



A Defense and Development of the Volitional Self-Contradiction Interpretation

Pauline Kleingeld¹

Received: 1 March 2023 / Revised: 1 March 2023 / Accepted: 3 March 2023 / Published online: 12 April 2023
© The Author(s) 2023

Abstract

Kant’s Formula of Universal Law (FUL) is generally believed to require you to act only on the basis of maxims that you can will without contradiction to become universal laws. In “Contradiction and Kant’s Formula of Universal Law” (2017), I have proposed to read the FUL instead as requiring that, for any maxim on which you act, you can will two things simultaneously, without volitional self-contradiction: (1) willing the maxim as your own action principle and (2) willing that it become a universal law. In the present essay, I reply to comments by Mark Timmons, Michael Walschots, Paola Romero, and Stefano Lo Re. In response to their comments concerning the application of the FUL, I expand the interpretive framework of the Volitional Self-Contradiction Interpretation. I argue that Kant also constructs the diagnostic volitional self-contradiction as a contradiction between (1) willing a tested maxim to become a universal law and (2) willing what humans, qua finite rational beings, necessarily will, namely the means to their actual and possible future ends. I also clarify how the two ways in which Kant specifies the test are related. Furthermore, I clarify the relation between the Volitional Self-Contradiction interpretation and other interpretations of the FUL, in particular the ‘Practical Contradiction’ and ‘Logical Contradiction’ interpretations, as well as its difference from the Golden Rule. I also address the objection that the volitional self-contradiction is superfluous, clarify the relation between the Formula of Universal Law and the Formula of the Law of Nature, and explain why the will of an immoral agent does not contain a volitional self-contradiction of the type at issue.

Keywords Contradiction · Formula of Universal Law · Immanuel Kant · Logical contradiction · Practical contradiction · Volitional self-contradiction

✉ Pauline Kleingeld
pauline.kleingeld@rug.nl

¹ Faculty of Philosophy, University of Groningen, Groningen, The Netherlands

1 Introduction

In ‘Contradiction and Kant’s Formula of Universal Law’ (Kleingeld, 2017), I focus on Kant’s principal formulation of the Categorical Imperative in the *Groundwork*, known as the ‘Formula of Universal Law’ (FUL):

Act only in accordance with that maxim through which you can simultaneously will that it become a universal law (GMS 4:421, original emphasis).¹

In restatements of this principle, both in the Kant literature and beyond,² the word ‘simultaneously’ is routinely omitted. The Formula is generally believed to require you to act only on the basis of maxims *that you can will to become universal laws* without contradiction. Accordingly, the FUL is associated with a test, conducted in moral reflection, to establish whether a contradiction would emerge. Although there is considerable disagreement about the precise nature of the test and the contradiction involved, there is widespread agreement that the crucial question concerning any maxim under scrutiny is:

MORAL QUESTION (on traditional interpretations of the FUL):

Can you will that this maxim become a universal law without this volition containing a contradiction?

Kant, however, includes ‘simultaneously’ or ‘also’ in virtually all formulations of the Categorical Imperative—and in *all* of its most prominent statements—in the *Groundwork* and elsewhere.³ In my 2017 essay, I proposed an interpretation according to which ‘simultaneously’ is essential to the meaning of the Categorical Imperative, such that omitting it alters the nature of the moral requirement.

The core of this interpretive proposal is to read the FUL as requiring that, for any maxim on which you act, you can will two things simultaneously. In the essay, I specify this requirement as follows: you should act only on maxims that you can *both* (1) will as your own maxim *and simultaneously* (2) will to become a universal law (‘through’ your adopting the maxim, immediately). In contrast to other interpretations, then, I propose that we locate the relevant contradiction, for any maxim that fails to satisfy this requirement, not *within* willing that the maxim become a universal law but *between* (1) willing the maxim as your own action principle and (2) willing that it become a universal law. If willing both things simultaneously would lead to a self-contradiction of the will—or synonymously, a ‘volitional self-contradiction’—

¹ All references to Kant’s work cite volume and page number of the *Akademie-Ausgabe*. Translations are taken from the edition listed in the References (Kant 1996), though I have sometimes altered them. Abbreviations: Anth: Anthropology from a Pragmatic Point of View (*Anthropologie in pragmatischer Hinsicht*); GMS: Groundwork for the Metaphysics of Morals (*Grundlegung zur Metaphysik der Sitten*); KpV: Critique of Practical Reason (*Kritik der praktischen Vernunft*); MS: Metaphysics of Morals (*Metaphysik der Sitten*); V-Met-L1/Pölitz=Pölitz Lectures on Metaphysics (*Metaphysik L1 Pölitz*).

² For representative statements from leading Kant scholars, Kantian ethicists and others, see Kleingeld 2017, 91–92.

³ For an overview, see Kleingeld 2017, 97–99.

then acting on this maxim is morally impermissible. The crucial question concerning any maxim under scrutiny thus becomes:

MORAL QUESTION (on the Volitional Self-Contradiction Interpretation):

Can you (1) will to act on this maxim *and simultaneously* (2) will that it become a universal law, without these two volitions contradicting each other?

This proposal runs counter to a long tradition of Kant interpretation, so I am honored that it has been taken seriously and am deeply grateful for Mark Timmons's, Michael Walschots's, Paola Romero's, and Stefano Lo Re's perceptive comments. I gladly note that none of them challenges my contention that the wording of the FUL permits this reading or that the contradiction, on my reconstruction of it, is a contradiction in the ordinary sense of the term. Instead, they challenge my claim that the Volitional Self-Contradiction Interpretation (VSCI) has interpretive and philosophical *advantages* over other readings. In this essay, I respond to their comments. Their comments concern aspects of my account of the application of the FUL (Sect. 2 below); its relation to other interpretations of the FUL, in particular the 'Practical Contradiction' and 'Logical Contradiction' interpretations, as well as its difference from the Golden Rule (Sect. 3); and the relation between the FUL and the Formula of the Law of Nature, as well as the role of volitional self-contradictions in moral judgment (Sect. 4).

In one respect, my response goes well beyond a defense and clarification of the view I presented in 2017. In Sect. 2 below, I expand the interpretive framework. On the basis of Kant's argument leading up to the FUL, and on the basis of his third example in the *Groundwork*, I argue that Kant also construes the internal structure of the volitional self-contradiction in another way. Again, one of the sides of the contradiction is 'willing that the tested maxim become a universal law', but now the other side is taken up not by willing the maxim but by willing a necessary object of willing as such. As I explain in Sect. 2.3, Kant assumes that humans, *qua* finite rational beings, necessarily will the means to their actual and possible future ends. On that basis, he specifies the crucial question to ask when testing any maxim as follows:

MORAL QUESTION (added in the expanded VSCI):

Can you (1) will the means to your actual and possible future ends *and simultaneously* (2) will that this maxim become a universal law, without these two volitions contradicting each other?

