Pauline Kleingeld*

Self-Contradictions of the Will: Reply to Jens Timmermann

https://doi.org/10.1515/kant-2021-0032

Abstract: In this article, I reply to Jens Timmermann’s critical discussion of my essay “Contradiction and Kant’s Formula of Universal Law”. I first consider Timmermann’s reasons for rejecting my interpretation of the Formula of Universal Law. I argue that the self-contradiction relevant to determining a maxim’s moral status should not be sought in the imagined world in which the maxim is a universal law. I then discuss Timmermann’s suggestion that something like a volitional self-contradiction is found within the will of the immoral agent. I deny this and clarify that the relevant contradiction is diagnosed counterfactually in moral reflection. Finally, I explain the differences between Timmermann’s account, Korsgaard’s Practical Contradiction interpretation, and my own Volitional Self-Contradiction interpretation.

Keywords: Formula of Universal Law, categorical imperative, contradiction, maxim, practical contradiction, volitional self-contradiction.

Introduction

In this paper, I reply to Jens Timmermann’s comments on my essay “Contradiction and Kant’s Formula of Universal Law”, both published in this journal (Kleingeld 2017; Timmermann 2018). I thank the editors of Kant-Studien for allowing me to do so. I am deeply grateful to Timmermann for his critical engagement with my essay and for countless rewarding discussions about Kant’s moral theory over the past fifteen years.¹

In the 2017 essay, I propose a novel interpretation of Kant’s Formula of Universal Law (FUL), the most prominent formulation of the Categorical Imperative in the Groundwork. The Formula reads as follows:

1 I also thank Sorin Baiasu for organizing the 2018 KOSAK conference at the University of Keele, where Timmermann and I presented earlier versions of the comments and reply.

*Kontakt: Prof. Dr. Pauline Kleingeld, University of Groningen, Faculty of Philosophy, Oude Boteringestraat 52, 9712 GL Groningen, Netherlands; pauline.kleingeld@rug.nl
Act only in accordance with that maxim through which you can simultaneously will that it become a universal law.² (GMS, AA 04: 421, emphasis in original)

In restatements of this principle in the literature, the word ‘simultaneously’ is routinely omitted without further discussion or justification. Accordingly, the FUL is commonly thought to demand that you act only on maxims that you can will to become universal laws without contradiction. The Formula is sometimes shrunk even further to the requirement that you act only on maxims that are ‘universalizable’. Kant, however, includes the word ‘simultaneously’ (or in some versions, ‘also’) not merely in the FUL but in virtually all formulations of the Categorical Imperative.³

In the 2017 essay, I propose that the Formula be read as demanding that you be able to will two things simultaneously concerning any of your maxims of action. You should be able not only (1) to will to act on the maxim but also, simultaneously, (2) to will that the maxim become a universal law (‘through’ your adopting it).⁴ In order to find out whether you can, you must examine whether these two volitions are compatible. I take Kant to argue that acting on a given maxim is morally impermissible if willing both simultaneously would lead to a self-contradiction of the will, or, in other words, to a volitional self-contradiction. There are two ways in which maxims can fail this test, corresponding to Kant’s distinction between perfect and imperfect duties.

Thus, I understand the FUL to be associated with a possibility test that is carried out in moral reflection, and I take the relevant self-contradiction to lie between willing the maxim as your own action principle and willing that it become a universal law. The addition of ‘simultaneously’ indicates that the test concerns the compossibility of two volitions.⁵ Consider Kant’s example of the

