the Categorical Imperative and the Formula of Humanity. The main formulation expresses the moral requirement from the perspective of the agent who deliberates what to do, while the second formulation expresses the same requirement from the perspective of the victim or person affected by the action (cf. O'Neill 1989, 141f). But, as Kant emphasizes, this difference is merely “subjectively rather than objectively practical” (GMS, AA 04: 436.11). The Formula of Humanity is more intuitively accessible, but the message is the same.

If this is correct, then Kant's procedure to derive moral duties is not just about universalizing maxims, and seeing whether a contradiction would occur, nor is it just about seeing whether one would be making an exception. Both of these tests are too wide, and run into the problems sketched above. Rather the question becomes whether one tries to exalt oneself above others, and whether one holds to more important from an impartial standpoint. What would be wrong is a form of arrogance, “in which we demand that others think little of themselves in comparison with us” (MS, AA 06: 465.11-3). It is a particular form of exception that Kant has in mind. What is morally wrong is not if one drives an alternative way to avoid the traffic jam, but if one drives on the emergency lane because one feels superior.

This answers one of the objections that are commonly raised against the Categorical Imperative (see Section 1 above). On the face of it, it is not clear why committing a contradiction should be a *moral* fault. But the particular form of making an exception just sketched is a recognizably moral idea. It is an idea of fairness, or of not being a free rider. One should not be selfish in that one holds others to standards to which one does not feel bound oneself. The presumption is that everyone has one voice, and is entitled to the same amount, unless there is a good reason for unequal treatment. One should not make an exception for oneself simply because one is oneself. This is a very popular idea. We react strongly if we feel that we have been treated unjustly, and we complain not just on our behalf, but even if we observe such a behavior towards a third person. The idea of fairness is also deeply engrained. It is more or less a human universal (cf. Darwall 2006, 175, 173), and even very small children hold moral rules to be universally valid (cf. again Nichols 2004, 5-7). The *moral* fault traced by Kant's test

does not lie in committing a contradiction, but in making an exception to a law one wills to be valid universally.

4. THE APPLICATION OF THE CATEGORICAL IMPERATIVE

So far I have argued against one of the objections against the Categorical Imperative. I have argued that the imperative expresses an idea which we do identify as a moral demand. But does this interpretation also answer the other objections? Is the demand not to exalt oneself, and not to make an exception to a rule one wills to be a universal law a reliable guide to derive concrete duties? Again, I shall first look at the Formula of Humanity, since it expresses the central idea in a more intuitive way. I shall then try to express the same idea in terms of the main formulation of the Categorical Imperative.

4.1 THE DERIVATION OF DUTIES WITH THE FORMULA OF HUMANITY

The central idea of the Formula of Humanity, I have argued, is that one should not exalt oneself, but regard everyone as being of equal, high importance. But this leaves open the exact procedure with which one can get from this general idea to more concrete duties. Does treating someone as an end in itself amount to respecting their decisions, e.g., if they agree to being thrown as a cannonball in a circus, or does it amount to not treating them in ways that are not befitting of human beings, irrespective of their decisions?⁶ How should one apply the Formula of Humanity?

There are different interpretations of the application of the Formula of Humanity. One suggestion is that the formula is about consent, and that one treats another as an end in itself if he or she can consent to the action. In the second example Kant gives in the *Groundwork* after the two formulas, he rules out making a false promise by saying: “For, he whom I want to use for my purposes by such a promise cannot possibly agree to my way of behaving toward him” (GMS, AA 04: 429.33-430.01). This suggests that the decisive question is whether another can consent to one’s proposed action. However, it seems that consent is not equivalent with all forms of treating someone as an end in itself, but that it only covers an important subset of treating others this way. For, first, consent does not seem relevant for duties towards self, and Kant, not surprisingly, does not mention consent in that context. In addition, second, it is not clear that consent is a reliable criterion for duties towards others.