In this specification of the FUL test, the relevant contradiction is again a contradiction between two (hypothetical) simultaneous volitions. Although its internal structure is different, it is a volitional self-contradiction in the full sense of the term, as developed in my 2017 essay. Moreover, as I clarify below, the two ways in which Kant specifies the test are intimately related, since he explains one in terms of the other.

2 How to Apply the FUL and Diagnose Volitional Self-Contradictions

The FUL articulates a single moral test that all your maxims ought to pass: you ought to act only on maxims that you can simultaneously will to become universal laws. The idea is not that you ought *actually to will* this but that you should be *able* to, without contradiction. The test associated with the FUL is a *possibility* test, to be carried out in moral reflection. It is used to determine the moral status of maxims, that is, of the action principles that serve as the major premises of your practical reasoning. You should make sure that your own maxims pass this test. But you can also apply the test to maxims that you attribute to others, for example in the context of moral blame or praise. And you can test maxims as a philosophical exercise, for instance to illustrate the application of the Categorical Imperative or just to see what would follow from this principle. Last but not least, you can use it to develop a metaphysics of morals, and this was Kant's own ultimate aim.

Maxims can fail this test at two stages. First, if a maxim cannot even be *conceived* as a universal law, then by implication you cannot will to act on it and simultaneously will that it become a universal law; your maxim fails the test. Consider Kant's example of the maxim of borrowing money and falsely promising to repay it, when one believes oneself to be in need (GMS 4:422). If, as Kant argues, this maxim cannot even be conceived as a universal law, then it is also impossible to will that it become one, and hence you cannot will to act on it and simultaneously will that it become a universal law.

Second, if a maxim can be conceived as a universal law, then you can go on to ask whether you can will to act on it yourself and simultaneously will that it become a universal law. If it turns out that these two (hypothetical) volitions would contradict each other, then the maxim fails the requirement at the second stage. Kant provides the example of a man who is indifferent to the needs of others and refuses to help them, even though he wishes to receive help from others when he needs it. By 'wish' Kant here means a disposition to will something under certain circumstances, circumstances that do not presently obtain.⁴ In the essay I show that it would be contradictory (1) to will to act on a maxim of not helping others while wishing to receive help from others when in need and (2) simultaneously to will that this maxim become a universal law.

The VSCI uses an ordinary sense of 'contradiction', and it accommodates Kant's distinction between perfect and imperfect duties by matching it with the two stages at which maxims can fail the test. Moreover, the interpretation fits Kant's own repeated assertion that all impermissible maxims, when submitted to the test, 'contradict themselves' and lead to a 'self-contradiction of the will'. For example, concluding his presentation of the principle of morality, Kant writes:

⁴ Kant here uses 'wish' not in the sense of a desire for something that is impossible or unlikely to happen (as in 'I wish I could fly', 'I wish I could undo the past') but rather in the sense of an 'inactive desire' (V-Met-L1/Pölitz, 28:253), which I here phrase in dispositional terms. Kant's use of 'wish' in the example is fitting, since the man in the example is not presently in need (Kleingeld, 2017: 108).

That will is absolutely good [...] whose maxim, if made a universal law, can never conflict with itself. This principle is, accordingly, also its supreme law: act always on that maxim whose universality as a law you can will simultaneously; this is the only condition under which a will can never be in conflict with itself, and such an imperative is categorical. (GMS 4:437)

Similarly, Kant claims that ‘*in every case*’ of a violation of duty, if we submitted our maxim to the test, ‘we would encounter a contradiction in our own will’ (GMS 4:424, emphasis added).

In response to this interpretive proposal, some of the critics insist that, in the case of maxims that fail the test at the first stage, the *salient* contradiction is the contradiction *in the conception* of the maxim as a universal law and that the volitional self-contradiction that I reconstruct is therefore *superfluous*. In Sect. 2.1, I explain this objection in more detail and provide a reply.

Regarding maxims that fail at the second stage, the critics do not so much challenge my reconstruction of the volitional self-contradiction as argue that it depends on my rendering of the relevant maxim in Kant’s egoism example. They argue that the maxim could or should be construed differently and doubt whether the VSCI could provide a satisfactory account in that case. Relatedly, they wonder whether the VSCI can accommodate the idea that the will has ‘essential purposes’. I present this set of objections and questions in Sect. 2.2, and in Sect. 2.3 I respond.

2.1 Contradictions in Conception and Volitional Self-Contradictions

I argue that the FUL presents a single moral test, but in the literature the Formula of Universal Law is often taken to require that maxims pass two separate tests: the ‘contradiction in conception test’ and the ‘contradiction in the will test’. This assumption is based on the following passage:

Some actions are such that their maxim cannot even be *conceived* as a universal law of nature without contradiction, let alone that one can also/still [*noch*] will that it *should* become one. In the case of other actions, however, that inner impossibility is not found, but it is still impossible to *will* that their maxim be raised to the universality of a law of nature, because such a will would contradict itself. (GMS 4:424)

Here, Kant distinguishes between two sets of impermissible maxims: those that cannot be conceived as universal laws without contradiction (and that, *a fortiori*, cannot be willed to become such), and those that can be conceived but not willed as universal laws without volitional self-contradiction.

The quoted passage raises the question of how the ‘contradiction in conception’ relates to the volitional self-contradiction that I call the ‘relevant’ contradiction. It seems to several critics that the *really* ‘relevant’ contradiction is the ‘contradiction in conception’ and that the problem Kant diagnoses for the first set of maxims is the inconceivability or ‘inner impossibility’ of their becoming universal laws, rather than the fact that they cannot simultaneously be willed to become such (Walschots, 2023,

Lo Re, 2023). Relatedly, others object that invoking a volitional self-contradiction here is *superfluous* (Timmons, 2023, following Kahn, 2019). After all, showing that a maxim fails the ‘contradiction in conception’ test suffices for establishing its impermissibility.

In response, I would like to insist that the *relevant* problem, in these cases, is not that your maxim fails the contradiction in conception test but that your maxim fails to satisfy the demand of the FUL. Relatedly, the demand of the FUL is *not superfluous*. Let me explain this by starting with Kant’s loan/false promising example.

In the example, the maxim is that of borrowing money when in need while promising falsely to repay it (GMS 4:422). Kant claims that it is impossible to conceive of this maxim as a universal law without contradiction. In a world in which the maxim were (and were known to be) a universal law, it would be impossible to misappropriate money through false promises, since potential lenders would not believe the alleged promises of those in need of money. Certainly, there is a ‘contradiction in conception’ here, but this is what *makes* it impossible to will to act on the maxim while simultaneously willing that it become a universal law. After all, willing this would mean willing to act on the maxim yourself while simultaneously willing a world in which it is impossible to act on the maxim. This combination of volitions would constitute a volitional self-contradiction, and this shows that the maxim fails the test.

