

Kant's Compass

Author(s): Jordan Howard Sobel

Source: *Erkenntnis* (1975-), Vol. 46, No. 3 (May, 1997), pp. 365-392

Published by: Springer

Stable URL: <http://www.jstor.org/stable/20012773>

Accessed: 24-05-2016 16:35 UTC

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Erkenntnis* (1975-)

JORDAN HOWARD SOBEL

## KANT'S COMPASS

**ABSTRACT.** Can I will that my maxim becomes a universal law? . . . It would be easy to show how common human reason, with this compass, knows well how to distinguish . . . what is consistent or inconsistent with duty. (Kant, *Foundations*, 403–4)

How exactly is this compass to work? Cases bring out connected difficulties to do, (1), with whether 'social contexts' are to be in or out of descriptions of actions maxims would have agents do – for example, 'disarming *alone*' and 'voting *when enough others would even if one did not*', or 'disarming' and 'voting' simply; and, (2), with a seldom noticed ambiguity of 'everyone's acting in accordance with a maxim' and 'a maxim's becoming universal law'. The paper argues dilemmatically for the inadequacy of Kant's test for maxims consistent with duty whatever policy for social contexts and manner of maxims becoming universal laws it is said to invoke.

### 1. INTRODUCTION AND PRELIMINARY REMARKS

#### 1.1. *The Compass*

The moral law is said to be "comprehensible . . . to the understanding" (*CPrR* 162).<sup>1</sup> Kant maintains that

human reason, even in the commonest mind, can easily be brought to a high degree of correctness and completeness in moral matters. . . . (391)

Elaborating he writes:

I do not . . . need any penetrating acuteness in order to discern what I have to do in order that my volition may be morally good. Inexperienced in the course of the world, incapable of being prepared for all its contingencies, I ask myself only: *Can I will that my maxim becomes a universal law?* . . . It would be easy to show how common human reason, with **this compass**, knows well how to distinguish what is good, what is bad, and what is consistent or inconsistent with duty. (403–4, bold emphasis added)

The compass is featured in the *Foundations* of 1785 – it is the principle of morality in its first, universality formulation (cf., 436):

There is . . . only one categorical imperative. It is: Act only according to that maxim by which you can at the same time will that it should become a universal law. . . . The universality of law according to which effects are produced constitutes . . . nature in the most general sense . . . , i.e., the existence of things so far as it is determined by universal

*Erkenntnis* 46: 365–392, 1997.

© 1997 Kluwer Academic Publishers. Printed in the Netherlands.

laws. . . . [Thus] the universal imperative of duty can be expressed as follows: Act as though the maxim of your action were by your will to become a universal law of nature. (421)

Something similar is in the *Critique* (second) of 1788 a rule for good and bad actions.

The rule of judgment according to laws of pure practical reason is this: ask yourself whether, if the action you propose were to take place by a law of the system of nature of which you were yourself a part, you could regard it as possible by your own will. Everyone does, in fact, decide by this rule whether actions are morally good or evil. Thus, people say: . . . if you belonged to such an order of things, would you do so with the assent of your own will? (CPrR 69, Abbott translation, emphasis added.)

And much the same surfaces in *The Metaphysics of Morals* of 1797:

The categorical imperative . . . is: Act upon a maxim that can also hold as a universal law . . . [y]ou can know whether [your maxim] holds objectively only in this way: That when your reason subjects it to the test of conceiving yourself as also giving universal law through it, it qualifies for such a giving of universal law. . . . [This is t]he supreme principle of the doctrine of morals. . . . (*MOm* 225; cf., 389, 392, and 453.)

### 1.2. *It is Primarily for Identifying Good and Bad Maxims*

Onora O'Neill says that the Categorical Imperative is a "test . . . for the moral acceptability of acts", a test that implies the injunction always to "*act on a maxim*" rather than to engage in "mere reflexes or reactions" (O'Neill 1989, p. 83) or (I assume) spontaneous acts that are not pursuant to plans. She may imply a somewhat different view in a chapter 'written much later' (126n), in which she indicates that for Kant all actions 'maximed': "Kant sees action as undertaken on certain principles . . . *maxims*" (129). If all actions are maximed, and especially if this is by definition, it is idle to enjoin that they be so.

Differing from O'Neill, I take Kant's compass to be offered primarily and directly as a test for good and bad, in the sense of consistent and inconsistent with duty, maxims and volitions (402–4 above),<sup>2</sup> not for good and bad, or right and wrong, actions. Furthermore, though the compass is supposed to be relevant to actions (consider CPrR69), it is not clear how this is to be. Perhaps we can attribute to Kant the view that an action is right in a situation if and only if it could issue from knowledge of the situation conjoined with a maxim for it that the compass would identify as good, and that an action is wrong in a situation if and only if it is not right in it. Actions could then be said to be good if and only if both right and pursuant to good maxims, and bad if and only if both wrong and pursuant to bad maxims. These conjectures would allow attributions of right and wrong, though not of good and bad, even to unmaximed actions. I do not read into Kant's

willing-to-be-universal-laws formulation of the categorical imperative an injunction to engage only in maximed actions. Nor am I persuaded that he was of the view that actions properly so called are necessarily maximed. In any case, leaving open several issues concerning actions and maxims, I shall in what follows consider relations of the compass only to maxims and volitions.

### 1.3. *Maxims*

A maxim is the subjective principle of acting . . . [and] contains the practical rule which reason determines according to the conditions of the subject (often its ignorance or inclinations) and is thus the principle according to which the subject acts. (421n)

Regarding the nature of maxims or subjective principles of action, I follow Schneewind, who says that maxims should be understood to be *plans or intentions*, possibly contingency plans, either particular or more or less general, that incorporate agents' subjective reasons for actions pursuant to them.

A maxim is a personal or subjective plan of action, incorporating the agent's reasons for acting as well as a sufficient indication of what act the reasons call for. When we are fully rational, we act, knowing our circumstances, in order to obtain a definite end, and aware that under some conditions we are prepared to alter our plans. . . . A full maxim . . . makes all this explicit. (Schneewind 1992, pp. 318–9)

Cf.:

To say that someone is acting on a certain maxim is to imply (if not to say) that he is acting for a certain purpose, or with a certain end in view, or with a certain intent; and to specify the maxim is to specify the purpose or intent of the action, as well as something of its circumstances. (Singer, 1961, p. 244.)

Kantian maxims, or 'subjective principles according to which subjects actually act' (421n), would, I assume still following Schneewind, incorporate precisely features that are for the agent relevant to his reasons for actions.

When we use the categorical imperative . . . we suppose that we are examining a maxim embodying the agent's genuine reasons for proposing the action, rather than [what are in the agent's view] irrelevancies . . . that might let it get by the categorical imperative [or, I add, that might let the categorical imperative stop it]. (Schneewind, 1992, p. 321.)

Christine M. Korsgaard says that "a properly formulated maxim" must not only specify a plan to act but the reason to act as planned: "Since the Formula of Universal Law is a test of the sufficiency of reasons, the maxim must include them" (Korsgaard, 1985, p. 24; see pp. 25 and especially 36). Barbara Herman says that "a maxim should provide a description of an

action (a proposed, intended action) *as* purposive voluntary activity initiated for the sake of an end . . . descriptions of an action for a maxim are . . . sufficient to justify the action from an agent's point of view. Correlatively, no description may be included in a maxim unless it fits the agent's conception of her action as willed. . . ." (Herman, 1993, p. 144).