One can distinguish between (a) actual, (b) hypothetical, and (c) possible consent as a moral requirement (see O’Neill 1989, ch. 6). Which of them should one employ? Each of them has its own problems as a moral guideline.

- (a) The most natural interpretation seems to be that others need to give their actual consent for something to be morally right. It seems that if someone does not agree to something, then one should not subject him to it. Conversely, one may hold that whatever one agrees

⁶ This is the case of Manuel Wackenheim.

to is morally right.⁷ But the problem is that people sometimes agree to less than is due to them, e.g., a freed slave might agree to wages which are far below what is labor is worth (cf. Williams 1973, 237). In addition, actual consent is dependent upon a proper description of the action. One might agree to some aspects, but not others (cf. O'Neill 1989, 106-9).

- (b) If it is not the actual consent of a person that is a reliable guide to what is morally right, one could reformulate the requirement by saying that it is the hypothetical consent that should count. It is not what a person actually agrees to, but what he or she would agree to if the person would be fully rational, for instance. But this form of consent is not fully self-explanatory. It depends upon a theory of what it means to be fully rational. A Utilitarian, for instance, might say that one could consent to being pushed from a bridge to stop a train that otherwise would hit five people (cf. Parfit 2011, 220, 190). The consent principle would then be compatible with a strong form of paternalism, where someone else decides what one could consent to, and what counts as the right thing to do would often violate the actual consent of the people affected. Paradoxically, hypothetical consent would be compatible with treating someone as mere means.
- (c) A third option is that the Formula of Humanity requires the possible consent of people affected by the action. One would have to give the other the actual power to agree or disagree in a way that coercion and deception deny this power. However, after an accident a person might be unconscious, and while he does not have the power to consent, this does not make the life-saving surgery morally wrong (cf. Parfit 2011, 190, 180). Furthermore, if one is in a life boat, and can either save one person drowning in one location, or five in another, but one cannot save both, it seems that one does not need to give the one the power to consent before rescuing the five.

What I have said so far does not mean that consent is not morally important. It often seems to be the decisive factor. If I drive your car, it seems that all that matters is whether I had your consent or not (at least in a non-emergency situation). But this leaves open the possibility that it is important under the guise of treating someone as an end in itself. If I am right that what the Formula of Humanity demands is that one does not exalt oneself but treat everyone as having equal, high importance, then consent will remain important. For treating others as ends in themselves would involve that one does not paternalistically try to govern their lives. In these areas, the consent of the affected is of crucial importance for how one is allowed to treat them, whether one can use their property or give medical treatment, for instance. But there are other areas of treating someone as end in oneself, e.g., duties towards self, where consent is less relevant.

The Formula of Humanity is then not just about consent. In addition, it is also not just about whether one *regards* the other as a mere means or an end in itself, but it is important that one *treats* them in this way. Otherwise the Formula of Humanity would not be a criterion for the moral rightness of maxims. Take, for instance, the example of the ruthless coffee buyer (cf. Parfit 2011, 216). The ruthless coffee buyer might regard the worker in the coffee shop as a

⁷ An extreme test case for this view is the cannibal of Rotenburg, Armin Meiwes, who killed and ate a person. If the victim consented to the action, does it make it morally permissible?

mere means. He would be prepared to kill the worker or steal the coffee without calms, but he finds that it would be easier just to greet friendly and pay for the coffee. In this case, the ruthless coffee buyer *regards* the worker as a mere means, but he *treats* him as an end in himself. The Formula of Humanity is not just a requirement about how one ought to regard people, but it is importantly also a demand for how one should *treat* them. The formula starts out by saying: “So act” (GMS, AA 04: 429.10).