Let me use an analogy to clarify why the *relevant* contradiction here is the volitional self-contradiction and not the contradiction in conceiving of the maxim as a universal law. In order to get your driver’s license, you need to pass a driving test. Suppose you take the test not even knowing how to start the engine, use the gas pedal, or handle the steering wheel. This on its own establishes that you fail the test, without your driving at all. But this does not mean that the driving test for aspiring drivers is superfluous. In fact, you just took and *failed this very driving test*, not a separate start-the-engine test (or use-the-gaspedal test, etcetera). Similarly, if a maxim cannot even be conceived as a universal law, it will fail the FUL test without your needing to examine whether you can simultaneously will it to become a universal law. But this does not mean that the latter requirement—stated in the FUL—is superfluous. In fact, your maxim fails to meet *this very requirement*, not a separate conceivability-as-universal-law test.

That this is Kant’s view is clear from the passage quoted above: the problem with the first set of maxims, he writes, is not simply that they cannot be conceived as universal laws without contradiction but that, *a fortiori*, they fail the criterion mentioned in the FUL (‘let alone...’). Taking the ‘contradiction in conception test’ to be one of two *separate* tests tends to create the false impression that the salient problem with these maxims is that they cannot be *conceived* as universal laws. The salient point is rather that they fail the criterion stated in the FUL.

If the salient problem with maxims in the first set is that they fail the test articulated in the FUL, and if Kant says that in *every* case of a violation of duty a ‘contradiction in our own will’ emerges during the test (GMS 4:424), then it should be possible to specify just *how* these maxims generate a volitional self-contradiction. This is why I spent considerable effort in my 2017 essay on showing *how* the maxim in Kant’s loan/false promise example generates a volitional self-contradiction when

submitted to the test. This reconstruction is not required for establishing *that* these maxims fail the test. We can see that they fail as soon as we know that they cannot become universal laws, just as the examiner sees that you fail the driving test as soon as it is clear that you know nothing about handling a car. But just as you fail the driving exam (not merely its start-the-engine part), the loan/false promising maxim fails the FUL test (not merely its conceivability-as-universal-law part).

2.2 Kant's Third and Fourth Examples: The Problem

In the 2017 essay, I used Kant's fourth example to illustrate how maxims that *can* be conceived as universal laws may fail to meet the requirement articulated in the Formula of Universal Law at a second stage of the test. In this example, Kant introduces a man in comfortable circumstances who wants to pursue his own ends while never helping others, even though he recognizes that people around him are in great difficulty. He further mentions that this man wishes to receive help from others when he needs it (GMS 4:423). Kant does not formulate the man's maxim explicitly, however. In the essay, I take the various elements together, which yields the following long but familiar maxim of ordinary egoists: 'I shall never help others in need of assistance, but I shall want others to help me when I need this to achieve my own ends' (Kleingeld, 2017:107). I will call this the 'long' version of the maxim. Kant asserts that a world populated by people acting on the maxim is not inconceivable,⁵ but the maxim yields a straightforward volitional self-contradiction when tested. In the world in which this maxim is a universal law, the man does not receive any help when he needs it; as per his maxim, however, he wishes to receive help when he needs it. Therefore, he cannot will to act on the maxim himself and simultaneously will that it become a universal law.

Walschots (2023), Timmons (2023), Romero (2023), and Lo Re (2023) do not contest that it is textually possible to construct the maxim as I do; nor do they challenge my claim that this would yield a volitional self-contradiction. They find it textually more plausible, however, to split Kant's description of the agent's attitudes into two parts. Only the first part would refer to the maxim: 'I shall never help others in need of assistance'. I will call this the 'short' version of the maxim. The second part—Kant's reference to the agent's wish to receive help when he needs it—would refer to an 'essential purpose' of the will, namely happiness.

The commentators point to Kant's statement that happiness is a necessary end of human beings (GMS 4:415). Moreover, they refer to Kant's explanation, in the *third* example, of why it is impermissible to adopt a maxim of not developing your useful natural talents (GMS 4:423). There Kant argues that this maxim fails the test because 'as a rational being [the man in the example] necessarily wills that all his capacities be developed, since they serve and are given to him for all sorts of possible purposes' (GMS 4:423). This raises the question whether the VSCI fits Kant's explanation of the third example and whether it can account for the structural similarity that Kant claims exists between the third and fourth examples.

⁵ Kant seems to conceive of the world of the universalized maxim as containing adults only; that is, he abstracts from the fact that human offspring does not survive without help.

The issue of how to construe the relevant maxim in the fourth example relates to a difficulty I noted in the 2017 essay, namely that it is not immediately clear how the short maxim of never helping yields a volitional self-contradiction when submitted to the FUL test (Kleingeld, 2017: 108, n.45). Yet Kant clearly deems the short version of the maxim to be impermissible (GMS 4:430n.). Hence the VSCI would be strengthened if it can show how, on Kant's view, the short maxim fails the test.

2.3 The Solution: Expanding the VSCI

I appreciate the textual point that we should expect Kant's explanation of the fourth example to match his explanation of the third. The VSCI will indeed gain 'interpretive force' (Walschots, 2023) if it can analyze his third and fourth examples in parallel fashion and solve the problem concerning the short maxim of never helping. Before I show how the challenge can be met, important clarifications are in order about Kant's argument leading up to the FUL. The argument immediately preceding the FUL turns out to be crucial to understanding the emergence of the volitional self-contradiction during the test.

Kant writes in the *Groundwork* that the will (*Wille*) is the capacity to act on the basis of principles (GMS 4:412). Thus, the will is the causal power of a rational agent, and as such it is subject to norms of rationality.⁶ Willing an end is being committed, on the basis of practical reasoning, to pursue, do or bring about something.

Kant claims that in the case of *finite* rational beings, willing an end entails willing the *means* to its realization (to the extent that those means are in one's power, GMS 4:414, 417). His contrast case is God, for whom willing something *immediately* establishes its actuality. Kant's claim concerns what willing entails for finite rational beings *qua* rational. Humans do not necessarily act rationally, however, so the claim should not be misunderstood as meaning that human beings always do in fact take the available means to their ends; they may be akratic, for instance.⁷ Nor, second, should the claim be misunderstood as referring to selfishness. It holds equally for all ends, regardless of whether these are altruistic or selfish, moral or immoral. Note, third, that willing the available means to one's ends is not properly conceived as a separate further end (let alone a separate 'essential purpose') of the will. Rather, it is already *entailed* in willing any end as such (insofar as the willing subject is rational).

Kant further claims that humans necessarily have happiness as their overarching end (GMS 4:415). He uses a formal notion of happiness as the condition of a rational being whose ends have been realized in the world (KpV 5:124). Thus, this condition includes the realization of an agent's moral ends, in addition to their contingent and merely subjective ends.⁸ Importantly, Kant argues, humans cannot predict with certainty what their future ends will be (GMS 4:418). As a result, they also cannot

⁶ This is rightly emphasized by Romero (2023) and elaborated by Walschots (2023).