Cases will be considered in which formulations of an agent's maxims or subjective principles for situations make explicit everything, including ends, that is for him relevant to what to do in this situation. I assume that maxims ready for the discipline of the compass should include precisely that. But must good and rational agents have somehow in mind maxims that 'make all of this explicit'? Moral particularist H. A. Prichard, who viewed Kant's compass as a paradigm of philosophy that rests on the mistake of supposing that proofs of any kind are ever possible for particular perceptions of duty, would deny it. I leave this question open and, without maintaining that there is a job that Kant's compass might do for good and rational agents, consider only whether supposing that there is such a job this compass is well-designed for it.

#### 1.4. *A Problem to be 'Bracketed'*

A problem I will not go into concerns the relation of the compass to Kant's realm-of-ends conception of morality. Morality according to that conception is that legislation that would, were its several rules universally obeyed, realize a realm of ends. Implied by this conception is a test for moral rules in which *whole systems* of rules are compared with an eye to that 'glorious ideal' (462). The compass, in contrast, would have maxims tested *one by one* for consistency with morality (duty, supposing "subjective restrictions and hindrances" 397), the test being, of course, whether or not a given maxim can be willed to become a universal law. But maxims that are *individually willable* as universal laws with an eye to that glorious ideal (which I note is not explicitly mandated in formulations of compass), might not be such that universal obedience to them all would realize a realm of ends, or therefore such that they are *collectively willable* with that glorious end in view. Maxims that are individually willable as universal laws with that end in view need not even be collectively *obeyable* as universal laws, let alone collectively willable. It is, I think, very doubtful that Kant's compass and his realm-of-ends conception of morality "are fundamentally only so many formulas of the very same law" (436). It is likely, I think, that these two "ways of presenting the principle of morality" (436) are not even consistent.

So much for this problem regarding the relation of Kant's compass to his realm of ends conception of morality. My concern in this paper is

not with the place of this compass in his system, but with the compass itself. I shall argue that if it works, it is at least not the simple device Kant supposed. Two cases taken together point to new difficulties that say that to use this compass one would need a better manual than Kant provides, of the production of which, for reasons to be indicated, I despair.

## 2. TROUBLESOME CASES

The first case features an apparently good maxim that, depending on exactly how the compass works, may be found by it to be morally bad and not consistent with the law. The second case features a maxim that one assumes Kant would condemn, which maxim the compass, depending how it works, may pronounce to be morally good and consistent with the law. Together, as will be indicated in Section 4, these cases point up a dilemma concerning exactly how Kant's compass could work for him.

## BRIDGES

### 2.1. *One Bridge*

Suppose it is common knowledge for several persons including me that:

Our travels through rough country have brought us, who are equal in weights, to a footbridge of precarious appearance strung across an awful gorge. We must, if we can, cross on this bridge. Those who do not cross will starve to death. Though we can all together move onto the bridge (it is wide enough), it would not long bear the weight even of any two of us. The only question is whether it will bear our weights singly. We are each considering going first in the interests of all – we are each considering venturing onto this bridge to demonstrate that it will hold under each of us if we go one at a time, or that it will not hold any of us. Of each it is true, that if he does not come forward no one else will, and that if he does he will be the only one to do so.

For definiteness, let it be common knowledge that my situation as I consider whether to try to cross or to hang back is the same as every other person's at this juncture, and that it has the following shape in which are indicated salient consequences for me, and in parentheses for all, if my options in each of two possible circumstances that are exclusive and exhaustive of possibilities. Numbers indicate orders of my preferences both for consequences for me, and for consequences for all without particular regard to me – as it happens I order alike personal consequences, and consequences for all without particular regard to which are mine.

Possibilities for the bridge.

	It will hold me.	It will not hold me.
Try to cross.	Cross safely. 1 (All others cross safely one by one after me.)	Crash to immediate death -1 # (All others starve to eventual deaths.)
Hang back.	Starve to eventual death. 0 # (All others starve to eventual deaths.) *	Starve to eventual death. 0 # (All others starve to eventual deaths.) *

(#Why -1 less than 0? Because I consider crashing to immediate death to be worse for a person than eventual death by starvation. \*Why would we all starve? Because if I do not test the bridge, no one else will.)

What has Kant's test to say about the maxim on which I am considering acting? That depends of course on what exactly my maxim is. Let it be the following relatively specific general maxim or plan:

In circumstances such as these, when if I do not do a necessary-for-the-good-of-all risky deed no one else will do it and if I do do it I will do it alone, to be brave and in the possible interests of all to do the deed.<sup>3</sup>

That seems to be not only a reasonable maxim from the standpoint of my interests, but a laudable and morally worthy maxim. Can I, however, as a knowledgeable member of this traveling troupe, will that this maxim should become a universal law? For one reason or another, depending on what we say this law would have me *do* in this case, it can seem that I cannot.

What exactly *would* the maxim have me do – what is that necessary deed? Suppose the 'necessary deed' intended is *to venture onto the bridge in an attempt to cross on it*, or for short but with the same meaning, *to try to cross on the bridge*. Suppose that precisely this is what the maxim would have me do. Then my maxim can become a universal law: it is possible for the acts it would have done all to be done together – we can all together try to cross on the bridge. But I cannot consistently with my end in the maxim will that these acts all be done, For it is given that I know that with everyone on it the bridge would collapse and we would all crash to 'immediate deaths, and that is contrary to the stated-in-it objective of my maxim which is the possible interests of all – there is, recall, the possibility of our all crossing safely. And so, taking the deed called for by the maxim to be trying to cross the bridge, one can say that it "would conflict with itself if it were made a universal law" (*MofM* 393).

Suppose, however, that the deed called for by the maxim is not merely to try to cross on the bridge, but *first and alone before anyone else to try to cross the bridge*.<sup>4</sup> Suppose that this is what we should, for purposes of an application of the compass, say the maxim would have me do. Then it seems that this maxim cannot even become a universal law, so that the issue of whether it can be willed as such does not arise. For there are *several* persons in its circumstances, each of whom can first and before everyone else try to cross on the bridge, and it is not possible that *each and every* of several should together do *anything* first and alone and before anyone else.

So in one way or another – depending on what my maxim, properly framed for testing, would have me do, and whether the ‘social context’ (what others are and would be doing) is properly part of what it would have me do – it seems impossible for me to will that my maxim should become a universal law. And yet, as stated, there seems nothing wrong with my maxim, however it is spelled out regarding the deed it contemplates. The willingness it must reflect to take on risks in the interests of all – it would have me sometimes say, “It is dangerous work, but someone has to do it” – far from being bad and condemnable and *inconsistent* with duty, is, one would have thought, very good and praiseworthy and even *beyond* duty.

N.B. I have written ‘*seems impossible*’ rather than ‘is impossible’ because there is a question whether the ‘logic for maxims becoming universal laws’ I employ here is the *right* logic. The form of this ‘logic’, and of an alternative ‘logic’ that I use in the election case of Section 2.5 below, are gone into in Section 3 below. That alternative logic can be made to work *well* to a good result in the present case. The *problem* I am starting here is that neither ‘logic’ works well to good results in *both* cases, and – Section 4 below – that there indeed may not *be* a right logic for ‘maxims becoming universal laws’ in Kant’s compass. I beg the reader’s patience.

