

## On Lying to Liars and Birthday Surprises

In several works, Immanuel Kant expresses a firm belief in an exceptionless duty not to lie, for instance he says “You shall absolutely speak the truth, is an imperative of reason, and in application a maxim which reason converts into a universal law.” (LE, 27: 496 (p. 264)) James Edwin Mahon, in *Kant and the Perfect Duty to Others Not to Lie*, argues that “no argument drawn from the Categorical Imperative can generate a perfect duty to others not to lie to others.” (655) In this paper, I will defend Kant’s view in response to the objections raised by Mahon.

I will begin with some definitions for terms used in this paper, followed by an explanation of Kant’s argument for the perfect duty not to lie, both using the Universal Law Formula and the Formula of Humanity. I will then present Mahon’s objection to the Universal Law Formula argument, and subsequently present my defense of Kant against this objection. Next, I shall present Mahon’s objection to the Formula of Humanity argument for the perfect duty not to lie, followed by my defense of Kant against this second objection.

To begin, I define the following:

**Maxim:** A principle that prescribes how one ought to act.

**Categorical Imperative (CI):** An imperative which declares an action to be objectively necessary without making any reference to a purpose. (GW 4:415)

**Universal Law Formula (ULF) of the Categorical Imperative:** ‘I ought never to act except in such a way that I could also will that my maxim should become a universal law’ (GW, 4: 402 (p. 57))

**Humanity as an End in Itself Formula (FH) of the Categorical Imperative:** ‘So act that you use humanity, whether in your own person or in the person of another, always at the same time as an end, never merely as a means.’ (GW, 4: 429 (p. 80))

**Humanity:** “The capacity to set ends through reason (GW 4:437), or the disposition that encompasses all rational capacities.” (PHI 304 Handout 6)

**Absolute Worth:** Most things have an exchange value or ‘price’, which can be compared to the price of other things. But ends in themselves have dignity, or a ‘value above all price’. (PHI 304 Handout 6)

**Duty:** “the necessity of an action from respect for law.” (GW 4:400)

**Contradiction in Conception (CC):** If a maxim cannot logically be a universalized law of nature, it fails the CC test, there is a logical contradiction in trying to conceive of such a maxim as universal.

**Contradiction in Willing (CW):** If a rational being cannot *will* for a maxim to be a universalized law of nature, it fails the CW test.

**Perfect Duty:** An exceptionless duty. When a maxim violates the CC test, acting on it violates a perfect duty.

**Imperfect Duty:** A duty that applies with some exceptions. A maxim that violates an imperfect duty passes the CC test but does not pass the CW test.

**Lie:** “An intentional untruthful declaration to another person” (SR 8:426 p.612)

Kant asserts that any maxim that allows us to lie to another person fails the Contradiction in Conception (CC) test, i.e. whatever purpose we mean to achieve through our lie, if such a maxim were universalized we could not possibly achieve that purpose. Consider the case where I ask my friend for money and promise to return it, knowing that I do not intend on keeping that promise; I utter an intentional untruth, i.e. a lie. Thus the maxim I follow is: It is permissible to lie to another person in order to gain money. However, if this maxim were universalized then everyone would lie to gain money. In that case, nobody would be willing to lend money because

they would know people lie about returning it. And so, in such a world it would be impossible for me to achieve my purpose: to borrow money. Hence, the maxim of lying to gain money leads to a contradiction in conception. Another more innocuous example is that of lying so as to not hurt people's feelings. Suppose your friend invites you to dinner, then asks you how the food was, and though the food was terrible, you lie so that your friend's feelings aren't hurt. The maxim followed is: I can lie if it prevents another person's feelings from being hurt. However, again if this maxim is universalized, no one would ever believe praise or affirmation since they know people are always aiming to sugar quote their response. Kant expands this to show that the act of lying for *any* purpose fails the CC test because if everyone began lying for that purpose, nobody would believe anyone lying in that specific way, "and thus my maxim, as soon as it were made a universal law, would have to destroy itself." (G, 4: 403 (p. 57))

This argument for a perfect duty to others to not lie comes from the Universal Law Formula. Kant also makes an argument for a perfect duty to not lie to others using the Humanity as an End in Itself Formula. He asserts that "the man who intends to make a false promise will immediately see that he intends to make use of another man merely as a means to an end which the latter does not likewise hold." (GW 4:430) In the example above, when I falsely promise to return the loaned money to my friend, I use them merely as a means to gain money, rather than respect their humanity, that is their capacity as free moral agents.