² “Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daß sie ein allgemeines Gesetz werde.”
³ For representative statements from leading Kant scholars and Kantian ethicists, and for a list of formulas of the Categorical Imperative, see Kleingeld 2017, 91–92 and 97–99. One important exception is the Formula of the Law of Nature (GMS, AA 04: 421), and I offer a possible explanation (110 n.48). More importantly, Kant’s second statement of the Formula of the Law of Nature does include ‘simultaneously’ (“simultaneously as universal laws of nature” [zugleich als allgemeine Naturgesetze], GMS AA 04: 437).
⁴ In many other passages, Kant articulates the idea of giving universal laws through one’s maxim, drawing on an analogy with political legislation (GMS, AA 04: 432, 433, 434, 438, 439, 440).
⁵ Further, the inclusion of ‘simultaneously’ is to invalidate willing the one after the other, as when you first will to act on a given maxim and then will that it become a universal law. If you first pocket my money with a false promise to repay me, you can then gladly will that your maxim become a universal law – universalizing it makes it henceforth impossible to obtain money in this way, but at that point you will have what you need. This option is blocked by the requirement
maxim of falsely promising, when one believes oneself to be in need of money, to repay a loan (GMS, AA 04: 422). On the Volitional Self-Contradiction interpretation, you should ask yourself whether you can will to act on the maxim and simultaneously will that it become a universal law. Given that this maxim cannot even be conceived as a universal law – as Kant argues – the answer is negative. I argue that this reading of the FUL has several significant interpretive and philosophical advantages. Among other things, the proposed account explains the relevant self-contradiction in terms of the ordinary notion of ‘contradiction’, and it explains how this self-contradiction is generated without presupposing substantive values or other extraneous assumptions.

In his critical commentary, Timmermann starts with an apt description of my interpretation of the FUL as involving the “requirement that we must be able, without contradiction, to will our maxim and simultaneously will that it be a universal law” (582). He agrees with my criticism of the standard way of contrasting ‘contradictions in conception’ and ‘contradictions in the will’. He finds my argument that all maxims that fail the test yield self-contradictions of the will convincing (though he often leaves out the ‘self’ in ‘self-contradiction’). He also concurs with my statement that the significance of the simultaneity condition in the wording of the FUL merits more scholarly attention.

Timmermann rejects the new interpretation of the FUL that I propose, however, endorsing Korsgaard’s Practical Contradiction interpretation instead. He further asserts that the idea of a ‘volitional self-contradiction’ should rather be understood as referring to a conflict within the will of an immoral agent.

In what follows, I first consider Timmermann’s reason for rejecting my interpretation of the FUL (section 1). I then discuss his suggestion that something like a volitional self-contradiction is found within the will of the immoral agent (section 2). Finally, I explain in more detail the differences between Korsgaard’s Practical Contradiction interpretation, Timmermann’s own account, and my proposed Volitional Self-Contradiction interpretation (section 3).

1 Who does the test, and where?

Timmermann’s primary reason for rejecting my interpretation of the FUL is his assumption that the contradiction that is relevant to determining the moral status of maxims must be sought within the imagined world in which the maxim is that you be able to will both ‘simultaneously’. Relatedly, Kant notes in the Critique of Pure Reason that one can say without contradiction that a man is young and that he is old, since he can be young first and old later, but that he cannot be young and old simultaneously (KrV, B 192).
a universal law. On my proposal, by contrast, the contradiction is not moved into that world. The agent who engages in moral reflection is to check for a contradiction between (hypothetically) willing to act on the maxim and (hypothetically) willing that it become a universal law. Examining the compossibility of these two volitions does not involve moving the contradiction into the imagined world of the universalized maxim.

Timmermann’s central argument in support of his claim that the contradiction must be located in the imagined world of the universalized maxim is that “the agent must be understood as part of that world” (581). On his view, this entails that Kant’s thought experiment requires that “I transfer myself into the world of universalization; and I take with me the [tested] maxim…” (586, cf. 585). The question “whether I can will the world of universalization” must then be answered by my imagined self in that imagined world, or, as Timmermann puts it, “in the world of universalization” (587–8). On his view, the moral status of maxims is “decided by the agent in that world” (595, Timmermann’s emphasis).

Thus, on Timmermann’s understanding of the test, morally reflecting agents must replicate themselves in thought. They need to imagine the world of the universalized maxim, and then they need to imagine themselves asking, in that world, whether they can will that the tested maxim become a universal law.

I did not explicitly discuss this conception of the test in my 2017 essay, but it is interesting and merits discussion. Four considerations seem to me to speak against it, however.

First, the fact that you must be understood to be part of the imagined ‘world of universalization’ does not entail – contrary to Timmermann’s assumption – that the test must be conducted in that world and that the relevant contradiction must be found there, in the will of your imagined self (591–595). Timmermann does not cite textual evidence to show that Kant indeed shifts the crucial question to your imagined self in the imagined world. The FUL as stated is addressed to you; Kant does not write that the crucial question must be answered – as Timmermann asserts – “by the agent in that world”.