## 2.2. *A Possible Problem with This Case – Strains of Ethical Egoism in Kant*

I court the problem to be discussed in this Section, 2.2, because it leads into the interesting territory of Kant’s egoistic tendencies, and notwithstanding that, as indicated in the next Section, 2.3, this problem is easily finessed by a small modification in the case.

2.2.1. There are aspects of the bridge case that can occasion qualms in the light of some Kantian texts. To begin, the reasonableness – asserted in passing – from the standpoint of private interest of its maxim can be challenged. For this maxim can conflict with the general prudential maxim to maximize ‘expected personal utility’. Taking numbers in the

bridge case matrix to be cardinal measures of personal values for possible consequences, trying to cross would maximize only if the probability that the bridge will hold me is not less than the probability that it will not. But, even if, as I am prepared to assume for argument, probabilities are such that this maxim is not perfectly reasonable from the standpoint of private interest, it seems reasonable enough from that standpoint still to qualify as morally good and consistent with duty.

I assume that the moral law, as understood by anyone who believes in it would sometimes constrain the pursuit of private interest and the maximization of expected personal utility. I assume that the moral law would sometimes call for self-sacrifice in the sense of nonmaximizing of expected personal utility actions. It seems that it might sometimes constrain for the good of others pursuit of private good, or maximization of expected personal utility. It seems that it might sometimes at least *permit* self-sacrifice in this sense for the good of others.

2.2.2. I am grateful to Edwin Mares for comments that suggest that Kant might demur, that he might say that moral laws lay down only imperfect duties to act in the interests of others, and that they never generate duties overall to act for others when so acting is contrary to the agent's interests and in this sense self-sacrificial. Mares cites the text:

a maxim of promoting others' happiness at the sacrifice of one's own . . . would conflict with itself if it were made a universal law (MofM 393).

Mares thinks that Kant might condemn the maxim of my bridge case as morally bad because it can, depending on probabilities, call for nonmaximizing of expected personal utility actions. Mares thinks that Kant might say not only that self-sacrifice merely for the good of others is never a duty, but that it is always inconsistent with duty and never even permitted.

Let me confess surprise at Kant's just quoted dictum. One wonders why Kant thought that any maxim to sacrifice one's own happiness for that of others would conflict with itself if it were made a universal law. It seems that universal obedience to a law that called for sacrifices for greater gains for others could be in everyone's interest. Should we say that Kant made a mistake here?

2.2.3. Does Kant think there is for a man who is 'weary of life' (421) a duty not merely *not to shorten* one's life, but to pursue happiness and maximize expected personal utility, that is, to *improve* one's life? There is some evidence that he does. There is some evidence that in his view there is such a duty, though not, as the duty not to shorten one's life is, a strict or perfect duty. He writes that "[t]here can be no imperative which would,

in the strict sense, command us to do what makes for happiness” (418). Kant’s reason for this cannot be that happiness is not possible for purely rational beings who would be devoid of inclinations: if that were his reason then preserving one’s life when one can end it would not be a duty because that is not possible for immortal rational beings. His reason is rather that, as he writes, “the task of determining infallibly and universally what action will promote the happiness of a rational being is completely unsolvable” (419).

If Kant recognizes an imperfect duty to pursue happiness, then Mares has documented a curiosity, namely, that Kant is inclined to say that this duty always trumps another imperfect duty, specifically, the imperfect “meritorious duty to others . . . [to] endeavor, so far as [one] can, to further [their] ends [and to contribute to their happiness]” (430). Thus stated, incidentally, this latter duty to others goes far beyond the requirement merely to help others “to struggle with great hardships” (423), and is in fact a general duty of utilitarian benevolence in which (as R. M. Hare would have them be) “the ends of any person . . . [are] as far as possible [made one’s own ends]” (430).

2.2.4. Kant, I note, is ambivalent regarding self-sacrifice for others. Consider:

The action by which a man endeavours at the greatest peril of life to rescue people from shipwreck, at last losing his life in the attempt, is reckoned on one side as duty, but on the other and for the most part as a meritorious action, but our esteem for it is much weakened by the notion of *duty to himself*, which seems in this case to be somewhat infringed. (CPrR 159, Abbott translation)

If the duty to others is only a meritorious duty, and the duty to preserve one’s own life is a strict duty, then the latter should quite override the former. And yet Kant’s words suggest that the latter duty may sometimes merely detract from the former one, so that the upshot in the case is still a duty net, albeit weakened, to attempt to rescue even at the probable cost of one’s life. Perhaps Kant would say that it makes all the difference whether one sacrifices oneself, which *violates* the strict duty to preserve oneself and is never permitted in the service of a meritorious duty, or risks sacrificing oneself, which merely *somewhat infringes* that strict duty and is sometimes permitted in the service of a meritorious duty. (On the need for the qualification, ‘in the service of a meritorious duty’, see CPrR 30, where it is implied that the duty not to make a false deposition sometimes/always? overrides the duty not to sacrifice oneself when threatened with certain death unless one makes a false deposition. Thanks to Marcia Baron for this reference.)

In any event, it seems from this passage in his *Critique of Practical Reason* of 1788, that a maxim to attempt to save the lives of others at a risk to one's own life would *not* be said by Kant to 'conflict with itself if it were made a universal law'. This makes further surprising his intrinsically surprising assertion in *Metaphysics of Morals* of 1797 that "a maxim of promoting other's happiness at the sacrifice of one's own . . . would conflict with itself if it were made a universal law" (MofM 393).

### 2.3. *Another Bridge*

Problems raised by Mares for my bridge case can be finessed for purposes of the point that case would serve, which is that the compass, depending on how it works, may condemn a maxim for reasons that have nothing to do with this maxim's sometimes calling for self-sacrifice in the sense of nonmaximizing actions. One way to avoid these problems and the need to deal with perplexing egoistic strains in Kant's philosophy, is to change the agent's situation. So I say that tigers will get me if I do not cross the bridge, and assume:

#### Possibilities for the bridge.

	It will hold me.	It will not hold me.
Try to cross.	Cross safely. 1 (All others cross safely.)	Crash to immediate death 0 (All others are eaten by tigers.)
Hang back.	Be eaten by tigers. 0 (All others are eaten by tigers.)	Be eaten by tigers. 0 (All others are eaten by tigers.)

Trying to cross, given the indifference now stipulated between crashing to immediate death and being eaten by tigers, maximizes expected personal utility *regardless of probabilities*. It maximizes uniquely if the bridge's holding me has positive probability. Otherwise it ties with hanging back. So trying to cross is never in situations of the kind now being considered a self-sacrifice in the sense of a nonmaximizing of expected personal utility action. To complete the finesse, amend the maxim on which I would act to:

In circumstances such as these, when, (i), if I do not do a necessary-for-the-good-of-all risky deed no one else will do it and if I do do it I will do it alone, and, (ii), I can do it without self-sacrifice, to be brave and in the possible interests of all to do that deed.

#### 2.4. 'What if Everyone Became a Doctor?'

Before moving on let me note that the place in my discussion occupied by one or another bridge case could be taken by a doctor case in an isolated community of a thousand for which there is no possibility of outside help. Assume this community would be best served by exactly one doctor, and that we can say to each of its members, "You can be the best that you can be, and do best for your community by being a doctor." Let the ground for this public good be, in the case of each community-member, in part that if he/she *does not* become a doctor for the community *no one else in it will*, and in part that if he/she *does* become a doctor *no one else in it will*.<sup>5</sup> A laudable maxim for a member of this community, albeit a maxim that arguably could not be willed to become a universal law, would be:

In circumstances such as these, when, (i), if I do not practice a profession not enough others will practice it, and if I do practice it not too many others will, and, (ii), I can be the best that I can be and do best for my community by practicing it, to practice it.

