

*“A free will and a will under moral laws are the same”:
Kant’s concept of autonomy and his thesis of
analyticity in Groundwork III*

Dieter Schönecker

I INTRODUCTION

In section 1 of *GMS* III, Kant claims that “a free will and a will under moral laws are the same” (447.6–10).¹ This claim expresses Kant’s concept of autonomy; after all, the concept of freedom is the “key” (446.6) to this concept. So if we understand freedom in its relation to morality, we understand autonomy. But how are we to understand this relation? On the standard reading, Kant’s claim is that a free will is under the moral law as a Categorical Imperative (CI); hence, once it is shown that we are free it is shown that we are obliged by the CI.² This interpretation, I will argue, is incorrect.

To holy beings the moral law is certainly non-imperative. But what moral law does Kant refer to when he claims that “a free will and a will under moral laws are the same” – is it the moral law as an imperative or the moral law in its non-imperative form? There is also no doubt that, according to Kant, the moral law as an imperative is synthetic. Now suppose the moral law in that claim is the non-imperative moral law – is this non-imperative moral law synthetic as well, or is it analytic? My answers are these: by stating that “a free will and a will under moral laws are the same,” Kant claims that with regard to perfectly rational and free beings,

¹ All textual references to the *Groundwork* are to, or are based upon, Allen Wood’s translation and edition (Yale University Press, 2002); changes were made in consultation with Wood. All references to the German text of the *Grundlegung* are to the edition by B. Kraft and D. Schönecker (Hamburg: Meiner Verlag, 1999). All other textual references are to *The Cambridge Edition of the Works of Immanuel Kant* (Cambridge University Press, 1992–). Numbers in parentheses refer to vol. IV of the *AA*, e.g., 447.6–10 refers to *GMS* 4:447, lines 6 to 10. I refer to the subchapters of *GMS* III as “sections” (section 1, etc.). I am grateful to Oliver Sensen for very helpful comments on the first draft of this chapter.

² Cf., for instance, O’Neill 1989: 294; Wood 1999: 171–76; and Guyer 2010.

or beings considered just as members of the intelligible world, the moral law is descriptive, not prescriptive; and therefore, it is not a synthetic, but an analytic proposition. This is to say that the ‘analysis’ of the very concept of such a being yields that by its very nature such a being always and necessarily wills morally. I call this claim about the analytic relation between freedom and the moral law *Kant’s thesis of analyticity*. This thesis is at the heart of Kant’s concept of autonomy. Unless we understand the former, we won’t understand the latter.

I will begin with three preliminary methodological remarks (the second section of this chapter). In the third section I will then lay out my interpretation of the overall structure of *GMS* III. This section as well as the fourth section on the analytic relation between freedom and morality basically reflect what I have set out elsewhere in German.³ However, whereas my earlier interpretations were quite superficial, I now try to get deeper into the text and its problems by a *kommentarische* interpretation.

2 THREE PRELIMINARY METHODOLOGICAL REMARKS

I have repeatedly argued that almost all literature on Kant suffers from *Textvergessenheit*, and that the remedy to this is to read Kant *kommentarisch*.⁴ Such a *kommentarische* interpretation follows, among others, two principles. First, it sharply distinguishes between the question of what a text *means* and the question of whether the arguments it manifests are *sound*. Considerations about the plausibility of a given theory are an important tool in interpreting a text; but if an interpretation *a* is available that is better than an interpretation *b* – roughly, an interpretation is better if it coherently (i.e., consistently and comprehensively) accounts for more textual observations than another – then *a* must be preferred over *b* even if *b* ascribes a theory to the text that seems more attractive than the theory based on *a*. For instance, the thesis of analyticity brings about the question of how evil actions are possible: if to act freely is to act morally, and vice versa, then there appears to be no room for evil actions that are free.⁵ If the best interpretation shows that bisubjunction to be Kant’s position, it should not be given up because it has an unwelcome implication (from ‘our’ point of view); maybe the implication is a reason to give up

³ Cf. Schönecker 1999; a very short English version was provided in Schönecker 2006.

⁴ Cf. Schönecker 2004 as well as Damschen and Schönecker 2012: 201–70.

⁵ I cannot address this problem here; cf., however, Schönecker 2011.

that bisubjunction but it is not a reason to give up the interpretation if the latter is the best available.

