False Negatives, False Positives, and Kant’s First Formulation

Abstract: Kant’s first formulation of the categorical imperative is criticized on the grounds that it makes action on morally innocuous maxims immoral (the problem of false negatives) and action on grossly immoral maxims moral (the problem of false positives). We argue that neither of these objections succeeds. Reflection on why these objections do not succeed reveals important distinctions regarding the categorical imperative and its formulations and the statement of maxims.

Keywords: Universalizability, Universal Law, Categorical Imperative, Maxim, Kant

Kant’s first formulation of the categorical imperative—the Principle of Universalizability or the Formula of the Universal Law of Nature—is widely criticized for two reasons. The first is that the principle makes acting on morally innocuous maxims immoral. Allen W. Wood (1999: 105) provides such an example when he considers a man who intends to buy but never to sell a toy train. This maxim, when considered as a universal law of nature, seems to make the action immoral, for if everyone bought but never sold his or her toy trains, then there would be no toy trains to buy (supposing, of course, that no more are being manufactured). Yet acting on such a maxim seems to be perfectly moral. Following Wood, we shall call this the problem of false negatives.

The second objection is that the principle makes acting on grossly immoral maxims moral. A common example is fanatics. This particular example is from Russ Shafer-Landau (2010: 154–55): there is no obvious reason that acting on the maxim that whenever a
person steps on my lawn without permission from me that person shall be killed could not be a universal law of nature. If anything, it only provides people with a good reason to walk on the sidewalk. Again following Wood, we shall call this the problem of false positives.

Our position is that these problems—at least as they have been stated and developed in the literature—do not succeed in refuting Kant’s first formulation. Reflection on why these objections fail reveals important distinctions involving the categorical imperative and its formulations, the statement of maxims, and the very idea of a universal law of nature. Moreover, it reveals interesting connections between Kant’s first and second formulations of the categorical imperative.

Our paper divides into two major sections. In the first section, we discuss the first formulation of the categorical imperative and what is involved in its application. Second, drawing on the insights of the first section, we show that the problem of false negatives and the problem of false positives as they are found in the literature are not problems after all. Of course, we do not deny that examples other than those considered here might be developed, and so we do not take ourselves to have settled once and for all a response to these problems.

The First Formulation and Universal Laws of Nature

The Categorical Imperative and Its “Formulations”

Kant is quite clear that there is only “a single categorical imperative” and that it is to “act only in accordance with that maxim through which you can at the same time will that it become a universal law” (GMM 4:421; see also CPrR 5:69). However, he thinks this statement is extremely general and that we need “to show what we think by it and what the
concept wants to say” (GMM 4:421; see also CPR 5:69, for we need to present the law “in concreto in objects of the senses and hence a law of nature”). In other words, being able to merely state the categorical imperative is not enough to render our comprehension of it clear. We also need to know what follows from it, its practical consequences or applicability to us.

A helpful analogy is with what Kant states in the first Critique regarding truth: it is one thing to know that “the nominal definition of truth...is the agreement of cognition with its object” but one also “demands to know what is the general and certain criterion of the truth of any cognition” (CPR A58/B82). The nominal definition of truth does not tell us anything about how truth links up with our practices, with our methods of inquiry, or with our beliefs. To make the conception clearer, we need to establish a criterion of truth. In like manner, it is one thing to know what the categorical imperative is and quite another to know how it links up to us and to our practices.

Kant aims to render our conception of the categorical imperative clearer by providing three different “formulations” of the categorical imperative. We put “formulations” in scare-quotes here because calling them formulations of the categorical imperative runs the risk of confusing these for three different statements of the categorical imperative.\(^1\) They are not. Kant states that the basis of the three formulations is the

\(^1\) We are thus in agreement with Wood (2008: 80–2), who argues that the formulations of the categorical imperative are not equivalent. However, we dissent from his interpretation insofar as he conflates the categorical imperative itself (the universal law formulation, on his view) with the first formulation (the law of nature formulation, which he regards as the “intuitional” expression of the universal law formulation) (see 66 and 82). Moreover, we
categorical imperative as stated above but that if "one want also to provide access to the moral law, it is very useful to bring one and the same action under the three concepts [i.e. the formulations] mentioned above and thereby, as far as possible, bring it closer to intuition" (GMM 4:437). In other words, the three formulations are different ways of rendering the one categorical imperative accessible to or applicable to us. They are different ways of understanding what predicating ‘moral’ of a proposed course of action entails. With respect to the first formulation, if an action is moral, then it could be a universal law of nature. With respect to the second, if an action is moral, then no human is used as a mere means and no human is not treated as an end in himself or herself. With respect to the third, if an action is moral, then the maxim must be such that one could legislate it and be subject to it in the universal kingdom.

Notice that the three immediately previous sentences are not biconditionals. We cannot conclude from the fact that a maxim could be a universal law of nature that action do not discern five, or even six, different formulations of the categorical imperative.

