



The Volitional Self-Contradiction Interpretation of Kant's Formula of Universal Law: A Response to Kleingeld

Michael Walschots¹ 

Received: 9 January 2023 / Revised: 9 January 2023 / Accepted: 19 January 2023 /
Published online: 28 January 2023
© The Author(s) 2023

Abstract

In this paper I critically engage with Pauline Kleingeld's 'volitional self-contradiction' interpretation of Kant's formula of universal law. I make three remarks: first, I seek to clarify what it means for a contradiction to be volitional as opposed to logical; second, I suggest that her interpretation might need to be closer to Korsgaard's 'practical contradiction' interpretation than she thinks; and third, I suggest that more work needs to be done to explain how a volitional self-contradiction generates both a 'contradiction in conception' and a 'contradiction in will.'

Keywords Kant · Contradiction · Categorical imperative · Formula of universal law · Happiness · Maxim

1 Introduction

Pauline Kleingeld's 'Contradiction and Kant's Formula of Universal Law' (2017) proposes a new interpretation of Kant's most prominent formulation of the Categorical Imperative, the Formula of Universal Law (FUL). While a number of existing interpretations face problems, Kleingeld argues that her new 'Volitional Self-Contradiction' (VSC) interpretation has many interpretive and philosophical advantages. Chief among these is the fact that the VSC interpretation respects what Kleingeld calls the 'simultaneity condition', namely the FUL's underappreciated requirement that a candidate maxim be willed *simultaneously* both as one's own private maxim and as a universal law, without contradiction.

✉ Michael Walschots
mhwalschots@gmail.com

¹ Martin-Luther-Universität Halle, Philosophisches Seminar, Halle, Germany

Kleingeld’s article has already received a significant amount of discussion in the literature.¹ In the following paper I wish to add to the conversation on Kleingeld’s interpretation by exploring certain aspects of her view that I feel are especially worthy of further discussion. It is my hope that this commentary is accessible to those who are not intimately familiar with Kleingeld’s paper, therefore I begin by summarizing the main argument of her article (2) before making three remarks (3–5). First (3), I seek to clarify what it means for a contradiction to be ‘volitional’ as opposed to logical. In contrast to what some interpreters have supposed, I argue in support of Kleingeld that there is a meaningful sense in which the will can contradict itself, which, although similar to a logical contradiction, is distinct in important ways. Second (4), I briefly take a closer look at the difference between Kleingeld’s interpretation and Christine Korsgaard’s ‘Practical Contradiction’ interpretation of the FUL. Although Kleingeld argues that her view has the advantage of not needing to presuppose ‘auxiliary assumptions’ such as essential purposes of the will, which is a core feature of Korsgaard’s interpretation, I argue that Kleingeld needs to accommodate one essential purpose if it is to have the interpretive force she hopes it to have, namely happiness. Finally (5), I turn to how Kleingeld’s view handles what is commonly referred to as the distinction between a ‘contradiction in conception’ (CIC) and a ‘contradiction in the will’ (CIW). On Kleingeld’s view there is only one kind of contradiction, a volitional self-contradiction, but there are still two ways in which this contradiction can be generated, which map onto the two classes of impermissible maxims that Kant describes in the *Groundwork*. I argue that more explanation is needed in order to properly understand the details of *how* a VSC is generated in both of these cases.

2 The Volitional Contradiction Interpretation of Kant’s FUL

The FUL reads as follows:

Act only in accordance with that maxim through which you can simultaneously [zugleich] will that it become a universal law (4: 421, emphasis in original).²

Kleingeld has described Kant’s FUL as “the most general and most important formulation of the principle of morality” (2019, 70). For example, it is described as “the” categorical imperative in the *Groundwork* (4:421), it is the formulation which most resembles the second *Critique*’s principle of morality (5:30), as well as Kant’s description of “the” categorical imperative in the *Metaphysics of Morals* (6:225–6; see Kleingeld 2019, 70). As a fundamental and core piece of Kant’s moral philosophy, it is perhaps no surprise that there is considerable disagreement in the literature over how we should interpret the formulation in a way that is faithful to Kant’s texts. However, interpreting the FUL is complicated for the additional reason that, as Kleingeld argues, when it comes to the contradiction that is said to arise from maxims that fail the requirements of the FUL, “Kant did not specify the nature

¹ See Timmermann (2018), Kahn (2019), and Timmons (this volume).

² All references to Kant’s works cite the volume and page number of his *Gesammelte Schriften* (see Kant 1900 ff.). In general, I have used the translations of Kant’s texts available in the Cambridge Edition of the Works of Immanuel Kant, and I have indicated where I have modified these translations.

of the relevant contradiction clearly enough” (2017, 110). Indeed, the FUL is often interpreted as requiring that one be able to will a candidate maxim as a universal law without generating a contradiction. Where interpreters disagree concerns the *nature* of the contradiction that arises when a maxim is universalized as well as the *location* of this contradiction, i.e. the specific features that contradict each other.³ The purpose of Kleingeld’s article is thus to offer a new interpretation of both the *nature* and *location* of the relevant contradiction that immoral maxims generate, according to Kant’s FUL (see *ibid.*, 90).