Second, theoretical claims are almost always part of a larger, comprehensive theory, and they are expressed in a certain context. This is a truism, of course, but a truism often neglected. As for Kant's theory of autonomy in *GMS*, it requires that we must not interpret section I without integrating it into a comprehensive interpretation of *GMS* III as a whole.

Third, Kant's theory of autonomy in the *GMS* is one thing, his theory thereof in other writings might be another. In this chapter, I will deal only with Kant's theory of autonomy in the *GMS* and not take into account, for instance, his *Religionsschrift*.

3 THE OVERALL STRUCTURE OF *GROUNDWORK* III: A SKETCH

In *GMS* I and II, Kant develops his ethical theory: if there is such a thing as morality, it must be understood in terms of duty and the CI. Whether there *is* morality is not decided at the end of *GMS* II; it may very well be merely a "figment of the mind" (407.17; 445.8). The crucial question is this: "How is a categorical imperative possible?" This question is already raised in *GMS* II (417ff.), but only the parallel question – how are hypothetical imperatives possible? – is answered there. In *GMS* II (425) that crucial question is raised again, but its answer is only given in section 4 of *GMS* III; the second paragraph of section 4 begins with the formulation "And thus categorical imperatives *are* possible" (454.6; my emphasis).

Kant's argument in *GMS* III has always been reconstructed as follows. A free will is a will under the CI; freedom must be presupposed as a quality of the will of all rational beings; human beings are rational beings; therefore, the human will is under the CI. Since in this interpretation premise 1 *is* proven in section 1, premise 2 *is* argued for in section 2, and premise 3 in section 3, the answer to the question of how a categorical imperative is possible would already be given in section 3. However, we have already noticed that it is only in the *fourth* section that Kant does provide the answer. At the end of section I Kant states the main problem how categorical imperatives are "possible" (447.15), and he explicitly says that the solution to this problem "cannot be directly indicated here" (447.22); rather the answer to the question of how categorical imperatives are possible "still needs some preparation" (447.25).

So how do we need to reconstruct *GMS* III? Here's my proposal. The question "How is a categorical imperative possible?" actually includes

three aspects: (1) why is the CI valid?;⁶ (2) why and in what sense may we consider ourselves to be free?; (3) how can pure practical reason bring about an interest in the moral law? As section 5 shows, the third question cannot be answered. The first and second questions, however, are being answered: in section 1, Kant puts forward his thesis of analyticity according to which a perfectly rational being, or a being considered just as a member of the intelligible world, by its very nature always follows the moral law; to such a will, the moral law is not, however, an imperative. Section 2 demonstrates by means of a transcendental argument – one cannot deny freedom by presupposing it – that a rational being, due to its ability to think spontaneously (i.e., freely), must consider itself practically free; section 3 then refers to the difference between the world of understanding and the world of sense in order to argue that the human being too must understand himself as practically free, i.e., as a being that has an intelligible will with the moral law as the law of its causality. But then Kant still thinks that the answer to that crucial question of why the CI is valid (binding) is *still not answered*; in section 3 Kant *still* asks “*from whence the moral law obligates*” (450.16). The answer to this question is found only in section 4. It is based on what I call the *ontoethical principle*. Very much abbreviated, it says: the world of understanding and thus the pure will as a member of this world of understanding are ontically superior to the world of sense, and therefore the law of this world and will (the moral law) is binding as a categorical imperative for beings that are both members of the world of understanding and the world of sense.⁷

In the world of understanding, the moral law is descriptive. It describes how a holy being acts (to wit, morally), or how a being would act if it were only a member of the world of understanding (to wit, morally). This is why Kant writes in section 4 of *GMS* III (453–54) that all actions of a human being, if it were *only* a member of the world of understanding,

⁶ By the ‘validity’ of the CI I mean its absolute and necessitating bindingness for beings that are both members of the world of understanding and the world of sense; to say that the CI is valid is to say that there is always and whatever my other interests are overriding reason for me to comply with it.

⁷ Cf. 453.31–454.5. The German sentence is highly challenging; for a detailed analysis see Schönecker 1999: 364–96. It is remarkable that to the present day this crucial sentence (453–54) is ignored even by those who try to interpret *GMS* III; Guyer 2010 and Rauscher 2009, for instance, do not even mention it. Of all the literature I know about *GMS* III published since 1999, Henry Allison’s new commentary on the *Groundwork* (2011) is the only one that pays close attention to the text. I am very grateful to Professor Allison, who made available to me the draft of his new book. He sharply (and sharp-wittedly) attacks my thesis of analyticity. I have no space here to reply to it (a reply will be published separately), but thinking about Allison’s own interpretation made me rethink my own and this helped me better see the important points.