Following Wood’s abbreviations, for which the interested reader is invited to consult his book, we think FUL is the categorical imperative and FU is a less felicitously expressed version of it. The first formulation is FLN, the second is FH, and the third is FRE, which is less felicitously expressed in FA. We should not be bothered by the fact that Kant sometimes expresses himself infelicitously. We all do, though we surely lack the genius of Kant. Second, we know that Kant composed his works hastily for want of time (which is not to say his ideas are hastily considered!). Lastly, Kant did not have the benefits of computers and modern printing presses; hence, infelicitous expressions sometimes had to remain as they were typeset simply for practical reasons.
on such a maxim is moral, for that would be to affirm the consequent. In like manner, we cannot conclude from the fact that in performing an action I treat no rational nature as a mere means and treat no rational nature as not being an end-in-itself that my action is therefore moral. In other words, the formulations of the categorical imperative can only tell us when an action is immoral and not when an action is moral. Indeed, in all of Kant’s writings on ethics, he never—not once!—uses the formulations of the categorical imperative to conclude that an action is moral. He only ever uses them to conclude that an action is immoral.2

To return to our previous point, the “formulations” of the categorical imperative tell us what the single categorical imperative entails. They do not restate what the categorical imperative is. To illustrate this point, it is helpful to draw an analogy. On analogy with the predicate ‘hot’, it one thing to know that to be hot is to have a highly intensive magnitude (see CPR A165/B208) and quite another to know that heat necessarily follows from the shining of the sun. The former is a pure transcendental claim about heat whereas the latter is a law of nature that incorporates content from an experience but is not learned from experience (PFM 4:312). As Kant explains, the latter is a judgment of experience but is not learned from experience because the concept ‘necessity’ is provided by the understanding and not found in the experience. Thus, this differs from the mere judgment that heat

2 This, of course, raises another and very significant problem: how do we know when an action is moral? Plausibly, although none of the three formulations is individually sufficient to inform us of when an action is moral, if we can infer from none of them that an action is immoral, then we can conclude that the action is moral. More shall be said regarding this in what follows.
follows the shining of the sun, which is a judgment of perception and involves no claims pertaining to necessity.

Now to draw the analogy to morality: it is one thing to know from pure practical reason that a moral action is such that one can will that his or her maxim at the same time become a universal law and quite another to know that a moral action must be such that it can necessitate actions *throughout the order of nature*. The former is a transcendental claim about morality. The latter incorporates empirical content from our experiences about people and their actions. Determining the morality of an action using the latter requires incorporating empirical content even though the morality of the action is not known *from* those experiences. Just as with empirical laws the concept of necessity is not found in experience, so with morality the concept of duty is not found in experience, a point Kant emphasizes when he claims that it is impossible to make out from experience a single case of action from duty alone rather than self-love (GMM 4:407). As Kant states, the first formulation is “not the determining ground of [one’s will]” but is “a *type* for the appraisal of maxims in accordance with moral principles” (CPrR 5:69). The determining ground is the categorical imperative. The type is the first (or second or third) formulation.

---

3 Kant is very clear about this in discussing the first formulation at CPrR 5:67–71.

Consistent with Kant’s endorsement of thoroughgoing determinism regarding the sensible world, he stresses that practical reason’s consideration of a possible action does not imply that the action is possible (CPrR 5:68). Nonetheless, in considering a maxim as a law of nature, “it is also permitted to use the *nature of the sensible world* as the *type* of an *intelligible nature*, provided that I do not carry over into the latter intuitions and what depends upon them but refer to it only the *form of lawfulness* in general” (CPrR 5:70).
In sum, the first formulation is not an alternative statement of the categorical imperative but is a type of the categorical imperative. As a type, its purpose is to render clear the application of the categorical imperative by considering its practical consequences for us. It tells us what the categorical imperative entails; it does not restate the categorical imperative. In applying the first formulation, we incorporate empirical content into our judgments of morality. However, a judgment of the morality of a possible action is not learned from experience but is made by practical reason.

**Interpretations of the First Formulation**

Kant’s first formulation of the categorical imperative is that the “universal imperative of duty” can be stated as “act as if the maxim of your action were to become by your will a universal law of nature” (GMM 4:421). We shall first make some comment on the nature of maxims and second on the various interpretations of the first formulation.

*Maxims*: Maxims consist of two parts. The first part is what a person wills to do. A person may will to kill himself. She may will to falsely promise. A person may will to live a leisurely life.

The second part is why the person wills to do the action that he or she does. One may will to kill oneself out of self-love, out of self-hatred, or out of love for country. A person may falsely promise out of a need for money, out a fear of punishment, or out of a desire for praise.

In stating why a person wills to do some action, we are not stating why that person chose to go about the action as he or she did but why that person chose to go about the
action *that* he or she did. If a bank robber is asked why he robbed a bank, he might say he robbed the bank because that is where the money is. Yet such a response tells us why the bank robber went about the action—thieft—*as* he did, not why he went about the action *that* he did. In contrast, if a bank robber says he robbed a bank out of a need for money, this tells us why he did the action—thieft—that he did. After all, he could have robbed a convenience store, restaurant, department store, etc. and yet done it for the same reason. The action that he did is thievery. The action as he did it is thievery from a bank. The reason why he did it is a need for money.

The importance of making distinctions in the statement of our maxims is especially evident with respect to Kant’s second example of borrowing money on a false promise. In that example, the maxim is: I shall falsely promise to repay a loan out of a need for money. It is important to note that the action in question here is the person’s false promising, not his or her borrowing of money. It is the money that is the person’s reason for falsely promising: he or she desires or needs money.