Kleingeld begins her paper by discussing three leading⁴ interpretations of the nature of the relevant contradiction before proposing her own reading. The first interpretation she discusses is the view attributed to Hegel and his followers (2017, 93), according to which a contradiction can only arise from universalizing a candidate maxim if agents presuppose certain values. The example that Kleingeld and Hegel discuss is theft: universalizing the maxim to steal only generates a contradiction if agents already presuppose the value of property. Here the *location* of the contradiction is supposed to lie between willing a candidate maxim as a universal law and the presupposed value (see *ibid.*, 93). The *nature* of the contradiction, on this interpretation, is unclear, or at the very least it is not specifically described as a logical contradiction, a practical contradiction, or otherwise.

According to a second leading interpretation, that of Christine Korsgaard, the nature of the contradiction generated by immoral maxims is specifically ‘practical’ in the sense that a purpose of the will is thwarted when the maxim is universalized. This can happen in two ways, for Korsgaard: the willing of a maxim as a universal law can thwart the very same purpose stated in the maxim itself when universalized (in which case a ‘contradiction in conception’ arises), or willing a maxim as a universal law can thwart a purpose that is essential to the will (in which case a contradiction in the will arises) (*ibid.*, 94; and see Korsgaard 1996, 97–8). The *location* of the contradiction, on Korsgaard’s interpretation, is therefore between the universalized maxim and a purpose of the will. As for the *nature* of this contradiction, the relevant contradiction is always a contradiction of the will: even though a will’s purpose can be thwarted in two ways, because we are always concerned with a purpose that the *will* has either necessarily or otherwise, the *nature* of the contradiction is always of the will, which Korsgaard calls a ‘practical’ contradiction.

The third leading interpretation discussed by Kleingeld is Onora O’Neill’s. On O’Neill’s view, similar to Korsgaard’s, we cannot will a maxim as a universal law in two ways: (1) “if the project stated in the maxim is *impossible* in a world in which the maxim is a universal law” (in which case a CIC would arise), or (2) “if it would be *inconsistent*, for agents who are aware of the background conditions of their own

³ In the literature on Kant’s FUL, Kleingeld’s article included, one gets the impression that there are two ways to understand the *location* of the contradiction involved. First, scholars often speak of the location of the contradiction as existing solely within the world of a universalized maxim. The more appropriate way to understand the location of the relevant contradiction, I think, and on which Kleingeld rightly focuses, concerns the two features (beliefs, volitions, etc.) that may or may not contradict one another.

⁴ Kleingeld notes that she does not provide an exhaustive discussion of all existing interpretations of the nature of this contradiction, and refers readers to the following sources for alternative interpretations: Korsgaard 1996, Timmons, 2006, and Galvin 2009.

agency, *to will* the universalized version of a maxim” (in which case a CIW would arise) (Kleingeld, 2017, 95). According to O’Neill’s view, there are famously two kinds, i.e. two distinct *natures*, of the contradiction involved: a contradiction in conception and a contradiction of the will. The *location* of the contradiction, on O’Neill’s view, is similarly in two different places: in the case of contradictions in conception, the contradiction is within the universal law itself, or within the world where the maxim is universal law (and so between the universal law itself and the project of this very same law), and in the case of contradictions of the will, the contradiction is between willing a maxim as universal law and what agents necessarily will in virtue of their status as rational agents.

The problem with all three of these leading interpretations, according to Kleingeld, is that they fail to explain two core features of Kant’s FUL: (1) that the contradiction involved is a *self*-contradiction, and (2) the formula’s ‘simultaneity condition’ (SC), i.e. the implicit condition that agents need to be able to will candidate maxims *simultaneously* as universal laws. I’ll discuss the SC in more detail below. Kleingeld concludes her survey of the leading interpretations of the FUL by claiming that they are all “dissatisfying” (ibid., 93) because they fail to explain these two core features. She therefore proposes a new interpretation of the FUL which acknowledges the importance of the SC and also takes Kant’s language of self-contradiction literally.

Kleingeld begins outlining her interpretation by observing that “Kant’s most prominent statements of the principle of morality, in each of his main works in moral theory, all include a simultaneity condition” (ibid., 97). Kleingeld describes the SC as: “the simultaneous compossibility of willing that a maxim be one’s own maxim and willing that it be a universal law” (ibid., 99–100). Perhaps even more clearly, the SC states that “the FUL can be read as requiring us to *act only on maxims that we can will as our own maxim and simultaneously will as a universal law*” (ibid., 100, emphasis in original). Kleingeld notes a few exceptions where the SC isn’t explicitly mentioned in Kant’s discussions of the FUL (see ibid., 98–9), but even in these cases she argues that the simultaneity condition might be “tacitly implied” (ibid., 110 note 48). To illustrate the SC Kleingeld considers first the maxim to borrow money while promising falsely to repay it (ibid., 100). Willing this as one’s own maxim and *simultaneously* willing that it be a universal law constitutes, she claims, a volitional self-contradiction (ibid.). The *location* of the relevant contradiction, on this interpretation, is “*between* willing that the maxim be one’s own action principle ... *and* willing that it be a universal law” (ibid.). Kleingeld contrasts this with existing interpretations, according to which the contradiction is *located* within or involved in the conception of the world where the maxim is a universal law (ibid.). On Kleingeld’s view, by contrast, the location of the contradiction is between what Rousseau calls the particular and the general will – terms which Kleingeld interestingly notes that Kant also employs in the *Observations* (see ibid., 100–101 and 20:145).