I now turn towards Mahon's first objection: the objection to the ULF argument for a perfect duty not to lie. To show that the ULF does not establish an *exceptionless* duty not to lie to others, Mahon uses the example of a maxim brought up by Christine M. Korsgaard:<sup>1</sup> *'When I believe that I am being lied to by another person I shall lie to the other person, in order to*

---

<sup>1</sup> *'The right to lie: Kant on dealing with evil'*, reprinted in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996) 137.

*counteract the intended results of her lie.*' (675) I subsequently refer to this as the Lying to Liars (LL) maxim.

The argument takes the following line of reasoning. When a person lies to another, the liar believes that the other person believes their lie. If the liar thought the other person knew they were lying, then they would know they aren't deceiving them, and thus they would not lie. As such, even if this maxim of lying to those who we believe are lying to us were to be universalized, the people who choose to lie in such a world would be convinced that their victim does not believe they are lying and so this maxim does not apply to them. Hence, if the victim follows this maxim they would be able to achieve their goal of counteracting the result of the other person's lie by lying back to them even in a world where this maxim is universalized.

To respond to Mahon's objection, I begin by unpacking the CC test. To show something is a perfect duty, one cannot conceive that such be the case, so we use reasoning roughly of the following form:

1. Jack wishes to achieve Goal A
2. To achieve Goal A, Jack needs to take action B
3. In a world where *everyone* takes action B in pursuit of Goal A, Goal A would be unattainable. (contradiction)
4. Thus to not take action B is a perfect duty.

There are three possible interpretations of the CC test. (PHI 304 Handout 5):

**Logical interpretation:** the resulting contradiction is internal to the universalized maxim itself

**Practical interpretation:** there is a practical contradiction between the goal you wish to achieve and what you would will in willing your maxim as a universal law of nature

**Teleological interpretation:** willing your maxim as a law of nature would contradict some natural purpose.

In Mahon's example of lying to liars, the Goal A is "to counteract the intended result of her (the liar's) lie". (676) What does this mean? I claim that there could be two distinct scenarios.

**Scenario 1:** Goal A is achieved by lying to those who lie and is defeated by lying to those who are truthful. For example, your goal is to administer some kind of retributive justice, where you wish to punish those who lie to you and reward those who don't.

**Scenario 2:** Goal A can be achieved by lying to someone, regardless of whether or not they are lying to you. But it can only be *universally* achieved by lying to liars. For example, you wish to trick people into giving you money. You make a false promise to return money to someone who you believe is lying to you for something. Now whether or not the person is actually lying, if they believe your lie and give you the money, your goal has been achieved.

To the first scenario, I respond with the practical interpretation of the CC test in mind. The crucial point to note is that in a universalized version of the LL maxim, we are acting based on the *belief that* someone is lying to us. This belief can be false. We could *think* that someone is lying to us, and consider it morally permissible to lie back to them. But we may be wrong in believing so - a case which Mahon does not address.

In this first scenario, in the cases when your belief that someone is lying is false, you will be defeating Goal A by following the maxim which was intended to lead to Goal A. For example, in the universalized version of LL, where you will always lie to those you believe are lying to you, you will inevitably defeat your goal of retributive justice by lying to those who you falsely believe are lying to you.

Now coming to the second scenario, I defend Kant in two ways. First, I respond with the practical interpretation of the CC test in mind. Consider the following propositions:

**Proposition A (Mahon implicitly accepts):** We cannot fulfill any Goal X by lying to someone who is *not* lying to us, in a universalized way.<sup>2</sup>

**Proposition B (Mahon's claim):** There is some Goal Y that we can achieve by lying to those who we believe are lying, in a universalized way.

But in any world where we universalize the LL maxim, Goal Y must be attainable based on our *belief* that someone is lying to us, not whether they actually *are* lying to us. Thus it must be attainable even in the circumstance that the other person is *not* lying to us, even though we may believe that they are lying to us. But by Proposition A, Goal Y is not attainable by lying to those who are *not* lying to us in a universalized way. And thus, in the universalized version of the LL maxim, Goal Y is made unattainable in certain circumstances by the maxim itself.

Another defense of Kant, in the second scenario, can be made using the teleological interpretation of the CC test.<sup>3</sup> Here I use the kind of argument that Kant makes in *On a Supposed Right to Lie from Philanthropy*. In response to Benjamin Constant's criticism that to lie to a murderer is ethically permissible because the murderer does not have the right to truth anymore, Kant responds by saying that with such a lie, "I bring it about, as far as I can, that statements (declarations) in general are not believed, and so too that all rights which are based on contracts come to nothing and lose their force." (8:426 SRL p 612) The argument Kant was making in this

---

<sup>2</sup> By "in a universalized way" I mean to indicate whether or not the maxim passes the CC test. This proposition is simply the part of Kant's ULF-based argument that Mahon does *not* object to. He only argues for the permissibility to lie to liars, implicitly accepting the impermissibility to lie to non-liars.