Crucially, Kant does not tell us why the person in question desires or needs money, for that is not part of the person’s maxim for falsely promising. Rather, it is part of the person’s maxim for borrowing money. These two maxims must be carefully discriminated: (a) I shall falsely promise out of a need for money and (b) I shall borrow money out of (for example) a need to pay my bills. The former, Kant thinks, is immoral. The latter likely is not. We must consider the maxim of each action in turn.4

4 These distinctions also provide a way to respond to those, like Wood (1999: 104–05) who worry that by jiggering with the maxim’s generality or specificity it is possible to make grossly immoral actions moral by specifying, for example, the time of the event. Such
Interpretations of the First Formulation: Christine Korsgaard (1996:77–105 and much more briefly at 1997: xix–xx) has quite rightly drawn our attention to the fact that the first formulation of the categorical imperative can and has been given three different interpretations pertaining to its application. Both for want of space and because our response to the problems of false negatives and false positives does not rely on the defense of any particular interpretation, our presentation of these interpretations shall not delve into the many important distinctions involved in applying these interpretations.

On one interpretation, the logical contradiction interpretation, acting on a maxim is immoral just in case it implies some sort of logical contradiction. This contradiction may be of three sorts. On one sort of contradiction, there is a contradiction in the very maxim itself. Kant, for example, seems to think there is a contradiction between the very action of suicide and doing it out of self-love. Apparently, that is because the principle of self-love aims at the furtherance of one’s life whereas to suicide is to do the very opposite (GMM 4:422).  

maxims smuggle in the action as it is done, which is not properly part of the maxim. See below for more comments with respect to the statement of our maxims.

5 Kant, like many of his interpreters (for some of the debate on this, see Korsgaard 1996: 100, Brassington 2006, Vong 2008, and Wood 2008: 172–73), must have had doubts about this conclusion, for in CPRR the would-be suicide does not suicide from self-love but “consider[s] himself authorized to shorten his life as soon as he [is] thoroughly weary of it” (5:59, emphasis added) and in MM Kant raises a series of “Casuistical Questions” wherein he considers scenarios where one seems required to take or risk one’s own life (6:423–24).
The second sort of contradiction that might arise is between the action and the social practice of which that action is a part. For example, false promising so as to borrow money involves the social practice of lending money. The social practice of lending money implies the expectation that the money should be repaid; after all, that is precisely what a loan is. However, in false promising, the recipient of the loan has no intention to repay the money. Hence, there is a contradiction between what is implied by the social practice and what the agent intends to do when the maxim is conceived of as a universal law of nature.

As casuistry is not a science but is to be “woven into ethics in a fragmentary way,” (6:411) Kant seems to be suggesting that taking one’s own life is sometimes morally permissible.

Our judgment is that Kant’s claim that self-love is the principle that furthers life is false. We think Kant was deceived by this argument: (1) Self-love is the principle that aims at accruing some benefit for the agent or avoiding some harm to the agent (see CPrR 5:22); (2) if one is accruing some benefit to oneself or avoiding some harm to oneself, then one is aiming to further one’s life (this seems to be implied at GMM 4:422); (3) So, self-love is the principle that aims at the furtherance of one’s life. But (2) is false, for taking one’s life, which does not further one’s life, is sometimes done to avoid some harm to oneself (future pain) or even to accrue some benefit to oneself (glory or the pleasurable thought that accompanies the promise of glory). Consequently, there is no contradiction in conception in Kant’s application of the first formulation to the suicide example in GMM. However, we also think the first formulation will prohibit suicide in various other circumstances, including the would-be suicide from CPrR who intends to take his life as soon as he is weary of it.
The third sort of contradiction that might arise is a contradiction between what all rational agents would will if they gave it due consideration and the proposed action. For example, Kant thinks that all rational agents would will to develop their talents as those talents serve them. However, a person who chooses to live a life of leisure rather than to develop his or her talents opts for an action inconsistent with what all rational agents would will if they gave it due consideration.

The second interpretation of the first formulation is the teleological contradiction interpretation. On this interpretation, what we are to do is to consider the natural purposes of what is involved in our actions. For example, the natural purpose of promising is to foster trust and cooperation; however, false promising frustrates this natural purpose. In like manner, the natural purpose of our talents is to serve us; however, not developing those talents is to undermine their very purpose.6

The third interpretation, the one favored by Korsgaard, is the practical consequences interpretation. On this interpretation, we are to treat the maxim as a hypothetical imperative. Next, we are to ask whether in a world where everyone acted on this hypothetical imperative, the agent’s aim could still be attained. If not, then the action is not morally permissible.

6 Scott Stapleford (2007: 308–312) has argued that the teleological contradiction interpretation is just a special instance of the logical contradiction interpretation. We find his argument compelling, but following Herman (1993) and Korsgaard (1996), whose works have become the touchstone for a discussion of these interpretations, we distinguish them here.
For example, from the maxim, “I shall falsely promise out of a need for money,” we state the hypothetical imperative that “if one is in need of money, then he or she ought to falsely promise so as to obtain it.” Now, if we imagine a universe in which everyone follows that hypothetical imperative, we find that the agent could not attain his aim. That is because all of his promises would be regarded as vain pretenses.

What is important to note about all of these interpretations (and consistent with our comments above) is that they all incorporate some empirical content in determining whether an action is immoral even though the determination of the immorality of the action is not drawn from experience. To begin, consider the practical consequences interpretation. Here, we imagine a world very much like our own, where people do not like to be duped and where they distrust those who have duped them in the past. However, we can just as easily imagine a world of fools who continue to lend money even on false promises or who take great pleasure in lending money on false promises. In those cases, the attainment of the agent’s aims would be possible.