According to Kleingeld, conceiving of the location of the contradiction as between the willing of one’s own maxim and the willing of the maxim as a universal law makes it possible to conceive of this contradiction as a *self*-contradiction in a straightforward way: “On the Volitional Self-Contradiction interpretation, in the case of a failing maxim, willing the maxim as one’s own maxim contradicts willing the maxim as a universal law. In this sense the maxim (*qua* maxim of the agent)

contradicts itself (qua maxim of all)” (ibid., 102). This interpretation also provides a relatively straightforward way of understanding the *nature* of the contradiction involved in immoral maxims. Kleingeld claims that the VSC interpretation “makes it possible to read the contradiction at issue in an ordinary sense of ‘contradiction’, namely, as willing A and simultaneously willing not-A” (ibid., 103).

Kleingeld offers two illustrations of how her interpretation is supposed to work by examining how the maxim of false promising and the maxim of egoism generate a VSC. Indeed, as Kleingeld states at various points in her article, a VSC can be generated in two different ways. The maxim of false promising, for example, generates a VSC because the universalized form of the maxim itself is “impossible” (ibid., 105). Willing this maxim as one’s own and simultaneously willing the maxim as a universal law is thus contradictory, because this maxim is not in fact tenable as a universal law. The maxim of egoism, on the other hand, generates a contradiction in a different way, namely in virtue of its content: if I will the maxim of egoism as a universal law, I contradict what I necessarily will as a personal maxim, namely that I would want others to help me when I am in need.

There is more to be said here and I will discuss the two ways in which a VSC is generated in particular in Sect. 5 below. For the purposes of the following commentary, the above summary of Kleingeld’s VSC interpretation of Kant’s FUL should suffice. The thrust of the article’s main argument is that the leading interpretations of the nature and location of the contradiction involved in maxims that fail the requirements of the FUL “fail to capture important elements of Kant’s description of the FUL” (ibid., 110), in particular the SC and the fact that the contradiction involved is a *self*-contradiction. The VSC interpretation, by contrast,

has significant interpretive and philosophical advantages. Importantly, it makes it possible to take Kant’s language of ‘self-contradiction’ at face value, to do so in an ordinary sense of ‘contradiction’, and to do so without additional presuppositions such as a commitment to substantive values, a conception of essential purposes of the will, or an account of the background conditions of human rational agency (ibid., 110).

In the next Sect. (3), I take a brief look at the nature of this ‘ordinary’ sense of contradiction involved in immoral maxims, on Kleingeld’s view. Then, (4) I suggest that Kleingeld’s view is not as different from Korsgaard’s interpretation as Kleingeld suggests, and indeed that the VSC interpretation *needs* to accommodate an important feature of Korsgaard’s view if it is to have interpretive force. In the final Sect. (5) I take a closer look at the two ways in which Kleingeld believes a VSC can be generated.

3 What is a ‘Volitional’ Contradiction?

As I mentioned in the above summary, a central piece of Kleingeld’s argument is that the VSC interpretation understands the contradiction involved in maxims that fail the requirements of Kant’s FUL in an “ordinary” (ibid., 103 and 110) sense of contradiction. On the one hand, this might surprise readers, since the contradiction involved does

not seem to be ordinary at all, but is rather a special kind of contradiction, namely a “volitional *self*-contradiction.” One might therefore ask: what exactly is a “volitional” contradiction as opposed to, say, a logical contradiction, if there even is a difference? Indeed, commentators in the past have suggested that the contradiction involved is simply a logical one. Perhaps most famously, J.S. Mill suggests this in *Utilitarianism* when arguing that the categorical imperative is unfit to evaluate candidate maxims: Mill claims that Kant “fails, almost grotesquely, to show that there would be any contradiction, any logical (not to say physical) impossibility, in the adoption by all rational beings of the most outrageously immoral rules of conduct” (Mill, 1969, 207).⁵ It seems to me, however, that a meaningful distinction can be made between a logical contradiction and a *volitional* contradiction, even though the difference might not always be clear. What I propose to do in this section of my commentary is clarify what it might mean for a contradiction to be “volitional”, and thereby argue that, on Kleingeld’s interpretation, the nature of the relevant contradiction is *not* logical, but of the will.