This difference relates to Kant’s distinction between the legality and morality of an action (cf. KpV, AA 05: 71.34; MS, AA 06: 214.18). Kant says that there are two requirements for an action to have full moral worth: “For, in the case of what is to be morally good it is not enough that it *conform* with the moral law but it must also be done *for the sake for the law*” (GMS, AA 04: 390.04-6). One should do the right thing, and do it simply because it is right. The first is a requirement about the right behavior, the second is a requirement about the right motive. At an earlier stage of his career Kant had separated these two as two different principles, the *principium diiudicationis*, or the “principle of appraisal of obligation,” and the *principium executionis*, or “the principle of its performance or execution.” (V-Mo/Collins, AA 27: 274) In the *Groundwork* there is only one principle, the Categorical Imperative and its formulas, but the difference is whether one acts in accordance with the law, or if one also acts for the sake of the law and out of respect for it. One can stop at a red traffic light, but there is a difference whether one does it in order not to get a ticket or because it is the right thing to do. The Formula of Humanity is not just about the right motive, and how one should *regard* people, but it is also about how one should treat them. It is also a *principium diiudicationis*. So how can one derive duties from this principle?

I will confine myself to duties towards others. These will be of two kinds: One should not treat others as mere means, and should treat them as ends in themselves. The first is a negative duty. Imagine two farmers who are about to harvest their crops. By themselves it would take each farmer three days to bring in their products. But if they cooperate they could bring the harvest in one day each instead of three. Therefore one farmer asks the other to help him with his harvest, and promises to help the other the next day. If he does not show up the next day, without providing a strong excuse, we would judge that he treated the second farmer as a mere means. Similarly, in the case of positive duties, it is not just enough to regard the other as deserving assistance, but one would also need to provide the help.

The procedure provides a rough guideline, the a priori framework, for determining concrete duties; it does not determine all cases without further empirical knowledge: “The different forms of respect to be shown to others in accordance with differences in their qualities or contingent relations – differences of age, sex, birth, strength or weakness ... – cannot be set forth in detail and classified in the *metaphysical* first principles” (MS, AA 06: 468.06-12) One needs knowledge of human nature, of particular societies (cf. V-Mo/Collins, AA 27: 466), as well as the persons involved to fully determine when and how one would treat someone as an end in itself. For instance, should one hold open the door for an elderly person? It depends. If the person is weak and grateful for the help, it could be a sign of respect. But if the person regards himself as having full strength, and wants to prove it, holding open the door might

disrespect him. Should one speak slowly and in simple sentences to a foreigner? It depends. If the foreigner hardly speaks the language, he might be grateful for this treatment. But if the foreigner understands the language perfectly, he might feel belittled by the way one speaks to him.

Kant can merely provide the general procedure with which one should approach these questions: One should not exalt oneself above others, but treat them as having an equal, high importance. Does this shed light on how one can derive concrete duties from the main formula of the Categorical Imperative?

4.2 THE DERIVATION OF DUTIES FROM THE CATEGORICAL IMPERATIVE

The Categorical Imperative has been confronted with serious objections. One charge is that the imperative is empty, unable to determine concrete duties. It has been argued that it (a) rules out too much, (b) too little, and (c) rules out maxims for the wrong reasons. So far I have addressed (c). I have argued that Kant's Categorical Imperative expresses an idea which we clearly recognize as a moral idea. One should not exalt oneself above others, but regard them as being of equal, high importance. It is this idea, I argued, that Kant wants to express with the universalization procedure as well. But can this interpretation avoid the problems of (a) and (b)? How can the revised universalization procedure be used reliably to derive concrete duties?

I have argued that Kant understands the universalization procedure not just as the question whether a contradiction would occur if everyone had the same maxim. Nor is it simply about detecting exceptions, as this test too would be too wide (cf. Section 3). Rather the procedure tries to identify whether one would make an exception for oneself to a rule that one otherwise wills to be universal. This procedure has two levels. The intuitive idea is that one should not be a free rider to a general rule. This is the popular idea of fairness. But for Kant it is not about fitting in or promoting conformity. We often admire those people the most who went against the prevailing custom, and pointed out forms of discrimination that had been tacitly accepted. Kant's further point is that one should not make an exception to rules which one cognizes as being "objectively necessary." (GMS, AA 04: 424.23f) So one first needs to find out which rules are objectively necessary. Kant's procedure is then, in a second step, about finding out whether one would make an exception to these necessary rules: "we would find a contradiction in our own will, namely that a certain principle be objectively necessary as a universal law and yet subjectively not hold universally but allow exceptions" (GMS, AA 04: 424.22-5).