Second, consider the teleological contradiction interpretation. This incorporates empirical content insofar as we must ascertain the natural purposes of our actions. Why should the natural purpose of promising be to build trust and cooperation? It need not be, for we may very well imagine a world of false promisers, people who only promise falsely and so use promising as a way to foster distrust and defection. Instead, what the teleological contradiction interpretation must do is appeal to how we in fact use promises.

---

7 Apparently uncomfortable with incorporating empirical content, Stapleton writes of the suicide example, “It is inconceivable that there could be a natural system in which the
Finally, consider the logical contradiction interpretation. Clearest is that the second and third kinds of contradiction incorporate empirical considerations. The third must appeal to what would be the result of rational deliberation among people situated like us. For we could imagine a rational creature that has all of its needs and wants readily supplied by nature. Why such a creature would will to cultivate its talents—except perhaps those involving the pleasures consequent on purely theoretical studies—is not clear, for that creature's talents will not aid it. The second must appeal to the social practices we in fact have and what is involved in them for the reasons stated above with respect to the teleological contradiction interpretation.

Less clear is the first sort of contradiction, contradiction in the maxim itself. However, even here, there is some supposition of what is implied in the principle of self-love. For what is it that people who act on the principle of self-love strive to do? They strive to accrue benefits and to avoid harms to themselves. And what is a result of doing this? On Kant's view, the result is the furtherance of one's life. Hence, he thinks that self-love is the principle that aims at the furtherance of one's life. Yet this consequence is only determined empirically. For we could just as easily imagine a universe where those who act out of self-love end up shortening their lives; for example, the universe where God mercilessly smites anyone who acts out of self-love is just such a universe.


Such a natural system may be unlikely, but it is hardly inconceivable.
For some critics of Kant, this raises a serious problem. They suspect that empirical considerations are really doing the work in Kant’s moral theory. Kant’s moral theory is supposed to be a *priori*, but we are incorporating empirical considerations to determine whether an action is moral or immoral.

However, this problem really arises from a failure to appreciate that the first formulation of the categorical imperative is not a statement of the categorical imperative. As stressed above, it is a type of the categorical imperative. Its purpose is to help us get clear about the practical consequences of the categorical imperative, of what is entailed by the single categorical imperative. That an action is immoral can be inferred from the fact that it could not be a universal law of nature. But in determining what cannot be a universal law of nature, we must draw on empirical considerations of nature just as we must do for any law of nature.

It does not follow from this that the moral law is itself an *a posteriori* law. To the contrary, practical reason provides the concept of duty in the judgment that falsely promising so as to obtain money is a violation of duty just as the understanding provides the concept of necessity in the judgment that heat necessarily follows the shining of the sun. To draw on empirical considerations in making universal and necessary judgments

---

8 The criticism presented in this paragraph is a criticism that has been raised in conversation rather than one clearly defended in the literature. A related objection—that Kant’s first formulation is an empty formalism—was first levied by Hegel (for a discussion, see Lo 1981). Obviously, if the first formulation is an empty formalism, then something else must be doing the legwork in Kant’s examples. Plausibly, that is some purportedly smuggled-in empirical content.
does not entail that the universality and necessity is found in the experience. To state it another way consistent with Kant’s comments in CPrR cited above, even if the type of the categorical imperative draws on empirical considerations it does not follow that the categorical imperative itself is a merely empirical, contingent judgment.

What Kant is asking us to do in applying the first formulation of the categorical imperative is to consider our maxims as universal laws of nature. As universal, the law is to apply to everyone. As a law, it is objectively necessary. As natural, it applies to this world and not to any merely conceivable world, just as the empirical law that heat necessarily follows the shining of the sun is a law of this world and not any conceivable world whatsoever (for as we all know from Hume we can conceive of worlds where heat does not follow the shining of the sun).

**Laws and Universalizability**

Now consider, for example, the maxim “I shall falsely promise out of a need for money.” If we state this as a law of nature in conformity with the example of heat and the sun, we should state it as: falsely promising necessarily follows the need for money. One might wonder whether we should understand the necessity here as *de dicto* or *de re*. Is the law of nature:

*Necessarily, if a person needs money, then he falsely promises*

or:

*If a person needs money, then necessarily he falsely promises?*

Consistent with our claims above, since the law is supposed to be a law of nature that draws on empirical considerations and so need not apply across all possible universes, it
should be given the latter interpretation. This, of course, is also consistent with Kant’s statement that heat necessarily follows the sun’s shining rather than that necessarily heat follows the sun’s shining.\(^9\)

A natural supposition is that there must be some singular reason why a maxim cannot be a universal law of nature. However, there is no good reason to make this supposition and the supposition is at odds with Kant’s own comments. Korsgaard herself states that “Language supporting all of these interpretations can be found in Kant’s texts, and different interpretations fit different examples better” (1997: xx).

For example, with respect to falsely promising, Kant states, “it would be make the promise and the end one might have in it itself impossible” (GMM 4:423). Here Kant indicates that the maxim of false promising cannot be a universal law of nature for two reasons. First, false promising made a universal law of nature would make the very promise impossible. Second, it would make it impossible to attain the end one has in making a false promise.