The first thing to note is that Kleingeld’s conception of a volitional contradiction appears to be no different from a contradiction of the will. Indeed, the only reason Kleingeld avoids calling the contradiction at the centre of her interpretation of the FUL a contradiction of the will is because this term is already associated with one of the other leading interpretations of Kant’s FUL, namely O’Neill’s interpretation with her now classic distinction between a CIC and a CIW (See O’Neill 2013, 136–193). Kleingeld’s reason for calling it a volitional contradiction as opposed to a contradiction of the will is thus merely to distinguish her view from O’Neill’s. Kleingeld’s volitional contradiction is thus a contradiction of the will and on her view the relevant contradiction in maxims that fail the FUL is a *self*-contradiction of the will.

With this in mind, the next question to ask is: what, then, is a contradiction of the *will*? Is this kind of contradiction in any way different from a logical contradiction? To answer these questions, it will help to take a brief look at the way in which Kleingeld very instructively and clearly describes the nature of the contradiction at issue in maxims that fail the requirements of the FUL. Kleingeld characterizes the contradiction in terms of *simultaneously* willing A and -A. She gives the example of “willing that I eat chocolate while simultaneously willing that nobody eats chocolate.” (Kleingeld, 2017, 103) In the second part of this statement, Kleingeld clarifies that what is claimed is not that nobody *else* eat chocolate, but that nobody, myself included, eat chocolate. Accordingly, to will that I eat chocolate and *simultaneously* that nobody eat chocolate (myself included) is to will two conflicting and contradictory actions. Of the utmost importance for Kant’s view, of course, is that it can be contradictory to will two conflicting *maxims* as well as actions, for example willing that I myself act on a maxim and that nobody does.

Conceiving of the relevant contradiction involved as simultaneously willing A and -A certainly suggests that the contradiction involved is logical, but it is not. In order to see how this is so, it will help to take a brief look at the nature of a logical contradiction. One way to characterize a *logical* contradiction is to say it concerns

⁵ For others who interpret the relevant contradiction involved in Kant’s FUL as a logical contradiction, see Korsgaard (1996, 78; 81–2).

the *truth* or, similarly, one's beliefs (see for example Korsgaard 1996, 79). Logically speaking, it is a contradiction for A to be true, or for one to believe A, and for -A to be true, and for one to believe -A (at the same time). A contradiction of the will, on the other hand, does not concern truth or belief, but *volition* or *desire*. Indeed, for Kant the faculty of desire [*das Begehrungsvermögen*] is one of the three fundamental faculties of the mind (together with the faculty of cognition [*das Erkenntnisvermögen*] and the faculty of feeling [*das Gefühl der Lust und Unlust*]). Put briefly, the faculty of desire concerns the subject's ability to cause or bring about objects. Kant identifies the higher faculty of desire with (practical) reason (see 5:29–30) and identifies the will (in the sense of *Wille*) with practical reason (4:412). The faculty of cognition, on the other hand, is concerned with knowledge, belief, and in general the subject's ability to represent objects.⁶ The will as practical *reason*, however, is still subject to the norms of rationality, which makes it possible to not only speak of a contradiction regarding one's knowledge, beliefs, or representations, but also of one's volitions and desires. A contradiction of the will, therefore, is still a contradiction and a violation of the norms of rationality, but concerns an entirely different faculty of the mind, for Kant, namely the faculty of desire, as opposed to the faculty of cognition.⁷

In light of the above we can also make sense of what it means for a maxim to be *impossible* as a universal law, or for it to be *impossible* to will A and -A simultaneously. When it comes to willing A and -A simultaneously, for example, this is not something we are *physically incapable* of doing, as Mill's language in the quote above seems to suggest. Quite the contrary, if I wanted to, I could indeed *simultaneously* will that my personal maxim is to make a false promise in order to obtain money, but also that *nobody* do so as a universal law. This (physical) possibility is completely open to me. The point, of course, is not that I physically cannot do this, but that I *rationally ought* not do this. The avoidance of contradiction is a norm of rationality, and in saying that the will is identical to practical reason and that we ought to avoid willing contradictory maxims, Kant is claiming that the will, too, is subject to the norms of rationality. Willing two contradictory maxims is thus not "impossible" in any physical sense, but is rather *impossible* in the sense that it is *incapable of rational consistency*. I thus *cannot* simultaneously will A and -A *if* I want to maintain my rationality. On Kant's view, which Kleingeld's interpretation captures nicely, it seems that as rational beings we are subject to the norms of rationality with respect to both knowledge and volition.⁸

⁶ For a helpful discussion of Kant's categorization of the faculties of the mind and their various functions, as well as the many passages where Kant discusses them, see Wuerth 2014, 221–228.

⁷ Another way of making the distinction I make here is to say that there is a difference between a *theoretical* contradiction and a *volitional* contradiction, as Korsgaard has done (see 1996, 94). This would allow one to say that a contradiction of the will is still *logical*. This is presumably what Timmons has in mind when he calls Kleingeld's volitional contradiction a "logical contradiction in the will" (see his contribution to this issue). However, because speaking of a logical contradiction often implies that we are talking about truth and one's beliefs and thus a theoretical contradiction, I think it helps to distinguish between a logical and a volitional contradiction as I have done here.