The problem with universalizing this maxim in this case (if one uses the 'logic' for maxims becoming universal laws that I have used in my bridge cases), is that (I now adapt lines drawn in Section 2.1 above) in one way or another, depending on what we say is the deed called for by this maxim, I, as a knowledgeable member of this community, cannot will that it should become a universal law. Which way depends on the scope of the infinitive phrase of the displayed formulation of the maxim. If the circumstance that – if I do not, not enough will, and if I do, not too many will – is part of that deed, then "the objection [is] that not everyone could do this" (Korsgaard 1985, p. 46n23).

#### AN ELECTION

#### 2.5.

Suppose that even though I am always ready to vote when my vote would matter and is needed, I am considering not voting in an important election because I realize that a good turnout is assured whether or not I vote, that one more vote would neither enhance the good effects of the assured good turnout, nor affect the outcome, and that I can use the time to do some private and/or public good. I realize that my case is not special. I realize that no one's vote is needed and that everyone, those who will vote and those who will not, has better things to do than vote. Would my volition not to vote be bad? Many would say so and say that, as it is everyone's, so it is my *duty* to vote. We expect Kant's compass to agree. My leanings, after

all, occasion the textbook ‘generalization argument’ challenge, “But what if no one votes who has better things to do?” I must want many persons to vote in this important election. I must want the good turnout of which I am confident. And we expect Kant’s compass to condemn my temptation to “take the liberty of making myself an exception” (424). But does it?

To find out we must fix the maxim on which I am considering acting. Let it be:

To vote when my vote would matter and is needed, as well as when I have nothing better in terms of private or public good to do, and otherwise to refrain from voting (and so to refrain from voting when a good turnout is assured whether or not I vote and I can use my time to better private or public purposes).

That would be a maxim of the contingency plan sort that incorporated reasons, those that tempt me not to vote in the case, as well as those that would move me to vote in other cases. Can I will that this maxim should become a universal law?

For an answer, consider first what it could be in an election-case for this maxim to be a universal law, what it could be for each person actually to be ‘moved, as by a law of nature, to do as it would have one do’ (Schneewind, 320). That it seems could be a *mixed* situation in which persons in numbers sufficient for a good turnout vote, while other persons who have better things to do do these better things rather than vote. In such a situation each person could be acting in accordance with that maxim, some by voting, and others by doing better things. The maxim I am considering can in this way become a universal law of nature. And there are no obvious barriers to my *willing* that it should become a universal law. The maxim implies interests on my part in private and public goods, and consistently with these interests, the situation just described would, by implicit hypothesis, maximize private and public goods though not necessarily any particular person’s private good. Some voters in that situation could be voting at personal costs because their votes matter and are needed.

And so it seems that Kant’s compass does *not* condemn my temptation to take advantage of the good citizenship of others, my temptation to make an exception of myself and not vote. It seems that his compass does *not* condemn what one supposes he would think was the *bad* maxim displayed above that I have stipulated is my maxim.

## 2.6. *False Promises*

For its purposes in the paper, the election case could be replaced by a false promises case in which the following condemnable maxim could, arguably be willed to become a universal law:

To make false promises for personal advantage, when personal advantage can be secured by false promises.

Were each person to be 'moved as by a law of nature to implement this maxim', it seems that precisely the number of false promises for personal advantages that the system can support would take place, no more and no less, and there is no obvious barrier to a person's willing that. In particular, willing that should not be a problem for someone who makes that maxim his maxim: arguably the expected gain for him in "[a] world of the universalized maxim" (Korsgaard 1985, p. 36) should exceed that in an otherwise similar world in which possibilities for gain by false promises are never capitalized.

## 2.7.

Let me relate elements of the discussion in Section 2.5 to certain maxims, some *prima facie* objectionable and some *prima facie* innocent, in cases considered by Herman and Korsgaard.

Korsgaard writes:

in the false promising case, the difficulty is that the man's end – getting the money – cannot be achieved by his means – making a false promise – in the world of the universalized maxim. The efficacy of the false promise as a means of securing the money depends on the fact that not everyone uses promises this way. (Korsgaard, 1985, p. 36; see Herman, 1993, p. 118n5, wherein this line of Korsgaard's is cited with approval.)

However, "in [a] world of the universalized maxim" to make false promises to get money – which, spelled out, must be the maxim to make false promises to get money *when and of course only when one can get money by making a false promise* – not everyone always uses promises this way. In that world, though everyone is always *ready* to use promises in this way, just so many promises are used in this way as the system will tolerate.

Similar difficulties attend claims made in (Herman, 1993, p. 138) concerning the maxims,

to save money by shopping in this year's after-Christmas sales for next year's Christmas presents,

and

to play tennis Sunday morning at 10:00 when neighbors are in church and the courts are not crowded.

Herman thinks that 'practical, contradictions-in-the-will' tests condemn these maxims. Implicit in her reasons is that in worlds in which these maxims were as if universal laws of nature, everyone would shop after Christmas for next Christmas (so there would not then be sales and money

to be saved, for it would then be next year's Christmas shopping season), and everyone would play tennis Sunday morning at 10:00 (so that the courts would then be crowded). However, these maxims, spelled out, have 'when and of course only when one *can* save money by shopping in this year's after-Christmas sales for next year's Christmas presents' and 'when and of course only when one *can* play tennis Sunday morning at 10:00 when neighbors are in church and the courts are not crowded'. In worlds in which there is universal conformity to these maxims, *some people but not all* will save money by shopping this year's sales for next year's Christmas, and *some people but not all* will play tennis Sunday mornings on uncrowded courts. The points, the purposes, that are plain in these maxims would *not* be undermined were they to be in this way as if universal laws of nature.

### 2.8. *Problems Posed by These Troublesome Cases for the Compass*

These cases – the bridge and elections cases (and their kin) – challenge the reliability of the compass. The *bridge* case presents an apparently laudable maxim that in one way or another, depending on what we say it would have agents do, seems to fail Kant's test. The *election* case presents a maxim that many would condemn as bad that seems to pass Kant's test.

More importantly, however, these cases point up problems with *exactly* how Kant's compass is supposed to work. They bring out that it is at best more difficult to use than he suggests. One area of difficulty concerns *social contexts* indicated by words such as 'first and alone' and 'when enough others will'. The place, if any, of social contexts, in what the compass would have maxims on becoming universal laws have agents doing, needs to be regimented.<sup>6</sup> The second connected matter that needs to be settled, to which I devote the next major section of this paper, concerns the '*logic*' of a maxim's '*becoming a universal law*' as the compass says that one should be able to will for one's maxims (403–4, quoted in Section 1.1 above). To the best of my knowledge, this second area of difficulty for Kant's compass has received no prior attention in the voluminous literature on Kant's compass, none at all.

## 3. WAYS OF MAXIMS BECOMING UNIVERSAL LAWS

Let me begin with introductory words of mine from 'Everyone's Conforming to a Rule'.