"*would* be perfectly in accord with the principle of the autonomy of the pure will" (453.26); if a human being were "alone" a member of the world of understanding, all its actions "*would* always be in accord with the autonomy of the will" (454.8). That autonomy and hence morality is a "consequence" (453.13; 461.17) is only true "under the presupposition of freedom of the will *of an intelligence*" (461.14; my emphasis). For only "*as intelligence*" (453.17; my emphasis) does a human being understand himself as a member of the world of understanding. In section 5 Kant argues that the human being must "think of him[self] as intelligence, also as thing in itself" (459.22) and hence of himself as the "*authentic self*" (*eigentliches Selbst*, 457.34; my emphasis). As such an 'authentic self,' a human being gives himself the law. That's why the moral ought is "*really* a volition" (*eigentlich ein Wollen*, 449.16; my emphasis), and that's why it is "his *own* necessary volition as a member of an intelligible world" (455.7; my emphasis). The moral law as a non-imperative law is the causal law of such a will; by strength of it we describe the volition of such a will.

Kant's summarizing answer to the question "from whence the moral law obligates" reveals the entire structure of *GMS* III:

And thus categorical imperatives are possible through the fact that the idea of freedom makes me into a member of an intelligible world, through which, *if I were that alone*, all my actions *would* always be in accord with the autonomy of the will; but since I intuit myself at the same time as member of the world of sense, they *ought* to be in accord with it, which *categorical* "ought" represents a synthetic proposition *a priori* by the fact that to my will affected through sensible desires there is also added⁸ the idea of precisely the same will, but one belonging to the world of understanding, a pure will, practical for itself, that contains the supreme condition of the first in accordance with reason. (454.6–15; my emphasis)

Later Kant points out again that the moral law "is valid [!] for us as [!] human beings, since [!] it has arisen from our will as intelligence, hence from our authentic self; *but what belongs to the mere appearance is necessarily subordinated by reason to the constitution of the thing in itself*" (461.2–6). The human being, Kant says, as a rational being is a member of the world of understanding, and "*since* in that world he himself only as intelligence is the *authentic self* (as human being, by contrast, only appearance of himself), those laws [of the world of understanding] apply to him immediately and categorically" (457.33; my emphasis). Recognizing my duty as that which I authentically will also motivates me to fulfill my duty: "The

⁸ As we will see, this 'addition' is why Kant calls the CI synthetic.

moral 'ought' is thus his own necessary volition as a member of an intelligible world and is thought by him as an 'ought' only insofar as he at the same time considers himself as a member of the sensible world" (455.7; original emphases removed).

It is obvious that Kant's deduction is open to severe criticism even if one accepts the basics of his transcendental philosophy. Kant avails himself of an ontological interpretation of his own distinction between thing in itself and appearance that otherwise is merely an epistemological distinction; and he is unable to account for evil actions. In any event, if I am right, then the interpretation I propose consistently and comprehensively accounts for the overall structure of the text as well as for its details, and it does so more than any other interpretation; as an *interpretation*, it must therefore be preferred over those other interpretations even if they yield an argument that might be more plausible.

4 WHAT IS THE RELATION BETWEEN FREEDOM AND THE MORAL LAW? KANT'S THESIS OF ANALYTICITY AND HIS THEORY OF AUTONOMY

Kant's thesis of analyticity: the bottom line

Kant's thesis of analyticity does *not* mean that the free will of a sensuous-rational being is "under" (447.7) the moral law if this is taken to mean that sensuous-rational beings are obligated by the CI. They are indeed, but that they are obligated is something that Kant after sections 2–3 has yet to demonstrate. After all, this is why he still raises the following question *after* he has argued for his thesis of analyticity (section 1) and the claim that freedom must be presupposed as a quality of the will of all rational beings (section 2): "But why ought I to *subject* myself to this [moral] principle?" (449.11; my emphasis). In section 3, Kant still asks "from whence the moral law obligates," a question to which "no satisfactory answer" (450.2) has been given yet. This structural observation also implies that a free will and a will under the CI are *not* 'the same,' and this renders the standard interpretation of the thesis of analyticity untenable.