⁸ It is an interesting question whether or not the faculty of feeling, Kant's third fundamental faculty of the mind, is also subject to the norms of rationality. There is some disagreement over the possibility of "rational feelings" in Kant's philosophy (see e.g. Cohen 2018), and perhaps there must be the equivalent to theoretical and practical reason for the faculty of feeling in order for feeling to be subject to the norms

It is indeed an advantage of Kleingeld’s view that it describes the contradiction involved in maxims that fail the requirements of the FUL in an “ordinary” sense of contradiction. At the same time, this “volitional self-contradiction,” i.e. contradiction of the will, is not identical to a logical contradiction, the latter of which I have described as concerning knowledge or belief. My aim in this section was, in support of Kleingeld’s interpretation, to outline a way in which it makes meaningful sense to talk about a contradiction of the will in contrast to a logical contradiction. A contradiction of the will concerns volition, and the willing of two opposites *simultaneously*. Although similar to a logical contradiction, a contradiction of the will does not signal a physical impossibility, but rather reflects a normative requirement of rationality. As identical to practical reason, the will, on Kant’s view, and our volitions as a result, are subject to the norms of rationality by nature.

4 Volitional Contradictions and Essential Purposes

According to Kleingeld, one of the main advantages of the VSC interpretation is that it allows us to understand Kant’s language of a *self*-contradiction of the will in a straightforward way. The VSC interpretation accomplishes this by explaining how a VSC does not require us to presuppose “auxiliary assumptions” (Kleingeld, 2017, 102) like substantive values, essential purposes of the will, or the background conditions of rational agency (see *ibid.*, 110 and 113); a maxim or the will⁹ simply contradicts *itself*. In this section I’d like to discuss one of these ‘auxiliary assumptions’ in a little more detail. In particular, I’d like to consider whether or not it is in fact an advantage that Kleingeld’s view does not require us to bring in essential purposes of the will. I hope to show that, if it is to have interpretive force, Kleingeld’s view must allow for the presupposition of at least one essential purpose, namely happiness. If I am right, this means that Kleingeld’s view must have an important feature in common with Korsgaard’s ‘practical contradiction’ interpretation. I begin with a brief summary of Korsgaard’s interpretation before discussing happiness as the one essential purpose that both Korsgaard and Kleingeld ought to allow for. I conclude by arguing that Kleingeld’s discussion of the maxim of egoism in fact presupposes this essential purpose. My analysis shows that either Kleingeld’s VSC interpretation is not as distinct from the ‘practical contradiction’ interpretation as her paper suggests, or more work needs to be done to accommodate happiness within the VSC interpretation.

of rationality. It is unclear what a contradiction of feeling would even look like, but I leave the exploration of such questions for another occasion.

⁹ Throughout her article, Kleingeld states that the volitional self-contradiction she has in mind is *either* of the will (see 2017 90, 101, 102, 103) *or* a maxim (*ibid.*, 97, 102, 90). At the same time, it seems that a contradiction of a maxim amounts to a contradiction in *willing* two (or more?) maxims that contradict each other, in which case a contradiction of a maxim reduces to a contradiction of the will. I don’t explore this any further in this paper but it would be advantageous for Kleingeld’s interpretation if it were indeed true that the relevant contradiction was always of the will, not of a maxim, and this strikes me as a better interpretation of Kant’s texts as well.

According to Korsgaard’s ‘practical contradiction’ interpretation, the relevant contradiction at issue in maxims that fail the FUL’s requirements consists in both willing and not willing an end (see Korsgaard 1996, 93). Similar to Kleingeld, this is just another way of saying that the relevant contradiction is always a “contradiction in the will”, which is not to be confused with a logical or theoretical contradiction (ibid., 94), which Korsgaard describes as *believing X and -X simultaneously* (see ibid., 79). More specifically, on Korsgaard’s view, “a thwarted purpose is a practical contradiction” (ibid., 96). This can happen in two ways. First, for the cases that usually fall under the umbrella of a CIC, a maxim’s own purpose is thwarted when the maxim becomes a universal law. A clear example is the false promising case: universalizing the maxim to make a false promise in order to borrow money thwarts the very purpose stated in the maxim, borrowing money, because the universalization of the maxim makes it such that no promises to repay money would be believed. Second, and this is the case that is relevant for my discussion in this section, the universalization of a maxim can thwart a purpose that is essential to the will, and these are the cases that usually fall under the umbrella of a CIW:

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. (ibid., 96).

Although she admits it is only a sketch, Korsgaard lists the following examples of essential purposes: “its [the will’s] general effectiveness in the pursuit of its ends” (ibid., 96); “its freedom to adopt and pursue new ends” (ibid.); and, in the example of developing one’s talents, the person who does not develop their talents thwarts the essential purpose of “the development of his rational nature” (see ibid., 104 note 24).