It has been said that in order for a rule to be a moral rule or a principle of rational action it must be possible consistently to will it as a universal law. . . . One problem with being clear about such ideas . . . is that there are a number of distinct ways in which universal

conformity to a rule, or everyone's acting according to it, can be understood, and universal conformity in one sense to a rule may be possible [and consistently willable] while in another sense or even in all other senses universal conformity to it would be impossible [or not consistently willable].

.....

Even when conformity to a rule is 'distributively,' or for everyone 'considered singly' [i.e., for any person] possible [and consistently willable], it can be impossible [or not consistently willable] for everyone (i.e., for all persons) 'collectively,' or 'taken together.' [See Sobel 1967] Indeed there are at least two senses in which everyone's (i.e., all persons') doing what is required by a rule can be [if not] 'collectively' impossible [then at any rate not consistently willable] – two distinct and *independent* senses. (Sobel, 1985b)

### 3.1. 'Universal Actual' and 'Universal Projected', Conformities

Different 'logics of universal conformity' – different understandings of what for purposes of the compass it is would be for a maxim to become a universal law of nature – are used in the bridge and election cases. These 'logics' are distinguished in (Sobel, 1985a) and (Sobel, 1985b). Present labels for them were introduced in (Sobel, 1987b), and they are used under these labels in (Sobel, 1988).

The bridge case supposes that a maxim's becoming a universal law of nature would consist in a combination or set of actions in which each person does what, in an initial actual situation, that law would have him do, which may depend on what, *in this initial actual situation*, others are doing and would do were he to do this or that. To identify this combination or set of actions, we determine for each person what, in the initial actual situation, that law would have him do – this combination or set of actions consists of the thus determined actions, one for each person in the situation. In this way 'everyone's acting in accordance with a maxim *M*' comes to,

each person's doing what, were he to do it in the actual situation, would have him acting in accordance with *M*.

Such combinations or sets of actions are 'universal *actual* conformities' (Sobel, 1987, p. 279, and Sobel, 1988, p. 234). These sets are not in every case what might be termed 'actual universal conformities'. A 'universal actual conformity' to a rule can be a set of actions in which, far from each action's actually taking place, none do (Sobel, 1988, p. 235). This way of universalizing leads in the bridge case (depending on what the maxim being tested by the compass would have agents do) either to an impossible combination of actions in which each of several persons first and alone tries to cross, or to a possible but unwillable combination of actions in which everyone tries to cross. (This unwillable combination of actions is, as it happens, also 'thoroughly unactual': in the case, no one tries to cross.)

The election case, in contrast, supposes that a maxim's becoming a universal law of nature leads to an at least possible combination or set of actions such that were it to take place each person would be conforming to the law given what everyone else was doing in this perhaps only projected and possible combination or set of actions, and, in it, would do if he were to do this or that. To see whether a possible combination or set of actions has this character, one considers whether each person would be conforming to the maxim, if all actions in this combination were to take place, which for some law may for each person depend on what others, *in this perhaps only projected possible situation*, were doing and would do were he to do this or that. In this way 'everyone's acting in accordance with a maxim *M*' comes to,

persons acting in ways such that, were they to act in these ways each would be acting in accordance with *M*.

Such combinations or sets of actions are 'universal *projected* conformities' (Sobel, 1987, p. 279, and Sobel, 1988, pp. 234–5). Necessarily, universal projected conformities are *possible* combinations of actions. In this, universal projected conformities contrast with universal actual conformities which can be impossible combinations of actions (sets of actions not all of which can take place together). But, like possible universal actual conformities, universal projected conformities need not be willable. This way of universalizing leads in the election case (thanks to the way in which social contexts are included in maximed actions) to combinations of actions all of which are willable.

I have said that every universal projected conformity is by definition a possible combination of actions. Let me add that though every universal projected conformity is a possible combination of actions, there is a *sense* in which universal projected conformities are in some situations impossible. There may, in a situation, not *be* a universal projected conformity to a law. For example, there is, in any situation involving more than two nations, no universal projected conformity to the law that would have a nation disarm alone for it is impossible for *both* of two nations to disarm alone (it is impossible for two nations *together* to disarm alone). And there is, in a situation in which each of two persons can preserve his life by, and only by, consuming a kind of pill of which there is just one, no universal projected conformity to the law that would have a person preserve his life when he can.

3.2. *Elaboration on These Two Ways of Maxims Becoming Universal Laws*

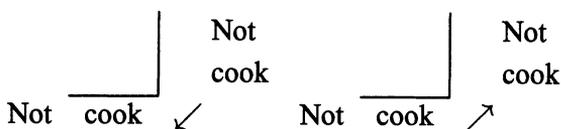
Control of this seldom noticed distinction between ways of becoming universal laws – this distinction between universal actual and universal projected conformities – is essential to elaborations and examinations of possible ‘categorical imperative procedures’. The ‘nice’ points of this section may help to clarify and fix this difficult to state and to grasp distinction.

3.2.1. *Universal actual conformities can be far from universal projected conformities.* It is possible for acts that a law would have persons do in a situation to be such that, were they all to take place, then no one would be doing what that law would then have him do. A universal actual conformity can be that far from being a universal projected conformity. For this point, consider a situation in which each of two persons, Row and Column, has promised to cook dinner if and only if the other does not cook dinner. For the ‘full independence’ of their actions let them be in ‘isolation kitchens’.

Suppose that neither will cook dinner. Think about a law to keep promises. Their actual interaction situation has the following shape:

		Column		
		Cook	Not cook	
Row	Cook	Neither keeps his promise.	Each keeps his promise.	Cook
	Not cook	Each keeps his promise.	Neither keeps his promise.	Not cook
		Cook	Not cook	

Vertical and horizontal arrows indicate what they will do, and angled arrows at corners what each would do were the other to do this or that. For example, the vertical arrow says that Column *will not* cook. The angled arrows on the southeast corner,



say respectively,

If it were the case that Row does not cook, then it would be the case that Column does not cook. $(\sim Cr \square \rightarrow \sim Cc)$	If it were the case that Column does not cook, then it would be the case that Row does not cook. $(\sim Cc \square \rightarrow \sim Cr)$
--	--

The pattern of arrows (vertical, horizontal, and angled) on the matrix makes explicit the independence of Row's and Column's actions in the actual situation.

The law to keep promises would in this situation have each cook dinner, for neither is going to cook each has promised to cook if and only if the other player does not cook. But were this universal actual conformity to the promise rule to *take place* – were both to cook dinner – then, in that *projected* situation in which each was cooking,

		*		
		Cook	Not cook	
* Cook	↗ ↘	Neither keeps his promise.	Each keeps his promise.	↖ ↙ Cook
Not cook		Each keeps his promise.	Neither keeps his promise.	Not cook
	↘ ↙	Cook	Not cook	

neither would be keeping his promise, just as neither is keeping his promise in the actual situation. Angled arrows at corners of this matrix indicate second order subjunctive conditionals. They indicate what, were the asterisked, universal-actual-conformity combination of actions, (Cook, Cook), to occur, each would do were the other to do this or that. The pattern of angled arrows and asterisks makes explicit that actions of Row and Column would still be independent in this projected situation. That they would be is part of what is meant by saying they are *fully* independent.

The universal actual conformity to the rule, (Cook, Cook), for the identification of which see the first embellished-by-arrows matrix, is what might be termed a 'universal projected violation', as the second embellished matrix makes plain. The universal projected conformities in this case are the mixed combinations of actions (Not cook, Cook) and (Cook, Not cook), as embellished matrices in which these combinations are asterisked could make completely evident.