At first glance, this modified procedure should stand a better chance of avoiding the objections that the Categorical Imperative (a) rules out too much, and (b) too little at the same time. For instance, (a), the tennis player at 10 am does not regard going to church as objectively necessary, or he would do so as well. The athlete who tries to be the best does not regard being an average athlete to be an objectively necessary law; and the person who pays off his credit card bills on time does not regard having credit card debts to be a necessary rule. If that is correct, then the modified universalization procedure does not rule out too much, at least not the same

kind of actions the simple versions ruled out. Similarly, (b), the person who refuses to take a bribe does not regard taking bribes to be an objectively necessary practice. Instead, he would regard not accepting them to be a necessary law, and by refusing them he would not make an exception for himself to such a law.

There is still another sense in which one might think that the Categorical Imperative rules out too little. Every society will have special customs which are defining of its group. The Categorical Imperative might not rule out many of those special customs, and proponents of these customs might hold this against the imperative. But the modified universalization procedure can handle much better the standard examples that have been brought forth against the application of the Categorical Imperative. This leaves the general charge that the imperative is empty, and unable to derive any duties. The fate of this question hinges on successfully identifying the laws to which one should not make an exception. So how can one discover which rules are objectively necessary?

It is to this purpose that Kant provides the two forms of contradiction, the contradiction in the will and the contradiction in conception tests (cf. GMS, AA 04: 424.03-10). The contradiction in the will test declares certain laws of positive duties (developing one's talents and helping others) to be objectively necessary. I shall again focus on the duty towards others. A maxim of helping others is practically necessary because one would rob oneself of the means to an end one has *if* the maxim were to be universal. One should imagine that everyone had one's proposed maxim because this is the overall moral question: Could my maxim be a universal law, or am I making an exception for myself to a necessary law? If everyone had the maxim of not helping others, this would mean that I never would receive help from others. But as a finite being I have to assume that at some point I will be dependent upon the love and sympathy of others. Even if one is very rich and could buy the help of others, one has to anticipate that one needs the help of someone to find a caregiver and transfer the funds to him or her. So although such a world could exist in which no one ever helps the other, one could not will it. "For, a will that decided this would conflict with itself ... by such a law ... he would rob himself of all hope of the assistance he wishes for himself." (GMS, AA 04: 423.31-5)

Kant is not talking about a particular wish that some people have but not others. As a finite being everyone has to assume that he will need help sometimes. The contradiction is not between a desire that is relative from person to person and its fulfillment, but in the general structure of willing: "Whoever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power." (GMS, AA 04: 417.08-10) Since – as a finite being – I do will the end of assistance, I also *necessarily* want to be helped at some point. The law of helping others is necessary as a means to an end, and has the objective necessity of a hypothetical imperative. If the end is given, the means is necessary. But does this not violate Kant's assumption that morality consists in a *categorical* imperative, and that morality should be purely a priori? The assumption is not violated. What is categorical and purely a priori is the demand not to make an exception to a necessary law, or the Categorical Imperative itself. But Kant argues that the imperative "needs anthropology for its *application* to human beings" (GMS, AA 04: 412.04f). The anthropological knowledge in

this case is that human beings are finite, and have to assume that they will need help at some point. Given this end, the law of helping is objectively necessary as a means.