It would be interesting to complete Korsgaard’s ‘sketch’ and outline a complete list of the will’s essential purposes. For my purposes here, I’d like to focus on one further essential purpose that we can attribute to the will, on Kant’s view, namely our own happiness. As Kant says in the *Groundwork*:

There is *one* end that can be presupposed as actual in all rational beings [...] and thus one purpose that they not merely *can* have, but that one can safely presuppose they one and all actually *do have* according to a natural necessity, and that is the purpose of *happiness*. (4:415, emphasis in original)

Presumably, Korsgaard is interested in listing the essential purposes of the will that go above and beyond this core end or essential purpose, that is, she is interested in the kinds of things a rational agent must necessarily be committed to *as* a rational agent, such as willing the necessary means to one’s ends. Nonetheless, Kant is clear that happiness is a *necessary* end or purpose of the human being that we can assume all human beings have. This is important, especially for Korsgaard’s understanding of the FUL, because a maxim is morally impermissible if, when universalized, it thwarts this purpose and thus gives rise to a ‘practical contradiction.’

What I would like to suggest is that Kant's understanding of happiness as a necessary end slightly complicates Kleingeld's VSC interpretation. Although Kleingeld considers it an advantage that her view does not require us to presuppose essential purposes of the will, the VSC interpretation needs to accommodate at least this one essential purpose of the will if it is to have interpretive force. It is not clear how the VSC interpretation can do so, however, in light of the fact that Kleingeld wishes to exclude all possible *auxiliary* assumptions in order to show how the will can contradict *itself*. At the same time, in her analysis of how the VSC view applies to the maxim of egoism, Kleingeld seems to acknowledge that Kant attributes happiness as an essential purpose to the agent in question: "Kant [...] asserts that the egoist wishes [wünscht] to be helped when he needs [bedarf] it and that he hopes that others will assist him." (Kleingeld, 2017, 107) The egoist's wish to be helped when needed is of course derivative of a more basic desire for one's own happiness. Put differently, the desire to receive help from others when we cannot help ourselves is simply desiring the necessary means to our own ends. According to Kleingeld, we can 'capture' the egoist's commitments in a maxim in the following way: "I shall never help others in need of assistance, but I shall want others to help me when I need their help to achieve my own ends" (ibid.). The egoist thus has a purpose that we presuppose, namely achieving their own ends, and it is only on the presupposition of this purpose that a contradiction arises when attempting to universalize the maxim of egoism. It is therefore unclear that Kleingeld's VSC interpretation actually avoids presupposing essential purposes or needs of the will. Again, this is not a bad thing, given Kant's claim that we have at least one essential purpose, but we need an explanation of how this essential purpose is compatible with Kleingeld's view if it is to be as persuasive as possible.

In this section I have attempted to show that Kleingeld's view needs to accommodate Kant's claim that rational beings have one essential purpose: their own happiness. Although she considers it an advantage that the VSC interpretation does not need to presuppose auxiliary assumptions like essential purposes of the will, Kleingeld's view needs to accommodate this one essential purpose if it is to have interpretive force. Furthermore, I have shown that her examination of the maxim of egoism actually acknowledges that a purpose of the egoist agent must be presupposed in order for the maxim of egoism to generate a contradiction when universalized. Kleingeld's view is thus similar to Korsgaard's in this respect, and indeed they *must* be similar in this way if they are to have interpretive force. In order to maintain her conception of the nature of the relevant contradiction at issue in maxims that fail the requirements of the FUL, Kleingeld thus needs to explain how it is possible to allow for at least one 'auxiliary assumption' about the will's necessary purposes.

5 The Two Ways of Generating a Volitional Self-Contradiction

One of the results of the VSC interpretation is that "some of the standard terminology used to describe Kant's position – though not used by Kant himself – starts to look less felicitous" (ibid., 111). One of the two pieces of terminology Kleingeld discusses, the other being universalizability (see ibid., 111 ff.), is the "well-known" (ibid.) distinction between a contradiction in conception (CIC) and a contradiction

in the will (CIW), introduced by Onora O’Neill in her *Acting on Principle*.¹⁰ This distinction and subsequent discussions of the details of the FUL were inspired by an important passage in the *Groundwork* where Kant distinguishes between two ways in which a maxim can fail the requirements of the FUL:

Some actions are such that their maxim cannot even be *thought* without contradiction as a universal law of nature, let alone that one can also/still [*noch*] *will* that it *ought* to become one. In the case of other actions, however, that inner impossibility is not found, but it is still impossible to *will* that the maxim [of such actions] be raised to the universality of a law of nature, because such a will would contradict itself. (GMS, AA 04: 424, emphasis in original)

According to Kleingeld’s interpretation of this passage, which she supports with textual evidence (see 4:437, 4:424) there are no longer two *kinds* of contradiction, but just one: “*all* failing maxims generate contradictions between willing the maxim as one’s own and willing it as a universal law. In other words, *all* failing maxims generate self-contradictions in the will, or, in the terminology adopted in this essay, volitional self-contradictions.” (Kleingeld, 2017, 111, emphasis in original) At the same, in order to explain the above passage Kleingeld argues that there are two ways in which a contradiction in the will can be *generated*. What I would like to do in this section is take a closer look at how Kleingeld understands the two ways in which she claims a VSC can be generated. Although Kleingeld *identifies* the two ways in which a VSC is generated, I argue that we need more details concerning *how* a VSC is generated in each case.