<sup>3</sup> There are arguments in support and against the teleological interpretation of the CC test, the main weakness being that it implies the existence of a natural purpose. In this paper, however, I set aside that debate and assume for the sake of argument that we agree on the teleological interpretation of the CC test.

passage had to do with the harm being caused to humanity by such a lie, but I use this idea to show how the LL maxim fails the CC test.

If everyone, universally, acted on the LL maxim, we would lie to those we *believe* are lying to us. But then from the perspective of a person who is *not* lying to us, we could always be *thinking that* they are lying, and thus be lying in response to them. Thus there will always be a possibility that the other person is lying to us even when we are not lying. Mahon's objection, as he himself identifies, rests on the fact that those who are lying to us inherently believe that we do not believe that they are lying to us, and thus will consider the LL maxim to be irrelevant. But Mahon does not consider the people who are *not* lying to us. These truthful people may think we believe that they *are* lying to us, and hence in turn we are lying to them because in this world the LL maxim exists as a universal maxim. And thus such doubt will bring it about that "statements (declarations) in general are not believed and so too that all rights based on contracts come to nothing and lose their force." (8:426 SRL p 612) And since all contractual agreements lose force, whatever goal we were attempting to achieve through our lie will not be attainable anymore. As such, the LL maxim fails the CC test in a teleological sense, where the natural purpose is to retain the force of rights based on contracts.

Now one may respond to the defenses I give above by altering the LL maxim to be as follows: "It is morally permissible to lie in the case that you are *certain* that someone else is lying to you." I call this the Lying to Liars when Certain (LLC) maxim.

Firstly, this is not Mahon's formulation, so we are already accepting that Mahon's argument is flawed. Now responding to the LLC maxim itself, what counts as certainty? Maybe hard, physical evidence? This itself is not quite enough, because as we know from Princeton honor code cases to crime cases, it is possible for physical evidence to be falsified or

misconstrued. So we still leave room open to be *wrong* about the fact that someone is lying.

Thus, if a maxim like this exists, the person who is not a liar but knows can be *thought to be lying* and thus lied back to based on some evidence that may incriminate him inauthentically, will always have reason to doubt the authenticity of what the other person is saying.

Okay, so maybe we can narrow it down even more: “It is morally permissible to lie in the case that you have unquestionable evidence that they are lying to you.” I call this the Lying to Liars when Unquestionably Certain (LLUC) maxim. For example, you personally saw someone at an event last night, but when you asked what they were upto, they said they were at home all day. In this case, we have two options:

1. Making an argument that we can never be *unquestionably* certain of any evidence that someone else is lying, because we could have been delusional, or hallucinating, or have seen a doppelganger, and so on ad infinitum. However, I find, and my intuition is that most people will, this line of reasoning unconvincing and somewhat of a sidestep.
2. Accepting that the LLUC maxim does not lead to a Contradiction in Conception and thus the Universal Law Formula of the Categorical Imperative does not establish an exceptionless duty not to lie to others.

I have argued that Mahon’s LL maxim fails the CC test. The modified alternative, the LLUC maxim, does pass the CC test. However, it would need to be successfully argued that it is possible to be unquestionably certain that someone is lying to you. Only then would it show that there is not a perfect duty not to lie to others.

Now, if we wish to further defend Kant’s view even in the face of the LLUC example, we need to turn to the alternative form of the Categorical Imperative, which is the Formula of Humanity. A liar is not any less valuable as a person than a non-liar, because people’s absolute

value comes not from their actions but their *capacity* as moral agents. If someone is *capable* of being rational and moral, then their lie doesn't make the humanity in them any less valuable. Lying to them involves using their humanity as a mere means rather than an end in itself, and thus lying to them violates the Formula of Humanity. Thus, using the FH, we establish that it is a perfect duty not to lie to others. Using this as a segue, I look at Mahon's second objection in which he argues that the Formula of Humanity does *not* successfully generate a perfect duty not to lie.