3.2.2. *A Universal Projected Conformity Can be Far from a Universal Actual Conformity.* Actions each of which would actually violate some law can make a combination of actions in which, were it to obtain, everyone would be conforming to this law. A universal projected conformity can that far from being a universal actual conformity. Consider the situation already detailed with this change: though (as above) Column will not cook, Row will.

		↓				
		↙	Cook	Not cook	↘	
→	Cook	Neither keeps his promise.	Each keeps his promise.	Each keeps his promise.	Cook	
	Not cook	Each keeps his promise.	Neither keeps his promise.	Neither keeps his promise.	Not cook	
		Cook	Not cook	↙		

Then Row's *not* cooking would violate the promise-rule, as would Column's *cooking*: each, recall, has promised to cook if and only if the other does not. The combination of violations (Not cook, Cook), this 'universal actual violation' of the promise-rule, would be as far as can be from the universal actual conformity to the promise-rule which is (Cook, Not cook). Even so, however, were the combination (Not cook, Cook) of actual violations to obtain, each *would be* keeping his promise. For that all that is required is that exactly one of the two parties should cook, it does not matter who, while the other player does not cook. So this combination of violation, though as far as can be from a universal actual conformity to the promise rule, is a universal projected conformity.

The two kinds of universal conformities are in various cases as far as can be from one another. However, it has been observed that though only possible universal actual and projected conformities can differ in these ways, under a certain condition *actual* universal actual and projected conformities cannot differ: if "all ... subjunctives [in a case] are ... 'strongly centered' [that is the mentioned condition] ... a set of actual actions [in the case] is a universal actual conformity to a rule if and only if it is a universal projected conformity to it" (Sobel, 1988, p. 235).

3.2.3. To further rehearse differences and relations between these kinds of universal conformity, universal actual conformity to the rule to maximize

utility (whether private or public) is, as in the first promise-case, as far as it can be from a universal projected conformity in the situation,

(i)

			↓		
		C1	C2	↖	
	R1	0	1		R1
→	R2	1	0		R2
		↖	↘		
		C1	C2	↗	

supposing that this situation is 'fully independent', that is, supposing, that is, that the following projected subjunctive relations between actions would obtain, were the combinations  $(R1, C2)$ ,  $(R2, C1)$ , and  $(R2, C2)$ , respectively, to obtain:

(ii)

			*		
		↖	C1	C2	↘
	R1	0	1		R1*
	R2	1	0		R2
		C1	C2	↘	

(iii)

			*		
		↗	C1	C2	
	R1	0	1		R1
*	R2	1	0		R2
		↖	C1	C2	↗

(iv)

			*		
		↗	C1	C2	↘
*	R1	0	1		R1
	R2	1	0		R2
		↘	C1	C2	

The universal actual conformity to the rule to maximize is  $(R1, C1)$ . In the actual situation – see matrix (i) above –  $R1$  maximizes, for if Row were to  $R1$ , Column would still  $R2$ , for a value of 1, which is better than 0, the value of  $R2$ . Similarly,  $C1$  would maximize in the actual situation. So universal actual conformity to the rule to maximize utility comes in this case to the combination of actions  $(R1, C1)$ . But – see matrix (iv) above –

if this combination of actions were to obtain, then neither Row nor Column would be maximizing expected utility.

Also, as in the second promise-case, universal projected conformity to the rule to maximize utility is as far as can be from universal actual conformity in the situation,

(i') →

			↓		
	↙	C1	C2	↘	
R1		0	1		R1
R2		1	0		R2
		C1	C2	↗	

supposing that this situation is fully independent so that,

(ii')

			*		
		C1	C2	↖	
R1		0	1		R1
* R2		1	0		R2
	↖	C1	C2	↘ ↗	

(iii')

			*		
	↗	C1	C2		
R1		0	1		R1
* R2		1	0		R2
	↖ ↘	C1	C2	↗	

(iv')

			*		
	↗ ↘	C1	C2	↘	
* R1		0	1		R1
R2		1	0		R2
	↘	C1	C2		

The universal actual *violation* of the rule to maximize utility is (R2, C1): see matrix (i'). But (R2, C1) is a universal projected conformity: see matrix (iii').

*Query:* Can the points made above by pairs of cases, be made by a single case? Yes. Consider a situation,

$$\begin{array}{c} \downarrow \\ \begin{array}{ccc} C1 & C2 & C3 \\ \rightarrow R1 & \begin{array}{|c|c|c|} \hline 0 & 0 & 1 \\ \hline \end{array} \\ R2 & \begin{array}{|c|c|c|} \hline 0 & 3 & 2 \\ \hline \end{array} \\ R3 & \begin{array}{|c|c|c|} \hline 1 & 2 & 0 \\ \hline \end{array} \end{array} \end{array}$$

Assume full independence. That is, assume independence for this actual situation, and for the related eight only projected situations. The universal actual conformity to the rule to maximize utility is  $(R3, C3)$ , and this is a universal projected violation of this rule. The universal projected conformity to rule to maximize utility is  $(R2, C2)$ , the situation's sole equilibrium. And this is a universal actual violation of this rule.

### 3.3.

In sum, a distinction was made in Section 3.1 between two senses of 'everyone's acting in accordance with a maxim', or in other words two ways of maxims becoming universal laws, actual and projected is made. This distinction has been explored in Section 3.2 with the object of 'fixing' it. It is not an easy distinction, but it is absolutely essential to issues concerning the intended and possible ways in which Kant's compass might work. Its importance indeed stems from its difficulty, and the ease with which it is possible in friendly discussions of the compass to let it, for good results, work now in one way, and now in the other.

## 4. CONCLUDING OBSERVATIONS AND A SCEPTICAL CONJECTURE

### 4.1.

If Kant had been sensitive to problems regarding social contexts, and whether these are, for purposes of his compass, to be incorporated into, or excluded from, what maxims would have persons do, *and* had been more sensitive to problems regarding the logic of maxims becoming universal laws; he might have found that to his way of thinking, issues regarding lying and promise-breaking for benevolent reasons, far from being easily resolved by even the commonest of minds, were very complicated. Possibly worse and more troubling for him, he might have found that his compass pointed to conclusions at odds with his most confident rigorous

intuitions. For example, there is no obvious difficulty with – in the universal projected manner exemplified in the election case – willing to become universal laws maxims that would have people, in order to do good, *to lie* and *to break promises* not always, but *exactly when whether or not one lies or keeps a promise there will be where one lives enough truth-telling and promise-keeping for confidence sufficient for language as a means for communication, and for cooperation and coordination*. This observation would have been disturbing for Kant, who was unreservedly against lying and breaking promises, if he had found himself drawn to this logic of maxims becoming universal laws.

There are furthermore reasons for thinking that Kant, if confronted with an articulated choice between the two, *would* have been drawn to the universal projected logic for maxims becoming universal laws. This logic is consonant with his ‘moral idealism’ (discussed in Sobel, 1987). He makes definitive of the moral law that universal obedience to it would realize a realm of ends, a situation in which everyone obeys the moral law to certain good effects for liberty and happiness. Kant says that this glorious ideal awakens a lively interest in the moral law (462–3), so that, keeping one’s eye on this prize, one may scrupulously will as directed by the law (pointed out, presumably, by the compass) whether or not all others or even any others ever do so, and so qualify oneself as “a universally legislative member of [what may well be] a merely potential realm of ends” (438) in which everyone would always do so and be himself a universally legislative member.