Why it would make the end to be attained impossible is sufficiently clear from the practical consequences interpretation. If the law of nature that falsely promising necessarily follows the need for money obtains, then everyone must falsely promise whenever he or she needs money just as heat must follow the shining of the sun. However, we all know (based on experience) that no one lends money to false promisors regularly. As the old saying goes, “Fool me once, shame on you; fool me twice, shame on me.” So, if

\(^9\) There are, obviously, important issues here concerning modal and deontological logic. Unfortunately, those issues are beyond the purview of the present paper.
false promising became a law of nature, false promisors would not be able to attain their ends.

We want to stress how the previous paragraph stated the notion of the universalization of a law of nature as opposed to the universalization of a law more generally. When we formulate a universal law generally, we are not to wonder what would happen if everyone acted on a certain maxim. We are not to ask what would happen if everyone falsely promised when in need of money. Rather, we are to suppose that whenever anyone is in need of money, necessarily he or she falsely promises. Now, it may happen that no one ever needs money, in which case no one ever falsely promises. In such a universe, precisely nothing would happen if it were a universal law that falsely promising necessarily follows the need for money just as in a universe without stars precisely nothing would happen if heat necessarily follows the shining of the sun. Hence, there is no reason to conclude that the proposed law could not be a universal law.

However, when we consider whether this can be a universal law of nature, we incorporate empirical content into our judgments (though once again the morality of the action is not found in the empirical content but is provided by practical reason in the judgment that action on some maxim is immoral). We are to postulate that the thoroughly determined sensible world in which we live and move and have our being is regulated by the law in question. We now suppose that in the world we know falsely promising necessarily follows the need for money. In that case, everyone who is in need of money falsely promises. As there are very many people in our world who are in need of money, we find that the maxim of falsely promising if made a universal law of nature would undermine the very end we have in falsely promising. However, as already stated, it need
not undermine that end in every possible universe. This drives home the point that the first formulation is a type of the categorical imperative, a way to “show what we think by it and what the concept wants to say” (GMM 4:421), a way of gaining “access” to it (GMM 4:437).

There is another point to consider here. Thus far, we have been presenting the universal law of nature to state what one must do when certain conditions are met, for example when one is in need of money. However, Kant often states the law as what could be done or what is permissible (see GMM 4:422 and CPPr 5:69) rather than as what must be done.

Our position is that Kant’s manner of expression is imperfect but there is no serious problem. To begin, Kant does sometimes state the law of nature as what is done (in the first example of suicide) or as what is necessary (in the third example of rusting talents) rather than as what could be done (in the second example of false promising). Now, on the one hand, a moral law is a command that necessitates an action out of respect for the law (GMM 4:416 and 4:400) and expressed as a law of nature a maxim is the necessity of an action given the laws and state of the universe. Hence, it states what must be done or what is necessarily done. However, if a maxim can be a universal law of nature, then action on that maxim may yet be permissible (“may yet” because, recall from the comments above, the formulations can only used to conclude an action is immoral, not that an action is moral) but need not be required, for it might nevertheless be imprudent to act on the maxim in question even if it is moral (see GMM 4:418). More importantly, when a maxim cannot be a universal law of nature, it is false that the action is permissible. When Kant expresses the law of nature as what could be done he is merely slipping from the proper
mode of expression when considering a maxim as a universal law (for example, can "falsely promising necessarily following being in need of money" be a universal law of nature?) and what we know after ascertaining whether a maxim can be a universal law of nature (for example, that one is not morally permitted—one could not—make false promises whenever he or she is in need).

Let us return to the topic of whether there is some singular way to rule out whether a maxim can be a universal law of nature (i.e. whether one of the interpretations Korsgaard identifies is the correct interpretation). We have already seen why Kant would say that the end to be attained would be impossible if the maxim were a universal law of nature. Now, why would Kant state that the universalization of the false promising maxim throughout nature would make the promise itself impossible? That is less clear, but there are at least two possibilities, both stemming from the logical contradiction interpretation. The first is that a universal law of false promising would undermine the social practice of lending money. If everyone falsely promised on his or her loans, no one would ever lend money. A second possibility is that there is a contradiction in the very notion of a false promise, just like Kant thinks there is a contradiction in the act of suicide out of self-love. That is because the very concept of a promise implies the intention to fulfill it (though one might of course fail). However, the agent in question has no intention to fulfill his or her promise.10

10 One might then wonder whether it is possible to make a false promise at all. We think the answer is “yes” but that a false promise is something like a spoiled promise. Just as there can be unjust laws on the books but a natural law theorist thinks unjust laws are not, properly speaking, laws at all, so also there can be false promises in the practice of promising even though false promises are not, properly speaking, promises at all.
The upshot is that a universal law of nature has the form of *de re* necessity and with respect to applying the first formulation we consider our maxims to legislate *de re* universal laws incorporating empirical considerations. A maxim might fail to be able to be a universal law of nature for many different reasons, not just one. There is no reason we must choose one among the three interpretations discussed in the previous section. All three can be reasons to conclude that such-and-such a maxim could not be a universal law of nature. By analogy, there are many reasons on the basis of which we can conclude that cold necessarily follows the shining of the sun is not a universal law of nature. One reason is simply our experiences. Another reason is the nature of electromagnetic energy and its effect on mean kinetic molecular energy. Notice, however, that either reason is individually sufficient to conclude that cold necessarily follows the shining of the sun is not a law of nature.