So how *are* we to understand Kant's thesis of analyticity? Let's develop the answer in several steps.

1. It is undisputed that according to Kant the moral law for human beings involves a "*necessitation*" (413.4) because they are always "subject to subjective conditions (to certain incentives)" (412.36); this is why to human beings the moral law is a *duty* or categorical *imperative*. To beings

that are not under such "limitations and hindrances" (397.8) and that therefore have a perfectly good will, the moral law is not an imperative:

A perfectly good will would thus stand just as much under objective laws (of the good), but it would not be possible to represent it as *necessitated* by them to lawful actions, because of itself, in accordance with its subjective constitution, it can be determined only through the representation of the good. Hence for the *divine* will, and in general for a *holy* will, no imperatives are valid; the *ought* is out of place here, because the *volition* is of itself already necessarily in harmony with the law. Hence imperatives are only formulas expressing the relation of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, e.g. to the human being. (414.1)

The moral law of a holy will I call the *non-imperatival moral law*; it is descriptive, not prescriptive, and it is therefore (I will argue), analytic, not synthetic.

2. Again, it is undisputed that Kant draws a line between a human being that is under the moral law as an imperative and a holy being that always acts in accordance with the moral law anyway. However, not as much attention has been paid to the fact that this line is not as firm as appears at first sight. For Kant also describes the human will, *as being a member of the world of understanding*, as such a perfectly good will. I have already provided textual evidence for this, and there is more. Right in the midst of the deduction in section 4, Kant writes: "As a mere member of the world of understanding, all my actions *would* be perfectly in accord with the principle of the autonomy of the pure will" (453.25–27; my emphasis). A bit later he says: "And thus categorical imperatives are possible through the fact that the idea of freedom makes me into a member of an intelligible world, through which, *if I were that alone*, all my actions *would always* be in accord with the autonomy of the will" (454.6–9; my emphasis). And at the end of the section 4 Kant concludes: "The moral 'ought' is thus his own necessary volition as a member of an intelligible world and is thought of by him as an 'ought' *only insofar* as he at the same time considers himself as a member of the sensible world" (455.7–9; my emphasis). So if we consider ourselves just as being members of the world of understanding, our will must be described as perfectly good; if we had only such a will, i.e., if we were only member of the world of understanding, we would always act morally. To such a will the moral law is a descriptive, non-imperatival moral law as well.

3. The question "How are all these imperatives possible?" is posed for the first time in *GMS* II (417.3). Kant himself clarifies what the question is really about, to wit, "how to think the necessitation of the will" (417.5). By

“all” imperatives Kant means both categorical and hypothetical imperatives, and since “necessitation” means both the necessity of imperatives as well as their motivational force based on their necessity, the question is this: why should and how can a being that is rational but also sensuous and hence under the influence of inclinations comply with imperatives that infringe upon and strike down these inclinations? The answer Kant gives with regard to hypothetical imperatives, though certainly disputable in itself, is quite revealing when it comes to our understanding of the thesis of analyticity: hypothetical imperatives are “analytic” (417.29; 419.10) *because*⁹ of the truth of the proposition that whoever wills an end, also wills, if she is perfectly rational, the means to the end;¹⁰ this latter proposition Kant also calls “analytic” (417.11; 417.23). Parallel to this, I submit, the non-imperative moral law is analytic whereas the CI is synthetic.

Five times Kant calls the CI a synthetic-practical proposition *a priori*.¹¹ But why does he call it ‘synthetic’ at all? Kant says little about it, and he says nothing about the question of whether the distinction between analytic and synthetic theoretical propositions is useful with regard to practical propositions, let alone with regard to imperatives. I don’t think it is, either from a non-Kantian or from a Kantian point of view. Imperatives cannot be synthetic simply because they are not propositions. However, in a sense that is not taken strictly (not based on meaning and the principle of contradiction) one can understand what Kant was driving at with the claim that the CI is synthetic.

In *GMS* II Kant speaks for the very first time of a “synthetically practical proposition” (420.14); right after ‘proposition’ Kant puts an explanatory footnote:

[Sy1] I connect the deed *a priori* with the will, without a presupposed condition from any inclination, hence necessarily (although only objectively, i.e. under the idea of reason, which would have full control over all subjective motivations).
 [Sy2] This is therefore a practical proposition that does not derive the volition of an action analytically from another volition [of an action] already presupposed (for we have no such perfect will), but is immediately connected with the concept of the will of a rational being, as something not contained in it. (420.29–35)

This footnote is not as easy to understand as it might seem; let’s read it *komentarisch*. To begin with, what does ‘This’ at the beginning of [Sy2]

⁹ Cf. the “*Denn*” in 417.30.