First, allow me to briefly summarize Kleingeld’s interpretation of the above passage and her account of the two ways in which a VSC can be generated. First, concerning the cases which usually fall under the umbrella of a CIC, Kleingeld claims that

a volitional self-contradiction emerges in cases in which universalizing a maxim makes acting on it impossible. If the universalization of a given maxim necessarily and immediately makes acting on this maxim impossible, then willing that the maxim be my individual principle and simultaneously willing that this maxim be a universal law constitutes a volitional self-contradiction. (ibid., 104)

Kleingeld’s claim here seems to be that there are some cases, such as that of false-promising and suicide, where “the maxim cannot coherently be a universal law” (ibid., 104) and thus “the action specified in the maxim would be *impossible* if the maxim were a universal law” (ibid., 105, my emphasis), i.e. “if the maxim is a universal law, acting on the maxim is *impossible*” (ibid., 105, my emphasis). This is most clear, again, from the example of making a false promise to borrow money: the action (false promising) is impossible if the maxim is universalized, because promises cannot be made if everyone expects them to be false. On Kleingeld’s view, if one wills a maxim as a universal law despite the fact that it is *incoherent* as a universal law, a

¹⁰ O’Neill (Nell) 1975, 2013 and see Kleingeld (2017: 111) and Timmons (2006: 164) for further discussion.

VSC is generated: “if the universalization of a maxim necessarily and immediately makes acting on the maxim impossible, then willing the maxim as one’s own and simultaneously willing it as a universal law yields a volitional self-contradiction” (ibid., 113). Thus, in the first kind of case, the relevant contradiction is still a VSC, but it is generated in a specific way, namely by the fact that the maxim is *incoherent* as a universal law, and that the action specified in the maxim is *impossible* if the maxim were to be a universal law. On Kleingeld’s view, calling this case a contradiction in conception is therefore a “misnomer” (ibid., 107) because the contradiction is still of the will, it is just generated by the *incoherence* of a maxim as a universal law.

In a second kind of case, which usually falls under the umbrella of a CIW, Kleingeld argues that rather than the maxim’s *incoherence* as a universal law, it is the maxim’s *content* that makes it such that, when universalized, a VSC is generated: “in the case of some maxims, it is contradictory to will the maxim as one’s own and simultaneously to will that it be a universal law, simply by virtue of its content” (ibid., 104). She offers the maxim of egoism as an example: on Kleingeld’s analysis, this maxim is *coherent* as a universal law, i.e. a world in which every rational being gives limitless priority to their own interests over those of others does not make it *impossible* to pursue one’s own interests. The problem in these cases is rather that “the content of the maxim makes it impossible, without volitional self-contradiction, to will to act on the maxim and simultaneously to will that the maxim be a universal law” (ibid., 107).

In order to explain the above passage from the *Groundwork*, and in order for the VSC interpretation to account for the two separate kinds of cases usually described in terms of a CIC and a CIW, Kleingeld thus specifies two features which can *generate* a VSC: (1) the incoherence of a maxim as a universal law, i.e. when the universalization of a maxim makes acting on that maxim impossible; and (2) the content of the maxim itself. Although Kleingeld has *identified* these two features, what I would like to suggest in the following is that it is unclear *how* exactly they each generate a VSC, and that once one starts to fill in the picture, the VSC interpretation starts to encounter difficulties.

Let’s begin with the first kind of case. Before addressing the question of *how* a VSC is generated in this case, let me first make a remark. According to Kleingeld’s analysis of the kind of case that usually falls under the umbrella of a CIC, if a maxim is incoherent as a universal law, then it would seem that, from the point of view of practical reason, the mere willing of that maxim as a universal law is problematic in itself, aside from the fact that there might be a contradiction between simultaneously willing the maxim as one’s own and willing it as a universal law. Even though Kant references a contradiction in *thought* in the passage quoted above, I think Kleingeld is right to suggest that Kant is primarily concerned with the problems this conceptual incoherence poses for *willing*. At the same time, what would be problematic here is willing a conceptually incoherent universal law in the first place. This, however, places the relevant problem for the first kind of case in a different location, namely in the act of willing an incoherent universal law itself, rather than between willing a maxim as one’s own and as a universal law simultaneously. Even if it is true that willing a maxim as a universal that is incoherent as a universal law *generates* a contradiction in the will, this seems to be the second mistake that the will makes in this process. The first mistake is willing a maxim as a universal law that is incoherent as a universal law.