Second, on Kant’s view, in the case of some maxims it is impossible even to conceive them as universal laws without contradiction, as with his well-known example of the maxim of falsely promising, when one believes oneself to be in need of money, to repay a loan. In the case of such maxims, you cannot even be ‘in the world of universalization’. If this world is inconceivable, there is nowhere for you and your tested maxim to ‘transfer into’ (to use Timmermann’s phrase), let alone for your imagined self ‘in that world’ to determine the moral status of the maxim.

Third, in the case of other maxims, the tested maxim can be conceived as a universal law. In such cases, however, it is hard to see how it can make sense for
your imagined self, in an imagined world in which such a maxim is a universal law, to ask whether the maxim can be willed to ‘become’ a universal law (as the FUL puts it). By hypothesis, it is already a universal law. On Kant’s conception, your will is your capacity to act on the basis of principles; that is, it is a form of causality, a capacity to pursue, do or bring about something for certain reasons (GMS, AA 04: 412). It is unclear what it could even mean for your imagined self in the world of the universalized maxim to assess the possibility of willing that a maxim that is a universal law become a universal law.

Fourth, in his discussion of the four examples in the *Groundwork*, Kant’s focus is consistently on the man who engages in moral reflection. In the first two examples, this man is said to realize that it is impossible for the tested maxim ever to be a universal law. In the second set of examples, the man is said to realize that although the maxim can become a universal law, it is impossible for him to will that it become one. At no point does Kant write that the man should mentally transfer himself into the world of the universalized maxim in order to answer the moral question there.

In sum, the moral status of maxims is to be evaluated by the agent addressed by the Categorical Imperative: by you, in moral reflection, as I claimed in the earlier essay (Kleingeld 2017: 102, 103), not by your imagined self in the imagined world. The question whether a maxim is morally permissible is asked and answered by you, in your armchair, at the barricades, at work, or wherever else you engage in moral reflection, together with others or alone. You must determine whether you can will to act on the tested maxim and simultaneously will that it become a universal law. If, in moral reflection, you find that this would yield a contradiction, this contradiction is ‘found’ counterfactually – as Kant’s subjunctives in the examples he discusses indicate. For example, he writes that in some cases it is impossible simultaneously to will that the tested maxim become a universal law, for this would mean that your will “would be in conflict with itself” [würde sich selbst widerstreiten] (GMS, AA 04: 423).

I would like to add a comment in response to Timmermann’s remark that the expressions ‘my maxim’ and ‘the maxim of your action’ in formulas of the Categorical Imperative apply only to “potential or available” maxims – maxims the agent is “free to adopt” (584) – rather than already adopted maxims. If Timmermann means that the formulas apply only to not-yet-adopted maxims, then he and I hold different views. Nothing in Kant’s argument rules out critically eval-

---

6 I agree with Timmermann that it would be “very odd if some formulations of the categorical imperative were to refer to maxims already adopted, and others to maxims merely under consideration” (584, n.3). It seems to me, however, that the formulas can apply to both.
uating maxims you have already adopted, say after receiving moral criticism (or after reading the *Groundwork*, for that matter). Nor does his argument rule out testing maxims attributed to others or maxims developed merely as a philosophical exercise, to see what follows from the Categorical Imperative – as Kant’s own examples illustrate.

2 The will of the immoral agent

Timmermann claims that the self-contradiction of the type I discuss (or something very much like it) is located “within the will of the immoral agent” (581). He ascribes to me the view that the volitional self-contradiction “is located ... within the [immoral] agent’s conflicted will” (582, cf. 588, 592, 593). He does not show which part of my argument supposedly commits me to this view, and the grounds for his ascription of it to me are not clear. On my interpretation, if it is impossible to will to act on a tested maxim and simultaneously to will that it become a universal law – that is, if the maxim generates a volitional self-contradiction when submitted to the FUL test – then acting on the maxim is morally impermissible. It would be fallacious to infer from this that, on my interpretation, if someone acts on a morally impermissible maxim, then their will contains a volitional self-contradiction. On my reading, Kant’s point is precisely that impermissible maxims are those that cannot simultaneously be willed as maxim and as universal law. Thus, immoral agents do not will both simultaneously. Perhaps Timmermann assumes that if the self-contradiction is not to be found in the will of the imagined agent in the imagined world of universalization – which is where he locates it – then it must lie in the will of an immoral agent. But why? It is unclear – to me at least – how this would follow from my argument.