Kant says that another type of contradiction occurs for laws that express negative duties. He describes it as a contradiction in conception. The example of a negative duty towards others, that one must not give a false promise, is often considered to be the example which works best even on a very general universalization test that merely looks for an exception (and not also for an exception to a necessary law). If someone wants to borrow from another with the promise to pay it back, but in fact has no intention of paying back the money, this maxim – if universalized – would contradict itself. For if everyone had the maxim of making lying promises, the lender would know that the borrower has no intention of making payments, and would not lend the money. Such a universalized maxim cannot be stated coherently without a contradiction.

While on the surface, this is a clear case, it is harder to explain where exactly the contradiction occurs. For instance, is it (a) because the lender would not believe the one making the false promise, is it (b) because the institution of promise keeping would vanish, or (c) because such a universalized maxim would defeat the purpose of the borrower? But if the problem is (a), that the lender would not believe the borrower, the solution presupposes that the borrower always makes false promises. But if he makes the promise just if he is in dire need of money and does not know how to pay it back, the example only works if the lender *knows* that the other is in dire need. If the problem is supposed to be (b), that the institution of promise keeping would vanish, then this presupposes a further argument for why this institution is valuable on independent grounds, or it does not rule out that the institution of bribery would vanish if no one accepted bribes, and that therefore refusing bribes is morally wrong. So maybe the reason why one finds Kant's example plausible is because of (c), that the borrower defeats his own purpose (cf. Korsgaard 1996, 92-4). But if this is the case, is the contradiction in conception test not just the same as the contradiction in the will test?

On this reading, the laws of negative duties are objectively necessary because they protect what is necessary for living as such. Duties not to kill, maim, or steal are necessary to prevent the state of nature. In such a state one could not develop any arts or flourish (cf. O'Neill 1989, 20-4), but more importantly one is in constant danger of being killed. A man would not only be "stunted, bent and twisted" (IaG, AA 08: 22.32), but life would be "solitary, poore, nasty, brutish, and short" (Hobbes 1651, 89). One could object that this interpretation would make moral duties only hypothetically necessary, they would only be valid *if* human beings do not want to be killed. But Kant seems to think that moral duties are categorically necessary, independently of what one wants. This objection would be correct if what human beings want would be the only material out of which one tried to build duties. If one started from the fact that most people wanted to live, by itself this wish would not generate moral obligation: "the only way this maxim can be binding is through its qualification as a universal law" (MS, AA 06: 393.19-22). The generation of duties is already guided by the Categorical Imperative which commands that one treat everyone as being of equal, high importance. It already demands to treat others (and oneself) as important. The necessary laws only come in afterwards, in order

to specify more concrete duties. In this task too, one “needs anthropology for its *application* to human beings” (GMS, AA 04: 412.04f).

But even if this is correct, does it not mean that there is no difference between a contradiction in the will and a contradiction in conception? In both cases the agent does not get what he necessarily wants as a finite being (to receive help, and to live and flourish). There is a difference in that negative duties are more foundational. One needs to live in order to live comfortably. One could live (mostly) without receiving help from others, but one always needs not to be killed in order to live at all. The list can be expanded to other negative duties in that one might always need to be secure from being maimed, and preserve a sense of what is one’s own. But here too Kant does not believe that there is an a priori way to determine all laws. One will need empirical knowledge to derive the more specific rules.

In sum: I have argued that the Categorical Imperative expresses a recognizably moral idea: One should not be a free rider, in that one makes an exception for oneself to an objectively necessary rule. Kant would agree with his critics that *by itself* this rule is empty. It just gives the form of moral obligation: “through the law ... *in genere*, no rule of dutiful action can then itself be determined, because this belongs to the matter” (V-MS/Vigil 27: 578.10-6). The moral law here resembles the categories of the understanding in Kant’s theoretical philosophy: The Categorical Imperative without content is empty (cf. KrV B75). In order to yield concrete duties it needs a matter. This matter is provided by general anthropological knowledge. People want to flourish and want the means to their ends. These are the laws we recognize as objectively necessary for beings like us. It is to these laws that one should not make an exception. This reading, if correct, has the advantage that it combines what we find plausible in Kant’s moral idea (the overarching rule of fairness) with a human face. Kant agrees that in order to derive concrete duties, it needs anthropological knowledge. Kant’s moral theory does not cover everything. It does not, for instance, decide between permissible customs of particular societies. However, I take this to be a strength rather than a weakness. Why should one expect one person in 18th century Königsberg to rule once and for all what is right in all cases? Kant formulated the deeply-ingrained law of fairness. It is up to us to find, implement and live by the particular rules that are permitted by the Categorical Imperative.