¹⁰ Cf. 417.8; 417.22; 417.31; for an interpretation of Kant’s answer with regard to hypothetical imperatives see Schöncker and Wood 2011: 103–5, 112–17.

¹¹ Cf. 420.14; 440.24–26; 444.35; 447.10; 454.11. Note that Kant does not always use the complete formula (‘synthetic-practical proposition *a priori*’).

refer to? Naturally, one would think it refers back to the preceding sentence, i.e., to [Sy1]. However, we must not forget that this footnote is to explain Kant's claim that the CI is a 'synthetically practical proposition.' The main text itself reads: "Secondly, with this categorical imperative, or law of morality, the ground of difficulty (of having insight into its possibility) is very great indeed. It is a synthetically practical proposition*" (420.12). Hence the 'proposition' mentioned again at the beginning of [Sy2] as a 'practical proposition' is the CI as a 'synthetically practical proposition' ('It'). [Sy1] is itself part of the explanation of CI as a 'synthetically practical proposition,' but [Sy1] is *not* the proposition itself that is referred to by "This."¹² Further, note that, although the entire formulation is "It is a synthetically practical proposition* *a priori*," the asterisk (*) is placed after "proposition"; so one expects Kant to account for the '*synthetic-practical*' character of the CI rather than for its *a priori* character, and for the most part, as we will see, this is true. Kant does discuss the syntheticity of the CI right away by introducing the idea of a 'connection' at the very beginning of [Sy1].¹³ Certainly, yet [Sy1] is about the *a priori* character of this 'connection' (the CI is *a priori* because there is no "presupposed condition from any inclination"); so the placement of the asterisk might very well be a mistake. In any event, the footnote is *mainly* about the synthetic character of the CI.

Some interpreters were tempted to think that in [Sy2] Kant compares categorical with hypothetical imperatives.¹⁴ In one respect this is true, because hypothetical imperatives *are* practical propositions that *do* "derive the volition of an action analytically from any other volition already presupposed"; after all, such hypothetical imperatives derive the 'volition of an action' (e.g., to practice the piano) from 'another volition' (e.g., to become a great pianist). However, Kant does *not directly* compare categorical with hypothetical imperatives here. It is *not* that Kant says in [Sy2]: "The CI is a practical proposition that – *unlike hypothetical imperatives* – does not derive the volition of an action analytically from another volition already presupposed (as in hypothetical imperatives in which the volition of the means is derived from the volition of the related end)." Rather, the parallel of that formulation in [Sy2] – "derive the volition of an action analytically from another volition" – is to the volition of a *holy*

¹² It is noteworthy, however, that Kant at this stage of *GMS* II has *not* yet introduced the formula of the CI, but just its "mere concept" (420.19); it's only right after the footnote that Kant derives that formula.

¹³ Cf. "I connect" ("*Ich verknüpfe*") in [Sy1]; in [Sy2] it says "connected" ("*verknüpft*"). Literally speaking, of course, 'synthesis' comes from 'connecting,' 'putting together.'

¹⁴ Cf. for instance Pieper 1989: 266 and Wagner 1994: 78.

being. For the volition ‘already presupposed’ in [Sy2] is the volition of a ‘perfect will’ as Kant says in parenthesis (“for we have no such perfect will”). Such a perfect will – a holy will or a will just considered as a will of the intelligible world – always wills the good.¹⁵

That parenthesis in [Sy2] (“for we have no such perfect will”) is very helpful indeed. Given Kant’s understanding of a perfect will as a will of a being that is not under “limitations and hindrances,” such a being’s will and volition is always moral and must not be connected to the moral law since it already is; to such a being’s will, the moral law is non-imperative. With regard to any volition of an action of such a being we know that it will be morally good; from the general character of such a will we can ‘derive analytically’ that any instance of its volition is good. If we take the concept of universalizability to be the key element in the moral law, then the following proposition is analytic:

(PW) All beings with a perfect will only have maxims that can become a universal law.