**Universal Laws of Nature and Degrees of Generality**

In order to set up a small part of our response to the problems of false negatives and of false positives, we have one more topic to discuss under the heading of universal laws of nature. It is that universal laws of nature can be more or less general. For example, heat necessarily follows the shining of the sun is a specification of the more general rule that heat necessarily follows being in close proximity to objects with high mean kinetic molecular energy. That law is less general than the law that every alteration necessarily has a cause. Kant writes, “Although we learn any laws through experience, these are only particular determinations of yet higher laws, the highest of which (under which all others
stand) come from the understanding itself *a priori* (CPR A126; see also CPR B165 and CJ 20:209–216).

In like manner, the universal law of nature that falsely promising necessarily follows the need for money is less general than the universal law of nature that falsely promising necessarily follows the need for anything whatsoever. In fact, Kant seamlessly moves from the former law to the latter when in his second example he states the universality of the law as that “everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it” (GMM 4:422, noting that this law makes no reference to needing money). Moreover, the last universal law of nature is less general than the law that “everyone [is] permitted to deceive when he believe[s] it to be to his advantage” (CPrR 5:69), as Kant alludes to the false promisor in the second *Critique*. This is more general because not all deceptions are false promises.

Of course, we might state that last law at an even higher level of generality. Even more generally, we might understand the law to be that a person may use another as a mere means whenever it is to his advantage. Here we glimpse a possible connection of the first formulation to the second formulation but with a crucial difference. On the second formulation, the immorality of treating another as a mere means derives from the fact that humans are of ultimate worth (see Wood 2008: 112) or that his or her practical aiming confers value on his or her ends (see Korsgaard 1996: 239–41).11 When considered with respect to the first formulation, however, the question is whether this maxim—out of a desire to advance my own interests, I shall treat others as mere means—could be a

11 See also Arroyo (2011) for a discussion of these two views and a criticism of the latter.
universal law of nature. We can make that determination independently of whether humans are of absolute worth and independently of how that worth is derived.

Obviously, we will want to know whether it can be a universal law of nature that treating others as a mere means necessarily follows an action involving treating others as mere means being in one’s advantage. That it is not is evident from the practical consequences interpretation. The end under consideration is the securement of one’s own advantages. The means is treating others as mere means. However, treating others as a mere means whenever it will result in those people shunning your company and your loss of the advantages of mutual cooperation over the long run. Hence, it is not to your long-term advantage to treat others as mere means and so the end is impossible to attain by way of the means.

Responding to the Problems of False Negatives and False Positives

The Problems of False Negatives

We are now in a position to respond to the problem of false negatives. Recall the problem of the lawn fanatic, who acts on the maxim “I shall kill anyone who steps on my lawn without permission, out of love for my lawn.” Could this be a universal law of nature, the law that murdering unpermitted lawn-steppers necessarily follows the love for one’s lawn?

Brian Skyrms’s study of the stag hunt and interaction drives this point home. He writes, “Stag hunters quickly learn to visit only other stag hunters. This is not surprising. Visits to hare hunters are not reinforced while their visits to stag hunters are” (2007: 95). As anyone knows from experience, we shun the company of those who treat us as mere means but seek out the company of those who treat us with respect.
Shafer-Landau thinks so. Even J.S. Mill thinks Kant’s theory sanctions “the most outrageously immoral rules of conduct” (236).

We have two responses to this objection. First, as we emphasized above, Kant nowhere uses the formulations of the categorical imperative to prove that an action is moral. It does not follow from the fact that a maxim could be a universal law of nature that action on the maxim is moral. Kant’s claim is that:

Action on a maxim is moral only if the maxim could be a universal law of nature.

And not:

If a maxim could be a universal law of nature, then action on the maxim is moral.

The first formulation of the categorical imperative is a type of the categorical imperative itself. It tells us what the categorical imperative implies for us, in concreto, in the sensible world. The formulation tells us what is entailed by morality, not what entails morality.

Our position is that all three formulations together must jointly be employed to determine whether an action is moral. We can conclude that an action is moral only if it passes the test on all three formulations of the categorical imperative. Each alone is sufficient to determine whether an action is immoral. None alone is sufficient to determine

\[13\] Thus our stance is consistent with, but not identical to, Wood’s position that each of the formulations is “entirely correct from a systematic standpoint” but that neither alone is sufficient to cover every possible example (1999: 107). We need the second and third formulations to give more determinate content to our conception of morality.
whether an action is moral.\textsuperscript{14} However, if a maxim passes the test of all three formulations, we can conclude with high degrees of certainty that action on the maxim is moral (though, perhaps, not with absolute certainty).

If this account is correct, it follows that while “murdering unpermitted lawn-steppers necessarily follows the love of one’s lawn” might be a universal law of nature, the action does not pass the test of the second categorical imperative. Such a murderer is not treating the lawn-stepper as of absolute worth, for the murderer regards the lawn as having more worth than the lawn-stepper. Consequently, action on the maxim is immoral, just not according to the first formulation.\textsuperscript{15}

\textsuperscript{14} Of course, this comment applies to the second formulation of the categorical imperative, too. This has implications for the scope of the moral community, for just because we do not fail to treat rational nature as an end in itself and never as a means when we torture animals and so do not violate the second categorical imperative, it does not follow that acting on such a maxim could be a universal law of nature or could legislate for the universal kingdom of ends.