⁷ In the passages cited here, Kant clearly relies on a conception of rationality. Thus, I would give a positive answer to Romero's question whether I would accept the idea that the FUL presupposes a conception of rationality. Moreover, Romero is right to point out that the principle of noncontradiction is in play in any event and that Kant sees this as an a priori principle of pure reason (Romero, 2023).

⁸ For further discussion of Kant's conception of happiness, see Kleingeld 2016.

predict which means they will require, for example which specific skills and forms of assistance (if any) they will need for realizing their future ends. But whatever their ends turn out to be, to the extent that reason has decisive influence on their will they necessarily will the means to their ends. Thus, they will the means to their possible future ends *in general*, that is, in the sense that they will the ability to realize their ends (to the extent that it is in their power), whichever ends they may turn out to have. What is meant is not that they will the *particular* means to the *particular* ends they will in fact turn out to adopt, since they cannot predict with certainty what their ends will be.

These claims concerning what finite rational beings *necessarily will*, together with the wording of the third example, suggest a novel construal of the volitional self-contradiction generated in Kant's third and fourth examples. This construal turns not on the desire for happiness as such but on the rational entailment relation between willing an end and willing the means, coupled with the claim that we cannot predict our future ends and the means we will need.

The third example concerns the maxim of neglecting one's natural talents. Kant asserts that a rational being 'necessarily wills' that his capacities be developed, since they serve him as means to all sorts of ends. This, he argues, explains why the man in the example, *qua* rational being, 'cannot possibly will' that this maxim become a universal law (GMS 4:423). Kant here suggests that (1) willing that this maxim become a universal law would *contradict* (2) willing what one necessarily wills *qua* finite rational being (*viz.*, the means to one's actual and possible future ends, including useful skills and capacities). Willing both simultaneously would constitute a volitional self-contradiction.

In the case of the fourth example, we get a structurally similar analysis if we take the maxim to be the short maxim of never helping, and if we take Kant's reference to the man's wish to receive help (should he need it) as referring to what he necessarily wills *qua* finite rational being. Then (1) willing that the maxim of never helping become a universal law would *contradict* (2) willing what one necessarily wills *qua* finite rational being (*viz.*, the means to one's actual and possible future ends, including help when one needs it). Thus, willing both simultaneously would indeed cause the will to 'contradict itself' (GMS 4:423), that is, it would be a volitional self-contradiction.

This analysis shows that we can proceed in either of two ways when formulating the volitional self-contradiction at issue in the third and fourth examples. If we make the maxim *long*, then it is easy to see how it generates a volitional self-contradiction. It is contradictory for the ordinary egoist (1) to will his long egoistic maxim as his own action principle and simultaneously (2) to will that his maxim be a universal law (Kleingeld, 2017, 107–108). After all, he does wish to receive help if ever he needs it, but in a world in which his maxim was a universal law, he would not receive help. If we keep the maxim *short*, then the general rational commitment to willing the means to one's ends occupies one side of a volitional self-contradiction, the other side of which is (hypothetically) willing that the maxim become a universal law. Whether we make the maxim long or short, it generates a volitional self-contradiction, although the internal structure of this contradiction will differ accordingly.

It seemed plausible to me, when writing the 2017 essay, to construe the maxim in Kant's fourth example in the long way, since this is a familiar maxim adopted by ordinary egoists. As a result, however, my original analysis left unaddressed the issue of how the short maxim could fail the test. In light of Kant's account of what finite rational beings 'necessarily will', it is now clear that his reference to the man's wish to receive help was not a contingent aspect of his choice of a particular example. Rather, it refers to willing the means to one's ends, which is something finite rational beings will necessarily. Moreover, it is the element that is essential to generating the volitional self-contradiction.

A revealing passage in the *Metaphysics of Morals* nicely clarifies how, on Kant's view, the short maxim of never helping yields a conflict between willing the maxim and willing the maxim to become a universal law. This is the passage in which Kant explains why the ('self-interested') maxim of never helping others fails the test. Here he first introduces the claim that all human beings wish to be helped when they are in need. He then describes the maxim at issue as the maxim 'not to help others in need'—that is, the passage concerns the short maxim (MS 6:453). He continues by arguing that if this maxim became a universal law, one would not receive help when in need. He then concludes that this shows that the *maxim* conflicts with the *maxim* as a universal law: 'the self-interested maxim conflicts with itself if it were made a universal law, that is, it is contrary to duty' (MS 6: 453).

What makes this passage particularly interesting is that Kant here clearly indicates that he sees the short maxim of never helping as leading to a conflict between the maxim and the maxim as a universal law, and that this conflict is *premised* on the claim that everyone (*qua* finite rational being) wishes to receive help when in need. *Without* the presupposition that everyone wishes to receive help when in need, it would remain unclear how the maxim could conflict with the maxim as a universal law. This was the problem that we started with, a problem diagnosed but left unaddressed in my 2017 essay. Kant's discussion in this passage from the *Metaphysics of Morals* now clarifies that this conflict *presupposes* that everyone (to the extent that one is rational) wishes to receive help when in need, since this is the premise with which he starts. *Given* the wish to be helped when in need, and considering that no one receives help in a world in which the maxim is a universal law, the 'maxim conflicts with itself as universal law', and one cannot will this maxim and simultaneously will that the maxim become a universal law.⁹

Thus, this passage also explains how the two ways of construing the volitional self-contradiction relate to each other. In the case of the *long* egoistic maxim, the contradiction between (hypothetically) willing the maxim and willing that it become a universal law is generated by the inclusion, in the maxim, of (rationally necessarily) willing the means to one's ends. In the case of the *short* egoistic maxim, this element is not included in the maxim, and in that case willing the means to one's ends con-

⁹ What about the maxim to *neither give nor accept* any help? Kant's account as reconstructed here suggests that the result of the test would be similar. The maxim would not be 'self-interested', however, but irrational, since a person adopting this maxim would set various ends while rejecting the necessary means (in cases where realizing these ends would require the help of others). To the extent that the person sets 'not giving or accepting help' as their highest end, overriding their moral duties to themselves and others, this would likely qualify as a 'passion' in Kant's sense, that is, as a form of evil (cf. Anth 7:265–267).

tradicts willing that the maxim become a universal law. Either way, the same three elements are in play in conducting the test: the idea that finite rational beings as such necessarily will the means to their actual and possible future ends, the maxim (long or short, depending on whether the former idea is included in the maxim or not), and the idea of the world in which the maxim is a universal law. Regardless of whether one takes the short or the long version of the maxim, the moral verdict when one conducts the test will be the same, since either way a volitional self-contradiction will result.