One could argue that even if this proposition is analytic, it would not follow that the moral law *itself* is analytic.¹⁶ But what then could it possibly mean that the moral law even in its non-imperative form is synthetic? What *is* the non-imperative moral law beyond PW? When it comes to perfect beings the law of their actions is such that all their actions are guided (ruled) by maxims that can become a universal law; but in what sense could this law be synthetic? Suppose we say:

(NIML) All maxims of beings with a perfect will can become a universal law.

If anything, then this is the non-imperative moral law. And I simply do not see in what sense this would be synthetic. We describe what beings with a perfect will *are*; and we do it by means of a law that is part of the meaning of ‘perfect will.’ Just as Kant claims that hypothetical imperatives are analytic ‘because’ of the analyticity of the proposition that whoever wills an end, also wills, if she is perfectly rational, the means to the end, so the moral law is analytic because of the analyticity of the proposition that all beings with a perfect will only have maxims that can

¹⁵ Note, further, that the comparison between hypothetical and Categorical Imperatives regarding the ‘presupposed condition’ has been discussed at length in the previous two paragraphs (419.12–420.11). There, Kant raises again (after 417.3) the question of “how the imperative of *morality* is possible” (419.12), and he argues that one reason why this is hard to answer is that the CI “is not at all hypothetical and thus the necessity, represented as objective, cannot be based on any *presupposition*, as with the hypothetical imperatives” (419.13; my emphasis). So the footnote does not need to, and does not, address the difference between categorical and hypothetical imperatives because it has already been addressed in the two previous paragraphs.

¹⁶ As Allison does; see his 2011: 168, 276ff.

become a universal law. On the other hand, the will of a being that is under "limitations and hindrances" must be *connected* to the moral law, and this connection ('synthesis') implies necessitation; to such a being, the moral law is an imperative. This imperativeness of the CI is the reason why Kant calls it synthetic. If there is syntheticity, there is imperativeness; and if there is imperativeness, there is syntheticity. The analysis of 'perfect being' yields that such a being always wills morally; in this sense PW is analytic. The analysis of 'imperfect being' does not yield such a result; yet it ought to will morally, i.e., its will must be connected with moral volition, and in that sense the moral law is 'synthetic.' This is all, and this is as good as it gets, i.e., as good as Kant's theory of analytic vs. synthetic practical propositions gets.¹⁷

So as crude as the basic idea of syntheticity might be, as simple it is.¹⁸ In the context of the footnote in 420, however, two more questions need to be addressed. First, why does Kant call the CI a '*practical* proposition'? It seems as if the 'therefore' (*also*, 420.32) at the beginning of [Sy2] indicates the reason for the 'practicality' of the proposition. However, the 'therefore' refers back to the 'I connect' and thus indicates a reason for the syntheticity of that proposition. The CI is a 'practical proposition' that, *unlike another* 'practical proposition' – to wit, the non-imperative moral law – connects the moral volition with the will of a being that is not perfect. A practical principle is not 'practical' because it involves a necessitation; it is 'practical' because it has to do with volitions and actions.¹⁹ Second,

¹⁷ Allison (2011: 168) writes that for Kant "it is an analytic truth that a perfect will would do whatever the moral law requires, because the thought of its failure to do so contradicts the concept of such a will." Nonetheless he adds in a footnote: "This does not mean, however, that the moral law is itself analytic" (Allison 2011: 168n.). But I think this is exactly what it means: the moral law as the CI is synthetic because it connects an imperfect will with the volition of the morally good; the non-imperative moral law is analytic because from the concept of a perfect being it follows that its volition is moral. Allison himself says: "The point rather seems to be that for a perfect will the connection between its volition and the course of action, which for finite rational agents is required by a Categorical Imperative, would be analytic" (2011: 168). But just as the synthetic connection is the reason to call the Categorical Imperative synthetic, this 'analytic connection' is the reason to call the non-imperative moral law analytic. Again, strictly speaking the entire talk of analytic vs. synthetic practical propositions makes little sense; but that's true on any reading. In the second *Critique* (*KpV* 5:31) Kant says that under certain conditions the moral law would be "analytic"; however, this might have a specific meaning in the context of Kant's theory of the "*Faktum der Vernunft*."

¹⁸ I say this with regard to the CI as a synthetic proposition; later I will distinguish between a methodological, a conceptual, and a propositional meaning of 'analytic' and 'synthetic.'