#### 4.2.

That this logic – the logic of universal projected conformity – coupled with a license to include social contexts in ‘maximed actions’, would have his compass pointing to conclusions at odds with some of his rigorous intuitions, while disturbing for Kant, could encourage Kantians with more liberal intuitions to favour this way of universalizing maxims, and the inclusive policy regarding social-contexts suited to it. That could seem a way of retaining Kant’s logical principles while avoiding, as many would wish to do, a number of his substantive conclusions “such as that lying, suicide, and political revolution are **always** prohibited” (Schneewind, 1992, p. 324, bold emphasis added) though the heavens fall. (See, regarding lies, ‘On a Supposed Right to Tell Lies from Benevolent Motives’ and *Lectures on Ethics*, p. 13, regarding suicide, (422), regarding obedience, *Old Saw* 305, and ‘What is Enlightenment?’ 37–8, and regarding revolution, *Old Saw* 300 and ‘Perpetual Peace’ 373n. Also consider *Old Saw* 286 where (*pace* Plato, *Republic* 331c) Kant says that it takes no “penetrating acuteness”

and is obvious to any “child of eight or nine” that it is wrong not to return deposits entrusted to one – this is wrong, he implies, no matter what the circumstances and consequences.)

Tempering that encouragement, however, could be the thought that, as there are no obvious difficulties with, in the manner exemplified in the election case, willing for laudable ends *good* lies *when enough others are telling the truth*, so there are also no obvious difficulties with willing in this manner *bad* lies, and in willing to become universal, maxims that would have one *for personal gain take advantage* of the social responsibility of others who as a matter of principle tell the truth, vote, pay taxes, keep their promises, and so on. There is no *obvious* difficulty in willing to become universal laws, maxims that would have persons free-riding and making exceptions of themselves for personal gain *when and only when they can do this without undermining the very practices and institutions they would exploit*. Such maxims can pass Kant’s test administered in the universal-projected-conformity-with-social-contexts-included manner. It is true that there cannot be situations in which everyone takes advantage of the social responsibility of others and makes himself an exception, but this is not a problem for that manner of universalizing. Such maxims can pass the test administered in that manner, first, because it is possible for the right numbers for general good to show social responsibility, and for the right numbers to take advantage of those who are showing it, and, second, because when this happens – when the numbers are in this way right – everyone is, as such a maxim would have him do, taking advantage for personal gain of the social responsibility of others *when he can without undermining the practices of value to everyone he would exploit*, and is otherwise showing social responsibility and being taken advantage of by others.

#### 4.3.

Certainly it would not be easy to assemble an adequate user’s manual for Kant’s compass. Of particular difficulty, I have argued, would be settling on a single logic of maxims becoming universal laws and on a single treatment for social contexts, and resisting the temptation for palatable results to suit logics and treatments to cases, and this without aid of guiding principles but simply *for* palatable results. For this temptation, consider the tantalizing prospects of better results in both the bridge cases (as well as in the doctor case) and in the election case (as well as in the false promise case) if the logics and treatments used in these cases are *switched!* It is possible suiting logics and treatments to these cases to get the results one wants. For half of this switch, fix the maxim in the bridge case to make plain that it calls for

my trying to cross on the bridge if and only if I would try alone. Universal *projected* conformities of a rule to this effect would have exactly one, any one, of the party testing the bridge, which is both possible and willable consistent with the other-regarding end of my bridge-maxim (see Section 2.1). For the other half, consider that in the election case universal *actual* conformity to the rule, when one has better things to do and one's vote is not needed not to vote, would come to the unwanted and presumably unwillable combination of actions in which no one voted.

One wants, however, a *single* logic and a single treatment. Or, failing that, one wants *principles* that assign to cases logics and treatments of social contexts. One needs, if not a single logic and a single treatment, principles and not merely the license to use whichever logic and treatment in a case gets the result you want regarding a possible maxim, if the compass is to provide *guidance*. There are hazards for logics and treatments, different hazards for different logics and treatments, between which Kant's compass no matter how it is refined may not be able to find a way.

My opinion, for what it is worth, is that Kant's own ship founders on these divergent logics and treatments, and that its compass cannot be fixed to yield acceptable to him, or to anyone, assessments of maxims in all cases. Which is not to say that there is nothing useful in its general idea that 'subjective principles of action' should be constrained by the discipline of their universalizations, somehow understood, being be willable. That can be, that has been, a fruitful idea, especially when, setting aside problems of Kantian *exegesis* and allowing what he says about and claims to do with his compass merely to inspire, theorists make of it something of their own. Cf.:

I am not . . . attempting . . . exegesis. . . . Nor do I claim to know just what Kant meant; I have been inspired by him to say certain things, but when I look to see where *he* said them, I always get lost. (Hare, 1981, p. vi.)<sup>7</sup>

#### 4.4.

It is in this positive moving-on spirit that, in closing, I record not a bad idea, albeit one of limited application, started by Ingvar Johansson in discussion of my bridge cases. The idea, for a broadly Kantian test that would approve of the apparently good maxim of these cases, *says goodbye to the business of willing maxims to become universal laws*, and instead makes the test of a maxim for an agent his willingness *that it should be everyone's maxim*. This is its major departure from Kant. To settle the issue whether he can will that his maxim should be everyone's maxim – now comes another departure from Kant – the idea instructs the agent to think about *the probable consequences of his maxim's being everyone's*

*maxim*. (Observe that that a maxim relevant to a situation is everyone's maxim does not entail that everyone acts on it, or even that everyone acts in accordance with it, in this situation.) The agent is to think about how its being everyone's maxim would probably work out in the real world in which, for example, when people find themselves on collision courses they are not stymied, but by zigging and zagging manage to coordinate their actions and to slip by one another to proceed on their ways. Relating this idea to the bridge case, it seems that members of the party who all shared my maxim could be expected in time, (1), all to realize the initial reluctance of others to venture onto the bridge, and to begin to move forward pursuant to this maxim of all which has an agent move forward when he thinks no one else will, (2), all to notice their several early movements, and to pause since this maxim has an agent venturing forward only when he thinks no one else will, (3), all to notice that they are again reluctant to venture forward, and to start up again pursuant to the maxim, and so on for some sequence of starts and stops. But not forever – nothing in the real world is forever. It is to be expected that someone will happen to separate himself from the pack and take the lead – perhaps after many starts and stops someone would trip and alone lurch forward leaving the rest behind – at which point the party's real-world coordination problem would almost certainly be solved.<sup>8</sup> The fortuitously selected hero would, pursuant to the maxim, continue on his forward way confident that everyone else pursuant to this maxim was holding back, the bridge would be tested as it needs to be, and, God willing, the troupe would be saved. In any case, and even if as it happens the bridge would fall under the weight of just one and all are lost, the maxim to do the brave thing in such a case is established as one that an agent with a little experience of interactions in the real world can will to be everyone's maxim.<sup>9</sup>

#### NOTES

<sup>1</sup> Citations, unless otherwise indicated, are by the pagination of relevant volumes of the Prussian Academy edition of Kant's works, and from various translations for which see References below. When unadorned they are to the *Foundations*.