As explained in the beginning, the FH asserts that we ought never to act in a way that uses humanity, in ourselves or others, as a means to an end rather than an end in itself. What does it mean to treat someone as an end in itself, rather than a mere means to an end? Mahon uses Korsgaard's definition. In order for an action to not treat another person as a mere means, "it must at least be the case that it is logically possible for the other person to consent to the action". [Mahon 679] With this definition, examples of actions which treat others as mere means include coercion, rape, murder and lying. In each of these cases, it is not possible for the victim to consent to the action. If the victim gives their consent to being raped, then it is no longer rape. If a victim consents to being lied to, then it is no longer a lie, because the victim knows they are not being told the truth. So the argument for the perfect duty not to lie based on FH is as follows. The other person cannot possibly consent to being lied to, thus lying involves treating the other person as a mere means rather than an end in itself. Thus it is a perfect duty not to lie.

Mahon rejects this conclusion by stating that while it is logically impossible to consent to being told a *specific* lie, it *is* logically possible to consent to being told an *unspecific* lie when one does not know what exact lie they will be told. He gives the following example. If my friend asks me whether they can lie to me in the future in order to throw me a surprise birthday party, it

is logically possible for me to consent to such a lie because despite my consent, my friend can still successfully lie to me in the future, unlike in the case of me consenting to a particular lie.

I argue that Mahon's argument is unconvincing in two ways. Firstly, I argue that it is *not* possible to consent to being told an unspecific lie, simply by virtue of what it means to give consent. Secondly, I argue that even if we accept that one can consent to being told an unspecific lie, such consent violates the Formula of Humanity.

For clarity, it is important to establish what we mean by the term "consent". Here, I am interpreting the term to mean *informed consent* defined as "an agreement to do something or to allow something to happen, made *with complete knowledge of all relevant facts*, such as the risks involved or any available alternatives."<sup>4</sup> I claim that Korsgaard's definition of what it means to treat another as 'mere means' is reflective of Kant's beliefs only if we think of consent as *informed* consent. The point of 'not treating someone as a mere means', for Kant, is to respect their humanity. Humanity consists in our capacity to be free, rational, moral agents. Thus humanity is reflected only in *informed* consent, when the person giving consent is fully aware of what it is that they are consenting to and is thus fully using their capacities as free, rational agents.

My first response to Mahon's argument is that it is simply *not* possible to consent to being told an unspecific lie. This is because the definition of (informed) consent necessitates full knowledge on part of the individual giving consent. I can not give informed consent to something that I do not have full knowledge of. I cannot possess full knowledge of an unspecific lie, otherwise it won't be unspecific. Thus I cannot give informed consent to being told an unspecific lie. There is a fundamental contradiction in the idea of informed consent and being told an unspecific lie.

---

<sup>4</sup> As defined by Cornell Legal Information Institute <[https://www.law.cornell.edu/wex/informed\\_consent](https://www.law.cornell.edu/wex/informed_consent)>

My second response to Mahon's argument is that even if we accept that it *is* possible to consent to being told an unspecific lie, giving such consent violates the Formula of Humanity. This is simply because giving such consent involves treating yourself as a means to an end and denying yourself your humanity. Mahon considers this possibility but claims that it is difficult to "demonstrate that it is possible to use humanity in oneself as a mere means," (684) rather than an end in itself. I think in this specific case, it is actually quite simple to understand what it means to use oneself as a mere means. Kant asserts that a person's value comes from their capacity to act freely and rationally as moral agents, which is what he terms "humanity". To consent to being deceived is to consent to being treated not as a free rational agent. Consenting to being told a lie is essentially consenting to not being asked for consent in the future. This is basically synonymous to consenting to not being treated as a free, rational agent in the future. Which is clearly giving up your own humanity. And the Formula of Humanity necessitates that you respect not just the humanity in others but also in yourself. Thus, consenting to being told an unspecific lie violates one's own humanity and thus violates the FH.

Hence, Mahon's case for lying to those who have 'consented' to being told an unspecific lie is unconvincing. He does not successfully show that the FH fails to lead us to a perfect duty not to lie to others.

In this paper I have responded to both of Mahon's objections. First, I have shown that Mahon's case for lying to those we believe are lying to us is invalidated by the Universal Law Formula, though an alternative version of such a maxim may be used to make a similar argument. I claim that this alternative maxim can be invalidated using the FH instead. Second, I have shown that Mahon's case for lying to those who have consented to being told an unspecific lie is both a definitional contradiction and a violation of the Formula of Humanity. Hence, I have

shown that Mahon has not successfully argued that “no argument drawn from the Categorical Imperative can generate a perfect duty to others not to lie to others.” (655)

## References

Kant, Immanuel, and Mary Gregor. "Kant Practical Philosophy." (1996).

Mahon, James Edwin. "Kant and the perfect duty to others not to lie." *British Journal for the History of Philosophy* 14.4 (2006): 653-685.