I believe this expanded account strengthens the VSCI considerably, both by showing how it applies to a wider set of maxims and by showing how the third and fourth examples can be read in a parallel fashion. Moreover, the expanded account further reinforces the VSCI's main thesis: the diagnostic contradiction at issue in the FUL test is a contradiction between two (hypothetical) volitions that are simultaneously entertained in moral reflection.

3 Practical Contradictions, Logical Contradictions, and the Golden Rule

3.1 The Practical Contradiction Interpretation

Walschots (2023) and Lo Re (2023) request more clarification on the relation between the VSCI and the Practical Contradiction interpretation (PCI) developed by Christine Korsgaard. Before I clarify this relation, let me first explain how I understand the PCI.

In her seminal essay 'Kant's Formula of Universal Law', Korsgaard argues that the contradiction that emerges in the case of impermissible maxims is best understood in 'a specifically practical sense of "contradiction"' (Korsgaard, 1996: 93). Willing that such a maxim become a universal law would equal willing the 'frustration' or the 'thwarting' of your end. Korsgaard argues that a volition with this internal structure is genuinely contradictory in a practical sense (93–94). She explains that it is a contradiction *within* a single volition rather than a contradiction *between* willing and not willing something, and she explicitly denies that a practical contradiction is best captured by saying that someone 'is acting as if she both did and didn't will the end' (93). Further, she points out that Kant claims that, rationally speaking, if you will an end, you also will the indispensably necessary means to it that lie in your power (GMS 4:417). Because Kant sees this as analytic, that is, as a matter of conceptual entailment, she argues, willing the frustration of your end is appropriately called a (practical) contradiction (Korsgaard, 1996: 93–94).

Korsgaard contends that practical contradictions emerge both with maxims that fail the 'contradiction in conception test' and with those that fail the 'contradiction in the will test'. In both sets of cases, willing that the maxim become a universal law equals willing the thwarting of your purpose. To show how these practical contradictions emerge, she introduces two further assumptions: she stipulates that every maxim must include a purpose and that the will has several 'essential purposes' (96–97).

The first assumption makes it possible to explain the practical contradiction involved in the first set of cases—maxims that fail the ‘contradiction in conception’ test—as follows:

the contradiction that is involved in the universalization of an immoral maxim is that the agent would be unable to act on the maxim in a world in which it were universalized so as to achieve his own purpose—that is, the purpose that is specified in the maxim. (92)

If your maxim includes a purpose that is impossible to achieve in the world in which your maxim is a universal law, then willing that it become a universal law comes down to thwarting your own purpose: a practical contradiction.

Korsgaard argues that in the second set of cases—maxims that fail the ‘contradiction in the will’ test—we ‘must find’ some essential purpose of the will in order to generate a practical contradiction. She writes:

If a thwarted purpose is a practical contradiction, we must understand the contradiction in the will test this way: we must find some purpose or purposes which belong essentially to the will, and in the world where maxims that fail these tests are universal law, these essential purposes will be thwarted, because the means of achieving them will be unavailable. (96)

As an example of such an essential purpose she mentions the will’s ‘general effectiveness in the pursuit of its ends’ (Korsgaard, 1996: 96–97). Her introduction of ‘essential purposes’ makes it possible again to locate a practical contradiction *within* willing that certain maxims become universal laws, namely when this would mean thwarting an essential purpose of your will. In my 2017 essay, I called the introduction of essential purposes an ‘auxiliary assumption’ on Korsgaard’s part, because she introduces them specifically in order to generate a practical contradiction and not as an idea found in Kant’s texts.

Korsgaard admits that the PCI has the unfortunate implication that there is a set of maxims that do not lead to practical contradictions when submitted to the test even though we would expect them to be impermissible on a Kantian account, and even though they are indeed rejected by the Formula of Humanity. These are maxims of ‘natural violence’, that is, maxims of violent actions such as killing, the efficacy of which does not depend on a social convention that disappears when the maxim becomes a universal law (as with lying promises). For example, willing that the maxim of killing out of hatred become a universal law does not involve willing the thwarting of your purpose, and hence on the PCI the maxim does not seem to fail the FUL test (Korsgaard, 1996: 100).

This problem is significant, and it applies to a wider set of maxims than those mentioned by Korsgaard. Consider the consequentialist maxim of using a smaller number of people, regardless of their consent, to save a greater number when the opportunity arises. Here universalization does not merely fail to thwart your purpose but positively *serves* it. If this maxim were a universal law, more lives would be saved. In the world in which it is a universal law, agents would push a heavy man in front of a

runaway trolley if this would save the lives of five others (other things being equal). Transplant surgeons would discreetly kill suitable passers-by to obtain the organs necessary to save the lives of many more patients. In this world, the agent's purpose of saving more people by using a smaller number of others would be realized (assuming correct calculations of course). No practical contradiction arises in willing that this consequentialist maxim become a universal law, so on the PCI it would pass the FUL test. Yet if acting on this maxim were permitted by the Formula of Universal Law, different formulas of the Categorical Imperative would yield contradictory results, since the maxim would clearly be prohibited by the Formula of Humanity.

It might seem as if this problem could be solved, on Korsgaard's behalf, by appealing to her notion of 'essential purposes' of the will.¹⁰ She introduces this notion specifically to produce a practical contradiction in Kant's third and fourth examples, suggesting that the 'general effectiveness in the pursuit of its ends' and the 'freedom to adopt and pursue new ends' might be thought to be essential purposes of the will (Korsgaard, 1996: 96). She does not develop this point in detail, but even if we grant it for the sake of argument no genuine practical contradiction would emerge in the case of the consequentialist maxim. For a practical contradiction to emerge, the means to achieving the will's essential purposes would have to 'be unavailable' in the world of the universalized maxim (Korsgaard, 1996: 96, quoted above). Yet it does not follow that in a world in which (by hypothesis) more persons have longer and healthier lives these means would indeed be unavailable.

Furthermore, given that the PCI requires reference to purposes in order to establish the thwarting relation characteristic of a practical contradiction, Korsgaard's account requires the stipulation that maxims must include purposes. Kant's own examples of maxims often do not include a purpose, however (e.g., the maxim 'not to help others in need', MS 6:453, or to 'endure no insult unavenged', KpV 5:19).

How does the VSCI compare to the PCI? Whereas Korsgaard defends her PCI primarily on the basis of philosophical considerations (Korsgaard, 1996: 80), my aim with the VSCI is both exegetical and philosophical. As for its exegetical advantages, the VSCI uses an ordinary sense of 'contradiction', as mentioned in the preceding sections, rather than Korsgaard's novel notion of a practical contradiction. This seems to be an advantage of the VSCI, since Kant also does not introduce a new sense of 'contradiction' in his discussion of the FUL.¹¹ Relatedly, the VSCI does not require the stipulation that maxims must include purposes. I consider this a further advantage of the VSCI as a reading of Kant's texts, since Kant also does not stipulate this.