<sup>2</sup> It is curious that Kant runs together "what is **good**, what is bad" and "what is **consistent**, what is **inconsistent** with duty" (404, bold emphasis added). For one expects not only consistency with but *regard for* duty to be necessary in Kant's view for the goodness of maxims, as he says it is for goodness of wills. The Compass just might afford a necessary and sufficient test for consistency with duty, and a necessary test for regard for duty. But it is not easy to see in it a sufficient test for that regard.

<sup>3</sup> Natural words for plans and intentions feature infinitive phrases. When packing for a trip I might, to indicate my particular plan, say it is *to go to New York* where no one knows me for a fresh start. Or to indicate a general plan I was implementing, I might say that it is, in

circumstances such as these when I have made a promise to be somewhere, *to keep it*.

<sup>4</sup> To avoid an ambiguity in the scope of 'try' I have not written 'to try to cross on the bridge first and alone before anyone else'. The intended scope of 'try', which could not in that formulation be easily signaled, is 'to cross on the bridge' and not 'to cross on bridge first and alone before anyone else'. My action is to be a case of trying. Its description is to include (i) that I am trying to cross on the bridge, and (ii) that I am doing precisely this – trying to cross on the bridge – first and alone before anyone else does it.

<sup>5</sup> This community is out of luck. It will not have a doctor.

Suppose there are just two people, Smith and Brown. Let '*Ds*' abbreviate 'Smith will become a doctor', and similarly for '*Db*'. It is stipulated that: (i) ( $Ds \square \rightarrow \sim Db$ ), (ii) ( $\sim Ds \square \rightarrow \sim Db$ ), (iii), ( $Db \square \rightarrow \sim Ds$ ), and (iv) ( $\sim Db \square \rightarrow \sim Ds$ ). Assume for an indirect proof to the contrary that, (v) *Ds*: derive for a contradiction  $\sim Db$  (from (i) by *MP*), and *Db* (from (iv) by *DN*, *MT*, *DN*). Therefore,  $\sim Ds$ . It follows similarly from (iii) and (ii) that  $\sim Db$ .

<sup>6</sup> There is discussion of social contexts in descriptions of acts for utilitarian generalizations in Sobel (1970).

<sup>7</sup> Hare's relation to Kant's compass contrasts with Korsgaard's. She considers three families of interpretations of 'The Formula of Universal Law' (Korsgaard, 1985, p. 26), and allows that support for each "can be found in Kant's texts" and that it is possible that "he was not aware of the differences among them" (p. 27). Two of these she criticizes, and one she defends "primarily on philosophical considerations" (p. 27, bold emphasis added). She defends this one, but in exactly what spirit? As the truth regarding maxims and duty, as the best and closest to the truth that can be made of Kant's relevant words, as the most edifying that we can make of these words, or what?

<sup>8</sup> "Consider Mackie's (1973) discussion of coordination, and . . . a symmetrical situation with two best equilibria:

2, 2	1, 1
1, 1	2, 2

Mackie asks, 'How do we solve this apparently insoluble problem?' In practice, he observes, 'one person happens to move . . . before the other [who] then adapts his own movements to fit in with that of the first' (p. 293)." (Sobel, 1994, p. 343n8.)

<sup>9</sup> Thanks to Willa Freeman-Sobel for many helpful suggestions and criticism.

#### REFERENCES

- Hare, P. M.: 1952, *Moral Thinking: Its Levels, Method, and Point*, Oxford University Press.  
 Herman, B.: 1993, *The Practice of Moral Judgment*, Cambridge, Mass.  
 Hume, D.: 1902, in L. A. Selby-Bigge (ed.), *Enquiries Concerning the Human Understanding and Concerning its Principles of Morals*, Oxford University Press.  
 Kant, I.: 1979, *Lectures on Ethics*, translated by L. Infeld, Methuen, London (reprint by Hackett Publishing Company). (Student notes of lectures delivered during 1775–1780.)  
 Kant, I.: 1958, *Critique of Pure Reason*, translated by N. Kemp Smith, London. (First edition 1781; second edition 1787.)  
 Kant, I.: 1983, 'An Answer to the Question: What is Enlightenment?', *Perpetual Peace and Other Essays*, translated by Ted Humphrey, Indianapolis. (First published in 1784.)

- Kant, I.: 1969, *Foundations of the Metaphysics of Morals*, translated by L. W. Beck, Indianapolis. (First published in 1785.)
- Kant, I.: 1956, *Critique of Practical Reason*, translated by L. W. Beck, New York. (First published in 1787.)
- Kant, I.: 1974, *On the Old Saw: That May be Right in Theory But It Won't Work in Practice*, translated by E. B. Ashton, Philadelphia. (First published in 1793.)
- Kant, I.: 1983, 'To Perpetual Peace: A Philosophical Sketch', in *Perpetual Peace and other Essays on Politics, History, and Morals*, translated by Ted Humphrey, Indianapolis. (First published in 1795.)
- Kant, I.: 1797, 'On a Supposed Right to tell Lies from Benevolent Motives', (Translated by T. K. Abbott, and included in his *Kant's Critique of Practical Reason and other Works on the Theory of Ethics*, Sixth Edition, London 1923. Abbott writes in a note: "This Essay was published in a Berlin periodical in 1797.")
- Kant, I.: 1991, *The Metaphysics of Morals*, translated by Mary Gregor, Cambridge. (First published in 1797.)
- Korsgaard, C. M.: 1985, 'Kant's Formula of Universal Law', *Pacific Philosophical Quarterly* 66, 24–47.
- Mackie, J. L.: 1973, 'The Disutility of Act-Utilitarianism', *Philosophical Quarterly*.
- O'Neil, Onora: 1989, *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge.
- Prichard, H. A.: 1970, 'Does Moral Philosophy Rest on a Mistake?', *Mind*, 1912, reprinted in W. Sellars and J. Hospers (eds.), *Readings in Ethical Theory: Second Edition*, New York.
- Schneewind, J. B.: 1992, 'Autonomy, Obligation, and Virtue: An Overview of Kant's Moral Philosophy', in P. Guyer (ed.), *The Cambridge Companion to Kant*, Cambridge.
- Singer, Marcus: 1961, *Generalization in Ethics*, New York.
- Sobel, J. H.: 1967, '"Everyone", Consequences, and Generalization Arguments', *Inquiry*, 373–404.
- Sobel, J. H.: 1970, 'Utilitarianisms: Simple and General', *Inquiry*, 394–404.
- Sobel, J. H.: 1985, 'Utilitarianism and Cooperation', *Dialogue*, 137–52.
- Sobel, J. H.: 1985, 'Everyone's Conforming to a Rule', *Philosophical Studies*, 375–87.
- Sobel, J. H.: 1987a, 'Cephalus: Republic 331c–d', *History of Philosophy Quarterly*, 282–90.
- Sobel, J. H.: 1987b, 'Kant's Moral Idealism', *Philosophical Studies*, 278–87.
- Sobel, J. H.: 1988, 'Maximizing Optimizing, and Prospering', *Dialogue*, 233–62.
- Sobel, J. H.: 1994, 'Utility Maximizers in Iterated Prisoners' Dilemmas', Chapter 15 in *Taking Chances: Essays on Rational Choice*, Cambridge University Press (reprinted with revisions from *Dialogue* 1976).

Manuscript submitted June 27, 1996

Final version received October 25, 1996

Department of Philosophy  
University of Toronto,  
215 Huron Street  
Toronto, Ontario  
Canada M5S 1A2