\textsuperscript{15} An important consequence pertaining to the logical contradiction interpretation of the first formulation follows from these considerations. Korsgaard (1996: 85–86) rejects this logical contradiction interpretation on the grounds that it cannot rule out actions of brutality, such as killing one’s child so as to get a good night’s sleep. However, even if that action should pass the test of the first formulation of the categorical imperative, it does not pass the test of the second formulation. Hence, if one wished to defend the logical contradiction interpretation alone (we do not, but if someone did), here is a ready response to Korsgaard’s objection.
Now we have little doubt that this interpretation of Kant will not be widely accepted because it is not the traditional rendering of Kant's moral theory (although we think it is the correct rendering and supported by textual evidence). So, we move to our second reply. Recall our previous comment that laws come in degrees of generality. Plausibly, the law “murdering unpermitted lawn-steppers necessarily follows the love of one's lawn” is a specific determination of the more general law that we may treat others a mere means to attaining a satisfactory state for ourselves. The satisfactory state is not being heartbroken by the ruthless treatment of one's lawn (after all, soil compaction from repeated stepping will kill one's lawn). The murder of the lawn-stepper ensures that one will not be heartbroken and so the lawn-stepper is used as mere means to keep from being heartbroken. Consequently, while the specific law “murdering unpermitted lawn-steppers necessarily follows the love one's lawn” does not obviously imply a contradiction, the law is itself a specification of a more general law that does obviously imply murdering the lawn-stepper is not morally permissible.

Our position is that this second reply is much weaker than the first reply, for we see no way to make principled determinations as to which laws are specifications of which laws. Nonetheless, we include this second line of response because we think some people may find it more convincing than the first. It highlights an interesting way to connect the first and second formulations, and we leave it open as a possibility worth exploring that there is a principled way to determine the specifications of various laws.

The Problem of False Positives
We turn now to the much more threatening, at least in our opinion, problem of false positives. The first formulation does tell us which actions are immoral. However, it seems as though a variety of apparently moral actions might turn out to be immoral. Wood (1999: 104–05) and Herman (1993: 138–39) provide several such examples. We shall consider the following three: the toy train example mentioned above; playing tennis on Sundays to avoid crowds; and selling one’s stocks when the Dow gains 1,000 more points. Our position is that action on none of three maxims turns out immoral on Kant’s First Formulation.

*Toy Train Buying and Never Selling:* The first example is the example of buying but not selling a no-longer manufactured (we shall just say vintage) toy train. What is the maxim here? Wood does not tell us the reason the person opts to buy a toy train, but presumably it is a love for them. Hence, the maxim is that “Out of a love for toy trains, I shall buy and never sell a vintage toy train.” Acting on this maxim, it seems, is immoral. For if everyone bought and never sold his or her vintage toy trains, then I would never be able to buy the toy train I desire.

However, this is not a properly stated maxim. Maxims consist of a single action and the reason it is done. Consider, again, falsely promising out of a need for money. The maxim of this action must be distinguished from the maxim that one will borrow money out of a need to pay one’s bills. In like manner, the action of buying a toy train and its maxim—out of a love for a vintage toy train, I shall buy it—must be distinguished from action of never selling a toy train—out of a love for a vintage toy train, I shall never sell it.
These are two separate actions and two separate maxims, and they must be separately evaluated to determine whether action on them is immoral.

We can think of no reason why action on either of the maxims should be immoral on any of the aforementioned interpretations of the first formulation *nor on any of the other three formulations of the categorical imperative*. Consequently, it is not immoral to buy and never sell a toy train.

Someone might object: suppose a tyrant has put a gun to your child’s head and demanded you sell your toy train. Is not it then immoral to never sell your toy train? The answer is: of course! But here the maxim is, out of love for my toy train, I shall have my child killed. That violates the second formulation of the categorical imperative much in the way murdering lawn-steppers does. Hence, it is immoral not to sell your toy train.

Someone will press this objection further: Doesn’t this mean that it is both not immoral and immoral to not sell one’s toy train? For if it is not immoral to never sell one’s train, then there is no occasion when it is not immoral to sell one’s toy train. However, there is an occasion when it is immoral not to sell one’s toy train.

There are two lines of response. The first is to admit that it is immoral to act on the maxim never to sell one’s toy train out of love for it, but it is not immoral to act on the maxim to *intend* never to sell one’s toy train, which is typically what people who never intend to sell their toy trains mean. As the old saw goes, we should never say never. A second line of reply is to defend a more general position regarding actions as abstentions from action over time. When your typical, sane person adopts as a maxim that he or she shall never engage in some action now and in the future, that maxim is really a general rule. Hence, to claim “I will never sell my toy train” is really just shorthand (among typical, sane
people) for “I adopt as a general rule never to sell my toy train,” full well knowing there might be unforeseen or unexpected exceptions that may be made to that general rule. Although these two lines of response are by no means identical, they have the same upshot as far as the current objection is concerned. It is at the very least not immoral to buy and intend never to sell a vintage toy train, which is the sort of plan of action your typical, sane person sets about when planning to by and never sell a vintage toy train.

Playing Tennis on Sundays: Herman provides another example of a maxim that seems innocuous but turns out to be immoral. The maxim is: out of a desire to avoid crowds, I shall play tennis on Sundays. Let us employ the practical consequences interpretation. Here, the aim is to avoid crowds, and the means is to play tennis on Sundays. However, in a world where everyone plays tennis on Sundays so as to avoid crowds, I can no longer attain my aim of avoiding crowds. Hence, it is immoral to play tennis on Sundays so as to avoid crowds.