CONCLUSION

I have argued that the Categorical Imperative expresses a requirement of fairness. This is an idea which we clearly recognize as being moral: One should not exalt oneself above others in trying to be a free rider. The imperative can be a guideline for decision procedures if one does not just look out for contradictions or any exceptions, but if one asks whether one would make an exception to a rule that we recognize as being objectively necessary for human beings.

ABSTRACT: Kant’s Categorical Imperative is commonly criticized as being empty, i.e., devoid of content, and therefore unable to generate any concrete duties. More specifically, the criticism is that the imperative rules out (a) too much as being morally forbidden in one respect, (b) too little in another, and (c) that it rules out maxims for the wrong reasons. In this paper I shall

argue that the Categorical Imperative expresses a recognizably moral idea, and consider how it can be used reliably to generate concrete duties. My claim is that the Categorical Imperative expresses the demand of fairness: One should not make an exception for oneself in the sense that one should not regard oneself as being something better. I shall argue that the main objections against the imperative can be answered if one asks the question: Do I (1) try to make an exception for myself (2) to a rule that I regard to be *objectively necessary*?

KEYWORDS: Kant, Categorical Imperative, Emptiness Objection, Derivation of Duties, Universalizing

REFERENCES

- Allison, Henry (2011), *Kant's Groundwork for the Metaphysics of Morals*, Oxford University Press.
- Bittner, Rüdiger (1974), "Maximen" in *Akten des 4. Internationalen Kant-Kongresses*, ed. by G. Funke, Berlin: de Gruyter, 485-498.
- Brentano, Franz (1889), *The Origin of our Knowledge of Right and Wrong*, ed. and tra. by R. Chisholm and E. Schneewind, London: Routledge, 1969.
- Darwall, Stephen (1977), "Two Kinds of Respect," *Ethics* 88, 36-49.
- : (2006), *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.
- Hegel, Georg Wilhelm Friedrich (1820), *Elements of the Philosophy of Right*, tra. by H.B. Nisbet. Cambridge University Press, 1991.
- Herman, Barbara (1993), *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press.
- Hobbes, Thomas (1651), *Leviathan*, ed. R. Tuck, Cambridge University Press, 1996.
- Höffe, Otfried (1992), *Ethik und Politik*, Frankfurt: Suhrkamp.
- Korsgaard, Christine (1996), *Creating the Kindgom of Ends*, Cambridge University Press.
- Nichols, Shaun (2004), *Sentimental Rules*, Oxford University Press.
- O'Neill, Onora (1989), *Constructions of Reason*, Cambridge University Press.
- Parfit, Derek (2011), *On What Matters*, Vol. I, Oxford University Press.
- Paton, H.J. (1947), *The Categorical Imperative*, London: Hutchinson, 1970.
- Sensen, Oliver (2011), *Kant on Human Dignity*, Berlin: de Gruyter.
- Sensen, Oliver (ed.) (2012), *Kant on Moral Autonomy*, Cambridge University Press.
- Timmermann, Jens (2000), "Kant's Puzzling Ethics of Maxims," *The Harvard Review of Philosophy* 8, 39-52.
- Williams, Bernard (1973), *Problems of the Self*, Cambridge University Press.
- Wood, Allen (2009), *Kantian Morality*, Cambridge University Press.

Recebido em / Received in: 22.10.13

Aprovado em / Approved in: 15.2.14;