Let me therefore clarify that I do not hold the view that, according to Kant, the actual wills of immoral agents contain volitional self-contradictions of the type I discuss. I explain in the essay that the volitional self-contradiction is found “in moral reflection”, when we “examine” a (failing) maxim in light of the FUL (e.g., 102, 103, 111), and I discuss the test as a possibility test throughout. On my reading, Kant is not saying that you ought actually simultaneously to will that your maxim become a universal law but that you should be able to without contradiction. In this last regard, my view does not differ from most other accounts of how Kant’s thought experiment is supposed to work.

---

7 For example, I do not describe Kant as arguing that “a will that adopts an immoral maxim is in conflict with itself” (cf. Timmermann 2018: 582).
There are several reasons for thinking that the actual wills of immoral agents cannot contain volitional self-contradictions of the type I discuss. For one thing, in his first two examples Kant argues that the maxims at issue cannot even be conceived as universal laws, let alone willed to become such (GMS, AA 04:422, 424). And in his discussion of the second set of examples, where the maxims could be universal laws, he argues that it is again “impossible to will” [unmöglich, zu wollen] that they become universal laws (GMS, AA 04: 423, 424). If it is impossible to will that these maxims become universal laws, then it is impossible for an immoral will actually to contain a self-contradiction of the type at issue – that is, it is impossible simultaneously to will to act on an impermissible maxim and to will that it become a universal law.

Furthermore, it is indeed evident that immoral agents do not will that their impermissible maxims become universal laws (unless they are very confused). After all, this would run counter to their acting on their impermissible maxims, maxims to which, by hypothesis, they are committed. Someone who wills to act on the maxim of promising falsely to repay a loan, to get rich fast, certainly does not simultaneously will that this maxim become a universal law, as this would make acting on the maxim impossible. Accordingly, Kant writes that in cases of our own violations of duty we find that “we really do not will that our maxim become a universal law”8 (GMS, AA 04: 424): we adopt one maxim for ourselves while willing another maxim to hold as a general law for others. And this combination of volitions does not yield a contradiction (GMS, AA 04: 424).

How, then, should we describe the will of an agent who violates the demands of the Categorical Imperative, if not in terms of self-contradiction? Imperatives, Kant writes, are “laws that must be obeyed, that is, must be followed even against inclination”9 (GMS, AA 04: 416). Adopting a maxim that you cannot simultaneously will to become a universal law constitutes not a contradiction but disobedience, transgression, offence, wrongdoing, and the like. Kant describes moral disobedience or transgression (Übertretung) in terms of ‘antagonism’: there is a “resistance of inclination to the prescript of reason (antagonismus)”10 (GMS, AA 04: 424), as Timmermann rightly notes.

Thus, the will of the immoral agent contains a disharmony of a different kind. Kant argues that the immoral agent is aware of violating moral requirements and feels the pangs of conscience. Conscience is the capacity for moral self-assessment, the ‘inner judge’ (MS, AA 06:438). It is the capacity to judge whether, in

---

8 “daß wir wirklich nicht wollen, es solle unsere Maxime ein allgemeines Gesetz werden, denn das ist uns unmöglich.”
9 “Gesetze, denen gehorcht, d. i. auch wider Neigung Folge geleistet, werden muß.”
10 “Widerstand der Neigung gegen die Vorschrift der Vernunft (antagonismus).”
your practical reasoning and acting, you have done what is morally demanded. Conscience is “practical reason holding the human being’s duty before him for his acquittal or condemnation in every case that comes under the law”\textsuperscript{11} (MS, AA 06: 400). But the self-condemnation of an agent is not the same as a self-contradiction of the will.