Setting aside the fact Kant thinks we should be in church on Sunday mornings, it is surely not immoral to play tennis on Sundays so as to avoid crowds. Is there a response to this objection? We think so and it is that the maxim has not been stated correctly. Above, we noted that it is important to distinguish among the action, the action as it is done, and the reason the action is done. In our previous example of thievery from a bank for money, the action is thievery, the action as it is done is from a bank, and the reason it is done is for money.

The tennis maxim results in a contradiction only by sneaking the action as it is done into the maxim whereas maxims properly consist only of the action and the reason it is
The action is really to play tennis at an unpopular hour, not specifically on Sundays, for to do the action on Sundays is how or as the action is done. In like manner with respect to the thievery example, the action is to rob a place that has money rather than a bank in particular (after all, you cannot rob a place that does not have the thing you plan to rob). To rob a bank in particular is the action as it is done. Hence, the tennis maxim properly stated is: out of a desire to avoid crowds, I shall play tennis at an unpopular hour.\textsuperscript{16}

There is an obvious objection: Are not “at an unpopular hour” and “a place that has money” still the action as it is done, and if so why should not those conditions be excluded from the maxim, too? The answer is that the reason an action is done will still stipulate general—not specific—conditions pertaining to how the action is done. The fact one wishes to avoid crowds stipulates that the action of playing tennis shall be done at an unpopular hour but not specifically on Sunday. In like manner, the fact one wishes to obtain money stipulates that the thievery shall be done at a place that has money rather than, say, an empty outhouse. So, when we state that the maxim excludes the action as it is done, we mean the specifics of the action and not the general conditions entailed by the reason the action is done.

To continue, once we are clear about the maxim, we find that there is no reason it could not be a universal law of nature. For, if everyone goes to play tennis at unpopular

\textsuperscript{16}Another maxim Herman discusses is “out a desire to save money, I shall do next-year’s Christmas shopping during this year’s after-Christmas sales.” However, this maxim also Sneaks in the action as it is done. The maxim properly state is: “Out of a desire to save money, I shall do next-year’s Christmas shopping when prices are lowest (or lower).” Of course, that maxim could be a universal law of nature. It very nearly is!
hours, then each person will adjust his or her schedule to when that person expects the fewest people to be there. Yet this law would actually result in a better distribution of court time! So, making the present maxim a universal law of nature would actually be great; no one will be frustrated because he or she has to wait a long to use the courts, so long as people coordinate their times well.

_Selling Stocks:_ We now come to our last example of a morally innocuous maxim that appears to be immoral on Kant’s first formulation. The final example is: To ensure my financial well-being, I shall sell all of my stocks after the Dow gains its next 1,000 points. (Here again Wood does not provide the reason the action is done, and so we have provided one consistent with the apparent purpose of the action so that the maxim is correctly stated.) This maxim, it seems, cannot be a universal law of nature, for if everyone sells his or her stocks after the Dow gains the next 1,000 points, then I will not be able to sell my stocks to anyone.

We respond that unlike the other two examples, which were based on mistakes in stating one’s maxim, this maxim can be a universal law of nature. Let us begin with the practical consequences interpretation. Here, the aim is to secure my financial well-being. The means is selling my stocks. We now suppose that there is a world where everyone sells all of his or her stocks when the Dow gains 1,000 points. Does that mean I cannot secure my financial well-being? The answer is no. For just because everyone decides to sell all of the stocks he or she has _when_ the Dow gains 1,000 more points does not entail no one buy any more stocks _after_ the Dow gains 1,000 points.
In our judgment, the only way that a contradiction can be generated here is to conceive of the maxim as being that everyone sells all of their stocks when the Dow gains 1,000 points and no one buys any more stocks. However, as with the toy-train example, these would be two separate actions and two separate maxims, and each would have to be separately evaluated to determine whether action on it would be immoral.

Let us consider the teleological contradiction interpretation. Is there a contradiction here? We see no reason why. The purpose of commerce generally and of the selling of stocks specifically is to secure one’s financial well-being (not to mention, hopefully, to aid others in need). Hence, selling one’s stocks when the Dow gains 1,000 points does not contradict the purposes of commerce or of stock-selling.

Finally, we can see no obvious contradiction in this maxim, whether in conception, in our social practices, or in what all rational creatures would will if they gave it due consideration. As the present example does not come out immoral on any of the interpretations of the first formulation of the categorical imperative, we conclude that it does not constitute a case of a false positive.

Conclusion
In conclusion and in sum, we have been arguing for an account to the first formulation of the categorical imperative not as a restatement of the categorical imperative but as a consequence of it. If a maxim could not be a universal law of nature, we may conclude that action on the maxim is immoral. However, we cannot simply conclude from the fact that a maxim could be a universal law of nature that it is moral to act on the maxim. Rather, for action on a maxim to be moral, the maxim must pass every formulation of the categorical
imperative. The examples employed in the problem of false negatives may be universal laws of nature; however, they violate the second formulation of the categorical imperative. In contrast, the examples provided for the problem of false positives either fail to properly state the maxim on which an agent plans to act or actually are able to be universal laws of nature. Consequently, neither the problem of false positives nor the problem of false negatives, at least as they have been developed thus far, refute Kant's first formulation of the categorical imperative.

**Works Cited**

*Kant References*


Other References