An important philosophical advantage of the VSCI is that it does not encounter problems with maxims of natural violence. Maxims of killing those who disturb one's sleep, killing others out of hatred (examples mentioned in Korsgaard 1996: 82, 100), or killing unsuspecting passers-by to save the lives of larger numbers of others (on the consequentialist maxim mentioned above) yield volitional self-contradictions when submitted to the test. A world in which such a maxim is a universal law is con-

¹⁰ I thank Guus Duindam for helpful discussion of this issue.

¹¹ I do not mean to deny that willing the thwarting of your purpose is a form of practical irrationality. But I believe it is unnecessary to introduce Korsgaard's new sense of 'contradiction' to explain the application of the FUL.

ceivable. But you cannot coherently will to act on a maxim of natural violence and simultaneously will that this maxim become a universal law. Consider Kant's account of what a finite rational being necessarily wills, and assume a maxim of natural violence (say, killing out of hatred, or killing some for the sake of saving a greater number of others). Willing such a maxim to become a universal law would mean willing that others end your life, annihilating your will, under the conditions specified in the maxim. When subjected to the FUL test, such a maxim would lead to the following result: you would (1) will to realize the actual and possible future ends of your own will and obtain the means required for this, and you would simultaneously (2) will a world in which others annihilate your will under the conditions specified in the tested maxim (and not conditional on your consent). This would constitute a volitional self-contradiction, and hence the maxim of natural violence fails the FUL test.¹²

The differences between the VSCI and the PCI, then, remain substantial. Korsgaard locates the relevant contradiction *within* hypothetically willing the maxim as a universal law, she introduces a new sense of 'contradiction' that requires further stipulations concerning maxims, and she acknowledges that the PCI has difficulties handling maxims of natural violence. The VSCI, by contrast, locates the relevant contradiction *between* two hypothetical volitions, it uses the ordinary notion of contradiction, it does not require stipulating that maxims must include purposes, and it does not let maxims of natural violence pass the test. I like to think that this gives the VSCI the exegetical advantage over the PCI, and probably the 'philosophical' advantage as well. Fully establishing the latter, however, would of course require more extensive discussion than I can include here.

3.2 The Logical Contradiction Interpretation

Like the PCI, the LCI locates the relevant contradiction within a single volition, viz., within the (hypothetical) willing of the maxim as a universal law. In contrast to the PCI, however, the LCI conceives of it as a contradiction in an ordinary sense. The VSCI agrees with the LCI in its use of the ordinary notion of contradiction but differs from the LCI in that it does not see the relevant contradiction as obtaining within willing the maxim as a universal law. Furthermore, the VSCI does not require an artificial thought experiment that requires at least one important step beyond what is found in Kant's explanations of the examples.

Mark Timmons defends a sophisticated version of the LCI that he calls the 'Logical/Causal-Law' interpretation. He locates the relevant contradiction *within* willing the world of the universalized maxim, by virtue of there being a contradiction '*in the hypothetical world* being contemplated' (Timmons, 2023, original emphasis). As for the first set of maxims that fail the FUL test, Timmons relies on the notion of a system of causal laws for showing how this contradiction is generated: it is a 'logical contradiction that results from trying to conceive of one's maxim as a universal law of nature' (Timmons, 2023). If it is impossible to act on the maxim when it is a law of

¹² It is sometimes assumed that if maxims of killing fail the FUL test only at the second stage, this would be an inappropriate result insofar as killing would 'merely' be a violation of an imperfect duty and would hence be permitted *sometimes*. I argue against this assumption in Kleingeld 2019.

nature, then in willing that the maxim become a universal law one would be willing (or attempting to will) two mutually contradictory states of affairs in one volition: a world in which people act on the maxim and a world in which people do not. Thus, the LCI locates the contradiction *within* a single (attempted) volition, namely within the (attempted) volition that the maxim become a universal law of nature.

In the case of the second set of maxims that fail the test, Timmons again sees the relevant contradiction as obtaining in the idea of the world in which the tested maxim is a universal law, but constructing this idea now requires a further step. He argues that ‘one is to import into that hypothetical world the information about the commitments of rational agency’ (Timmons, 2023). If a maxim’s becoming a universal law is inconsistent with the rational commitments of agents in that world (as, for example, in the case of the maxim of never helping), then the idea of this hypothetical world contains a contradiction. As a result, willing that such a maxim become a universal law contains a contradiction. Again, the logical contradiction resides ‘*in the hypothetical world being contemplated*’ (Timmons, 2023).

Elsewhere, I have argued against the claim that the relevant contradiction is to be sought ‘in the hypothetical world’, as Timmons puts it (Kleingeld, 2021, 614–615). Furthermore, the construction of the thought experiment includes a step that is not found in Kant’s examples. On the VSCI, the commitments of rational agency do not have to be ‘import[ed] into the hypothetical world’ in a separate step but are simply left where they are to begin with: on the side of the agent who engages in moral reflection. The contradiction, if one emerges, is found not within the idea of the hypothetical world. Rather, it is a contradiction between the reflecting agent’s hypothetically (1) willing a certain maxim and (2) willing that it should become a universal law, or between (1) willing the means to one’s actual and possible future ends and (2) willing that a certain maxim become a universal law (see above, Sect. 2.3).

3.3 The Golden Rule

The Golden Rule expresses the criterion of good action in terms of the subjective preferences of the person who applies it. In the negative formulation used by Kant, this criterion is: Do not do to others what you do not want them to do to you. The Golden Rule targets actions, not maxims, and the results of applying it often vary from person to person. Depending on the context, a hedonist, a masochist, a socialite, a misanthrope (and so on) could reach different results. Thus, the Golden Rule does not yield objective moral standards. Furthermore, it yields no duties to oneself, and as for duties to others, it yields only duties of omission.

In all these respects the Golden Rule differs from the Categorical Imperative,¹³ and this is no different on the VSCI. According to the latter, the criterion is not subjective preference but the possibility of willing two things simultaneously without contradiction. It is a test of maxims, not actions. It is to give the same outcome regardless of who performs the test. It is to identify duties to oneself and to others, and duties of omission and of commission.

¹³ See also Kant’s comment at GMS 4:430n.

Nevertheless, Kant's handling of the egoism example has sometimes been read as a version of the Golden Rule argument. Timmons suggests that my analysis of the egoism case makes it look like a 'reversibility test characteristic of the negative formulation of the Golden Rule' (Timmons, 2023). Timmons also argues that on my account the egoist maxim would be rejected because 'it would be *imprudent* to will a world entailing that no one helps you, yet you will to be helped' (Timmons, 2023). This is not how I present (or intend) the case, however. My analysis proceeds in terms of the agent's maxim, not their action (as the Golden Rule would suggest), and I do not explain this maxim's failing the FUL test in terms of imprudence or dispreference. Instead, I argue that it would be *contradictory* (1) to will to act on a maxim of never helping others while wishing to receive help when you need it and, simultaneously, (2) to will a world in which no one ever helps.