Kleingeld might admit this and reply, however, that the *relevant* contradiction is the one that exists between the private and the general will. If we were to grant this (even though I think one can argue this point), the next question to ask would be: *how* exactly does willing a maxim as a universal law, which is incoherent as a universal law, in fact generate a VSC? Kleingeld argues, in line with her interpretation, that this scenario is equivalent to willing A and -A simultaneously or, more specifically, “it amounts to willing *that I act* on the maxim and simultaneously willing *that nobody acts* on the maxim (or, alternatively, willing *that I can act* on the maxim while willing *that nobody can act* on it)” (ibid., 104). My question here is: if I nonetheless attempt to will a maxim *as* a universal law that is incoherent as a universal law, but I cannot because the maxim is in fact incoherent as a universal law, am I necessarily committed to willing the *opposite* of that maxim as a universal law? I’m not so sure. If a maxim is incoherent as a universal law, by definition I cannot will it as a universal law, since it does not actually make sense as a law. I must, therefore, be willing something else as a universal law when I attempt to do so. We need an explanation of why this must necessarily be the opposite of the maxim I attempted to universalize, such that a VSC is generated. Otherwise, again the primary issue, it seems to me, is the fact that we try to will a maxim that is incoherent as a universal law in the first place, not that doing so generates a VSC.

For the second class of cases, Kleingeld argues that a VSC is generated due to the maxim’s “content.” In these cases, the maxim under consideration is coherent as a universal law, but the content of the maxim makes it such that a VSC is generated when one tries to will it both as one’s own and as a universal law simultaneously. But *how* does the content of a maxim generate a VSC, when one tries to will it simultaneously as one’s own and as a universal law? Not *all* content generates a VSC when a maxim is universalized, for in this case no maxim could be a universal law. The question is, then: what is the relevant *kind* of content that it would be contradictory for us to will in this way? One answer here, of course, is the answer offered above, namely that a VSC is generated in those cases where a maxim has the kind of content such that, when we will it as a universal law, we contradict something we necessarily will as rational beings. Another option is the kind of content which, when the maxim is universalized, it contradicts a necessary end of *nature*. This is one way in which to understand why the maxim of suicide fails the requirements of the FUL, i.e. because its universalization would contradict the final end of nature, namely to preserve and promote rational nature.¹¹ Kleingeld might resist both of these kinds of explanations, however, because they both smuggle in ‘auxiliary assumptions’ about the essential purposes both of the will and of nature. We therefore need an account of the relevant *kinds* of content that Kleingeld has in mind in order to explain *how* a VSC is generated by maxims with that content, if her interpretation is to be successful.

¹¹ See for example the analysis of the maxim of suicide offered in Klemme 2017, 158–9.

6 Conclusion

In this paper, my aim has been to offer some comments on Kleingeld's VSC interpretation of Kant's most prominent formulation of the categorical imperative, the FUL. As I have tried to argue here, there are aspects of Kleingeld's view that are extremely plausible, but others that require further explanation if they are to be ultimately persuasive. It is my hope that my remarks will advance the discussion on what is to my mind one of the most plausible interpretive options available on a core feature of Kant's moral philosophy. I also hope it will encourage others, as it has done for me, to explore the intricacies of both Kant's FUL and Kleingeld's interpretation of it in more detail.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing Interests The author declares that he has no conflict of interest.

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/>.

References

- Cohen, A. (2018). Rational feelings. In K. Sorensen, & D. Williamson (Eds.), *Kant and the Faculty of feeling* (pp. 9–24). Cambridge University Press. <https://doi.org/10.1017/9781316823453>
- Galvin, R. (2009). The Universal Law Formulas. In T. Hill Jr. (Ed.), *The blackwell guide to Kant's Ethics* (pp. 52–82). Oxford University Press.
- 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. (1900 ff.). *Gesammelte Schriften* (Akademie Ausgabe). Vol. 1–22 edited by Preussische Akademie der Wissenschaften, Vol. 23 edited by Deutsche Akademie der Wissenschaften zu Berlin, Vol. 24 ff. edited by Akademie der Wissenschaften zu Göttingen.
- 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>
- Klemme, H. (2017). *Kant's 'Grundlegung zur Metaphysik der Sitten': Ein systematischer Kommentar*. Reclam.
- Korsgaard, C. M. (1996). *Creating the Kingdom of Ends*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139174503>
- Mill, J. S. (1969). *Collected Works of John Stuart Mill Volume X: Essays on Ethics, Religion and Society*. Edited by J. M. Robson. University of Toronto Press.
- O'Neill (Neill), O (1975) (2013). *Acting on Principle: An Essay on Kantian Ethics*. New York. 2013 Second Edition: Cambridge University Press. <https://doi.org/10.1017/CBO9781139565097>
- 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. (2006). The Categorical Imperative and Universalizability (GMS, 421–424). In C. Horn and D. Schönecker (Eds.), *Groundwork for the Metaphysics of Morals* (pp. 158–199). De Gruyter.
- Wuerth, J. (2014). *Kant on mind, Action, and Ethics*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199587629.001.0001>

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.