In sum, the notion of a volitional self-contradiction does not describe the content or structure of the will of an immoral agent. It is used to describe a contradiction that emerges counterfactually, during a test used in moral reflection, in cases of maxims that fail this test. If a maxim fails the test, this means that adopting the maxim is morally impermissible. But if a maxim is impermissible, this does not mean that there is a self-contradiction in the will of someone who acts on it. \textit{Mutatis mutandis}, the same is true on standard interpretations of the FUL: immoral agents are not usually conceived of as actually embracing ‘contradictions in conception’ or as actually having ‘contradictions in the will’.

\section{Practical contradictions and self-contradictions of the will}

Timmermann defends Korsgaard’s ‘Practical Contradiction’ reading of the FUL. Since he criticizes my reading and applies its core notion of a volitional self-contradiction to the will of the immoral agent, he claims that my proposal is not a “rival” but a “supplement” to the Practical Contradiction interpretation. In the previous two sections, I defended my reading of the FUL against Timmermann’s objections and argued that my conception of the ‘self-contradiction of the will’ does not describe the will of an immoral agent. In this section, I situate the Volitional Self-Contradiction interpretation of the FUL in relation to Korsgaard’s and Timmermann’s accounts.

Korsgaard’s interpretation of the FUL turns on a “specifically practical sense of ‘contradiction’” (Korsgaard 1996: 93), that is, a different \textit{notion} of ‘contradiction’. She sees this as distinct from the ordinary notion, which she calls a ‘logical’ contradiction. The latter consists in simultaneously claiming \textit{A} and not-\textit{A}, or, applied to the will, willing \textit{A} and not-\textit{A}. By contrast, on Korsgaard’s view, a ‘practical contradiction’ consists in willing the thwarting of one’s own purpose. She develops this idea on the basis of Kant’s claim that willing an

\footnote{11 “Denn Gewissen ist die dem Menschen in jedem Fall eines Gesetzes seine Pflicht zum Lossprechen oder Verurtheilen vorhaltende praktische Vernunft.”}
end rationally entails willing the indispensably necessary means to it that lie in your power (GMS, AA 04: 417). Since Kant presents this as an analytic claim, that is, as a matter of conceptual entailment, she argues that willing the frustration or thwarting of your end is appropriately called a practical contradiction (93–94).

Korsgaard argues that a contradiction in this specifically practical sense is not best understood as a contradiction between simultaneously willing and not willing the end (93). Rather, it is a contradiction that occurs within a single volition: in willing the frustration of your end. This makes the Practical Contradiction interpretation seem particularly well suited to accounting for the contradiction that emerges in the case of maxims that fail the FUL test, on the common understanding of the formula. After all, on the common reading the Categorical Imperative demands that we act on maxims that we can will as universal laws. Korsgaard’s Practical Contradiction interpretation provides an elegant and ingenious account of how, in the case of maxims that fail the test, willing a maxim as a universal law, simply as such, can contain a (‘practical’) contradiction.

On her account, there are two ways in which practical contradictions can be generated, corresponding to Kant’s distinction between perfect and imperfect duties, but in both cases the practical contradiction is found in willing an immoral maxim to become a universal law. In some cases, the purpose specified in the maxim would be impossible to achieve in a world in which the maxim were a universal law (as in Kant’s false promising example). As a result, willing this maxim to become a universal law amounts to willing the thwarting of your own purpose – a practical contradiction (92). In other cases, the purpose of the maxim is not thwarted if the maxim is a universal law, but, Korsgaard suggests, a practical contradiction may still be generated with some maxims if we assume that there are “essential purposes of the will” (96). One such purpose, Korsgaard suggests, may be effectiveness, and she assumes that this helps to generate a practical contradiction for Kant’s example of never helping. If the maxim of never helping were to become a universal law, this would thwart your will’s effectiveness, since you would not receive the help you need. Thus, your willing this maxim to be a universal law would amount to willing the thwarting of your purpose – again, a practical contradiction.12

The main difference between Korsgaard’s reading and the Volitional Self-Contradiction interpretation, for the purposes of this reply, lies in the use of different notions of ‘contradiction’ and, as a result, in different accounts of the salient contradiction.13 On my proposal, the relevant contradiction lies not within a single

12 I discuss other aspects of the relation between Korsgaard’s Practical Contradiction interpretation and the Volitional Self-Contradiction interpretation in Kleingeld (forthcoming). In that essay I also discuss the Logical Contradiction interpretation.
13 In Kleingeld (forthcoming) I argue that there are several additional advantages.
volition (that is, within willing the maxim to become a universal law) but between two, namely between (1) willing that the maxim be one’s own action principle and (2) willing that the maxim become a universal law. Since it involves two elements, that contradiction can be explicated in terms of willing A while also willing not-A. As a result, the ordinary notion of ‘contradiction’ suffices.