Thus, on the VSCI, the reason why the maxim fails the FUL test is neither that it would be imprudent to will a world without help nor that the egoist prefers to receive help. To be sure, it may indeed *also* be imprudent, and the actions of Kant's egoist may *also* fail the Golden Rule test. But what matters for the purposes of moral evaluation is that the maxim yields a volitional self-contradiction when submitted to the FUL test.

4 Two Further Replies to Timmons

4.1 The Formula of Universal Law and the Formula of the Law of Nature

My focus, both in the 2017 essay and in the present discussion, has been on the FUL throughout. This is the most prominent statement of the Categorical Imperative in the *Groundwork*. Yet Kant's four examples are introduced to illustrate the Formula of a Law of Nature (FLN). How should we conceive of the *relation* between the FUL and the FLN? I did not address this issue in detail in my 2017 essay, and Timmons is right to raise it. For one thing, the first statement of the FLN is one of the very few formulations of the Categorical Imperative that does *not* include 'simultaneously' or 'also'. Furthermore, as Timmons notes, Kant attributes a unique status to the FLN in the 'Typic' of the *Critique of Practical Reason*.

As for the omission of 'simultaneously' from the FLN, Kant's second statement of the FLN does include it ('*simultaneously* as universal laws of nature', GMS 4:437). Thus, there is no reason to assume that Kant considered the simultaneity condition inapplicable to this formula. In the 2017 essay, I offer a possible explanation for its omission from the initial statement (2017:110, n.48).

As for the relation between the two formulas, in the *Groundwork* Kant introduces the FUL as the most general formulation of the Categorical Imperative. The conception of law in the FUL remains abstract: Kant does not specify a particular *kind* of 'universal law'. He then proceeds to offer several reformulations, using what he calls 'analogies' with laws of nature (FLN) and with political legislation (Formula of Autonomy and the Formula of a Realm of Ends). As Kant explains, each of these reformulations serves to 'represent' the abstract FUL in more concrete terms, that is, in terms of laws with which we are familiar (GMS 4:436–437). He explains that

the reformulations are ‘useful’ for *introducing* the principle of morality. He adds, however, that it is ‘better’ to use the ‘general formula’ of the Categorical Imperative for purposes of moral *evaluation* (*in der sittlichen Beurteilung*) (GMS 4:437), and he then offers another general formulation that—like the FUL—refers abstractly to ‘universal law’ (GMS 4:436–37). Because Kant presents it as the principal formulation of the supreme moral principle in the *Groundwork*, my focus is on the FUL.

In the *Critique of Practical Reason*, the principal statement of the supreme moral principle again includes an abstract notion of universal law and a simultaneity condition. Kant formulates the ‘fundamental law of pure practical reason’ as follows: ‘So act that the maxim of your will can always simultaneously hold as the principle of a universal legislation’ (KpV 5:30). In contrast to the *Groundwork*, however, Kant does not immediately proceed to offer a series of reformulations using analogies from the realms of nature and politics. Instead, he adds a separate section later on, entitled the ‘Type of pure practical judgment’. Here, he designates a ‘law of nature’ as the formal ‘type’ of a moral law (KpV 5:69–70). He explains this by saying that, for the purposes of moral judgment, you should ask whether ‘your will’ would ‘agree with’ (*einstimmen*) living in a world in which ‘everyone’ acted on your maxim. He also refers to this procedure as a ‘comparison’ (*Vergleichung*) between your maxim and a universal law of nature (KpV 5:69). These locutions are clearly consistent with the idea, central to the VSCI, that we should examine the simultaneous compossibility¹⁴ of two volitions rather than merely the possibility of willing a maxim as a universal law.

In sum, the principal statements of the Categorical Imperative in the *Groundwork* (FUL) and the *Critique of Practical Reason* (the ‘fundamental law’) contain an abstract notion of universal law. In both works, the law of nature serves to ‘represent’ this abstract notion in more concrete terms. Because the main statement of the supreme moral principle is the abstract formulation, this is what I have focused on. Finally, and most importantly, for determining the nature and location of the contradiction that emerges in the case of impermissible maxims, it makes no difference whether we consider the abstract general principle or its concrete type (in terms of universal laws *of nature*). After all, insofar as the FLN is a more concrete representation of the FUL, the nature of the relevant contradiction must be the same.

4.2 Moral Judgment, Inconsistency, and Conscience

As mentioned at the beginning of Sect. 2, whether you test actual or counterfactual maxims, your own maxims or those of others, for practical or theoretical purposes, the test as such is and remains a possibility test. The question is whether you *can* simultaneously will that the maxim become a universal law without this leading to a volitional self-contradiction. The contradiction emerges, if it does emerge, in moral *reflection* when you submit a maxim to the test. This raises the question of how this counterfactual contradiction, which is diagnosed in reflection, relates to the actual

¹⁴ For an explanation of why this is not a pleonasm, see Kleingeld 2017: 99–100, n. 32.

inner tensions in agents who are aware that they are acting on morally impermissible maxims.¹⁵

Timmons, himself a defender of a version of the LCI, rejects the VSCI but nevertheless claims there is a good use to which the idea of a volitional self-contradiction can be put. He suggests that its ‘true relevance’ pertains not to *moral* judgment and evaluation but to practical deliberation more generally. The idea of a self-contradiction of the will serves to articulate a general coherence constraint on one’s volitions. This coherence constraint ‘only rules out a *combination* of attitudes’ (Timmons, 2023), and according to Timmons it puts pressure on agents to ‘either elevate moral reasons over reasons of self-love or vice versa’ (Timmons, 2023). Thus, it puts pressure ‘on individuals contemplating acting contrary to duty’ (Timmons, 2023; cf. Romero, 2023), and Timmons equates this with the voice of conscience (Timmons, 2023). On his view, this pressure should not be ‘conflated’ with the logical contradiction central to the procedure for reaching *moral* judgments.

Timmons sees the FUL (or rather, in his case, the FLN) as a test to be run through a separate thought experiment, and he sees conscience as prompting agents to *run* this test (to ‘*raise the question* of moral permissibility’, Timmons 2023, emphasis added). Running this test then yields a specifically moral verdict. Thus, he seems to claim that we are *first* alerted by our conscience to the fact that a contemplated course of action violates general coherence constraints, and we *then* turn to the Categorical Imperative to find out whether the course of action is morally permissible.

I see two difficulties with Timmons’s view. First, it is not clear to me what role the notion of a ‘volitional self-contradiction’ plays in the general coherence constraint on attitudes as Timmons conceives of it. Which combination of attitudes is the constraint supposed to rule out as incoherent? The volitional self-contradiction as I presented it consists in a contradiction, emerging counterfactually in moral reflection, between willing a maxim and willing that it become a universal law. This, however, is certainly not the combination of attitudes entertained by agents who contemplate acting immorally. They contemplate adopting *one* maxim *themselves* while willing that *another* maxim be a law for *everyone else* (GMS 4:424)—and these two volitional attitudes are compatible. Immoral agents encounter a volitional self-contradiction only in moral reflection, when they ask themselves whether they can will *the same* principle both as their own maxim and as a universal law (that is, as a principle for everyone, not just for others).¹⁶ They then realize that they cannot. And certainly they do not.¹⁷

Second, Kant describes *conscience* as an inherently *moral* capacity, not as concerned with general deliberative consistency. Conscience is ‘the representation of

¹⁵ Lo Re (2023) wonders whether, on my account, the disharmony in the immoral will constitutes a contradiction, as suggested by Timmermann (2018). I deny this, as I explain below and (in more detail) in my reply to Jens Timmermann (Kleingeld, 2021).

¹⁶ Relatedly, *moral motivation* should not be explained simply by reference to counterfactual volitional self-contradictions, but in terms of *respect for the moral law itself*. In response to comments by Timmons (2023) and Romero (2023), therefore, I would like to insist that, on Kant’s account, ‘moral pressure’ does not derive from confronting a contradiction merely as such, but from recognizing the binding force of the moral law itself.

¹⁷ For a more detailed discussion of this point, see my reply to Timmermann (Kleingeld, 2021).

duty’ (MS 6:438). In the *Groundwork* examples, the man’s conscience prompts him to ask whether his manner of acting is ‘impermissible and against duty’ (GMS 4:422). Conscience is the capacity for moral self-assessment, the ‘inner judge’ (MS 6:438). It is ‘practical reason holding the human being’s duty before him for his acquittal or condemnation in every case that comes under a law’ (MS 6:400).¹⁸ For Kant, bad conscience is not the awareness of inconsistency among your attitudes but an awareness of *transgression*—the awareness that you are (or were or will be) violating a law that you recognize as unconditionally binding.

5 Conclusion

The FUL requires that you act only on maxims that you can ‘simultaneously will’ to become universal laws. Simultaneously with willing *what*? In Kleingeld (2017) I pointed to textual evidence in support of the following answer: simultaneously with willing that the maxim serve as one’s own action principle. In this essay, in response to the thoughtful questions and objections from Walschots, Timmons, Romero, and Lo Re, I have pointed to textual evidence in support of a second answer: simultaneously with willing what you necessarily will *qua* finite rational being, namely the means to realizing your actual and possible future ends. I have also argued that the two answers are intimately related. This new argument constitutes a crucial supplement to the original account, since the expanded VSCI now makes it possible to address the most significant question left open before, namely whether ‘short’ maxims such as the maxim of never helping would generate a volitional self-contradiction when tested. I have addressed several other important questions and comments as well. More work remains to be done, however, if only to spell out further details concerning the application of the FUL and to show whether the VSCI can handle the canonical set of objections that has become a staple of the literature on the FUL .

Acknowledgements I thank Sorin Baiasu for organizing the 2018 KOSAK conference at the University of Keele, where Romero, Timmons and Walschots first presented their comments. I am grateful to Christian de Weerd, Fiorella Tomassini, Guus Duindam, Jan van Houten, Janis Schaab, Leon van Rijsbergen, Lisa Janssen, Marijana Vujošević, Michael Gregory, Sam White, Samuel Kahn, Sandra Kempen, Suzanne Jacobi, Vinicius Pinto de Carvalho and the editors of this special issue (Sorin Baiasu and Micha Gläser) for helpful comments. I am grateful to the Dutch Research Council (NWO) for financial support.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

¹⁸ For further discussion of the role of conscience in Kant’s moral philosophy, see Vujošević 2014.

References

- Kahn, S. (2019). Defending the Traditional Interpretations of Kant's Formula of a Law of Nature. *Theoria* 66(158), 76–102. <https://doi.org/10.3167/th.2019.6615804>.
- Kant, I. (1996). *Practical Philosophy*. Translated and edited by M. J. Gregor. Cambridge University Press. <https://doi.org/10.1017/CBO9780511813306>.
- Kleingeld, P. (2016). Kant on 'Good', the Good, and the Duty to Promote the Highest Good, in T. Höwing (Ed.), *The Highest Good in Kant's Philosophy* (pp. 33–49). De Gruyter. <https://doi.org/10.1515/9783110369007>.
- Kleingeld, P. (2017). Contradiction and Kant's Formula of Universal Law. *Kant-Studien*, 108(1), 89–115. <https://doi.org/10.1515/kant-2017-0006>.
- Kleingeld, P. (2019). A Contradiction of the Right Kind: Convenience Killing and Kant's Formula of Universal Law. *The Philosophical Quarterly*, 69(274), 64–81. <https://doi.org/10.1093/pq/pqy038>.
- Kleingeld, P. (2021). Self-Contradictions of the Will: Reply to Jens Timmermann. *Kant-Studien*, 112(4), 611–622. <https://doi.org/10.1515/kant-2021-0032>.
- Korsgaard, C. M. (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139174503>.
- Lo Re, S. (2023). All Bad Things Come in Threes? On FUL and Its Contradictions – Comments on Walschots's Response to Kleingeld. *Philosophia*. <https://doi.org/10.1007/s11406-023-00639-8>.
- Nyholm, S. (2017). Do we always act on Maxims? *Kantian Review*, 22(2), 233–255. <https://doi.org/10.1017/S1369415417000036>.
- Romero, P. (2023). From Volitional Self-Contradiction to Moral Deliberation: Between Kleingeld and Timmons's Interpretations of Kant's Formula of Universal Law. *Philosophia*. <https://doi.org/10.1007/s11406-023-00644-x>.
- Timmermann, J. (2018). A Tale of Two Conflicts: On Pauline Kleingeld's New Reading of the Formula of Universal Law. *Kant-Studien* 109(4), 581–596. <https://doi.org/10.1515/kant-2018-4003>.
- Timmons, M. (2023). Making Sense of Kant's Formula of Universal Law: On Kleingeld's Volitional Self-Contradiction Interpretation. *Philosophia*. <https://doi.org/10.1007/s11406-023-00615-2>.
- Vujošević, M. (2014). The Judge in the Mirror: Kant on Conscience. *Kantian Review*, 19(3), 449–474. <https://doi.org/10.1017/S1369415414000181>.
- Walschots, M. (2023). The Volitional Self-Contradiction Interpretation of Kant's Formula of Universal Law: A Response to Kleingeld. *Philosophia*. <https://doi.org/10.1007/s11406-023-00621-4>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.