This does not mean that I “reject” the notion of a practical contradiction (Timmermann 2018: 587). Nor do I deny that willing the thwarting of your purpose is a form of practical irrationality. It is. On my interpretation of the FUL, it is simply unnecessary to introduce the novel notion of a ‘practical contradiction’. I take this to be an advantage, since Kant also does not introduce a new notion of ‘contradiction’ to explain the application of the FUL.

How does Timmermann’s own account of the FUL relate to Korsgaard’s? On Timmermann’s account, the “principal problem with immoral maxims” is that universalization would “make the candidate maxim means-ends irrational” (587–588). It is not immediately clear how to apply the criterion, thus understood. For instance, Kant’s diagnosis of the problem in the second example – the case of the maxim of false promising – is not that universalization makes the maxim means-ends irrational but that universalization is impossible. Furthermore, it is not immediately clear how the maxim of indifference to others – which is how Timmermann construes the maxim in Kant’s fourth example – becomes “means-ends irrational” when universalized. Universalization does not make your acting on this maxim ineffective. Indifferent people can still be wholly indifferent to others even when all others are indifferent to them.14

It seems that Timmermann’s account of the means-ends irrationality at issue should be understood not so much in terms of Korsgaard’s notion of a practical contradiction but in terms of his own background account of practical deliberation as a two-stage process. He sees practical deliberation as triggered by inclination and initially oriented towards our long-term interest. Maxims are “first formulated as a result of prudential, means-ends reasoning” (584). Only subsequently, in the next “stage”, does the agent ask whether the maxim is endorsed by pure practical reason (585).15 In the process of answering this new question,
the agent is to determine whether universalization of the maxim runs counter to
self-interest. In this way, Timmermann’s account of the contradiction is anchored
in the prudential commitments he sees as basic. Here is how he explains the rel-
evant contradiction that emerges during the test:

[A] will placed in the world of the universalized maxim would contradict itself in the sense
that a rule initially devised on prudential grounds ceases to be means-ends rational. (592)

He also writes:

[T]he contradiction in the will ... is a contradiction in a fairly ordinary sense: the will as
imagined in the world of universalisation is committed to willing and to not willing the
maxim. It is committed to willing the maxim because it was the will’s sole rational posses-
sion as it entered the sphere of pure reason; it is committed to its opposite because it turns
out not to be means-ends rational ... in that sphere. (589)

Rather than explicating the contradiction in terms of a practical contradiction,
Timmermann does so in terms of ‘willing and not willing’ the maxim and in terms
of what he calls a “fairly ordinary sense” of ‘contradiction’. Thus, his analysis
of the contradiction seems to differ much more significantly from Korsgaard’s
account than his endorsement of her Practical Contradiction interpretation
would suggest.16

I cannot enter into a discussion of Timmermann’s broader background
picture here, since this would go far beyond the scope of a reply to his discus-
sion of my essay. Moreover, for similar reasons, Timmermann was unable to
develop this picture in full detail in his commentary. It is clear, however, that
his conception of the procedure for determining the moral status of maxims
is based on his background account of practical deliberation. It seems, then,
that the disagreements between us run much deeper than a mere difference of
opinion on how best to interpret the FUL and that we therefore have much more
to discuss.17

16 Timmermann’s construal of Kant’s fourth example here may in fact be closer to Mark Tim-
mons’s Logical Contradiction interpretation; see Timmons (forthcoming).
17 I would like to thank Fiorella Tomassini, Janis Schaab, Joel Anderson, Leon van Rijsber-
gen, Marijana Vujosevic, and Michael Gregory for helpful comments. I am grateful to the Dutch
Research Council (NWO) for research support.
References