¹⁹ Cf. 412–13, where Kant speaks interchangeably of 'laws' and 'principles' and introduces them independently of the question of whether they are necessitating or not. It is tempting to think of Kant's definition in the first *Critique*: "Praktisch ist alles, was durch Freiheit möglich ist" (*KpV* B82:8). However, Kant has different concepts of freedom in the first *Critique*, one of which is 'naturalized,' and this is later given up; cf. Schönecker 2005.

what exactly is ‘derived analytically’? For reasons that are brought to light in section 1, the volition of a perfectly free and rational being is always moral; so whatever *particular* ‘deed’ or ‘volition of an action’ is required by the moral law, this will, simply as a perfectly free will, wills it anyway.

I have outlined in what larger context Kant’s thesis of analyticity stands. Let me add one more piece to the puzzle. In section 3 Kant famously discusses the notorious thread of a “circle” (450.18; 453.4).²⁰ Up to the point where Kant introduces that alleged circle, the human being as an intelligence had not been justified; and even if we may assume that the human being is an intelligence, i.e., a member of the world of understanding, it would still be *begging the question* to simply assume that *thereby* the validity of the CI has been proven. So at the end of section 3, Kant may claim that as

a rational being, hence one belonging to the intelligible world, the human being can never think of the causality of its own will otherwise than under the idea of freedom; for independence of determinate causes of the world of sense (such as reason must always attribute to itself) is freedom. Now with the idea of freedom the concept of *autonomy* is inseparably bound up, but with the latter the universal principle of morality. (452.31)

Here Kant repeats his thesis of analyticity from section 1: with freedom autonomy is ‘*inseparably bound up*,’ and with autonomy morality. This is why Kant after his solution to the alleged circle concludes:

For now we see that if we think of ourselves as *free*, then we transport ourselves as members into the world of understanding and cognize the autonomy of the will, together with its *consequence*, morality; but if we think of ourselves as *obligated* by duty, then we consider ourselves as belonging to the world of sense and yet at the same time to the world of understanding (453.11–15; my emphasis)

The first part of this passage really is once more nothing but a reformulation of the thesis of analyticity; this is why Kant says that ‘morality’ is a ‘*consequence*’ [*Folge*] of autonomy.²¹ The second part after the semi-colon, however, makes clear why for sensuous-rational beings the moral law is an imperative.²²

²⁰ As a matter of fact, it’s not a *circulus in probando* but a *petitio principii*; for a detailed analysis see Schönecker 1999: 317–58.

²¹ Cf. 461.14: “Under the presupposition of freedom of the will of an intelligence, its *autonomy*, as the formal condition under which alone it can be determined, is a necessary *consequence*” (second emphasis mine).

²² In the context of the ‘circle’ Kant also says that “freedom and the will giving its own laws [*Gesetzgebung*] are both autonomy, hence reciprocal concepts” (450.23). But he also speaks in the very same context of his “inference [*Schlusse*] from freedom to autonomy and from the latter to the moral law” (453.4). But if freedom and autonomy are reciprocal *concepts*, how then can we

So the bottom line is this: The analysis of the concept of a perfectly free and rational will, or of a will that is just considered as a member of the intelligible world, shows that such a will necessarily follows the moral law. A perfectly free and rational being, or an intelligible will, is "under" (414.1) the moral law just as an imperfect sensuous-rational being. However, such a being or will is not necessitated by the law and therefore, for such a being or will, the moral law is not an imperative. For imperfect beings, or for beings both with an intelligible and sensuous will, the moral law is an imperative; and this is also why Kant calls it "synthetic." The moral law, inasmuch as it is not an imperative, is not a synthetic but an analytic practical proposition.

Kant's thesis of analyticity: an interpretation of section 1

Let's now turn to a *kommentarische* interpretation of section 1 (though I can only address a very limited part of it). Its title is: "The concept of freedom is the key to the definition [*Erklärung*] of autonomy of the will" (446.5). It is of the utmost importance to see that 'autonomy' (and its laws: *auto-nomy*) is a term that is *not* limited to finite beings. The term makes its first brief appearance in 433, where Kant explicitly says that it "leads to a very fruitful concept depending on it, namely that of *a realm of ends*" (433.15). To this realm of ends, which is ruled by the laws of morality, i.e., by the laws of autonomy, also belongs "a fully independent being, without need and without limitation of faculties" (434.5) that Kant calls "*supreme head*" (433.36). Again, Kant repeats his thesis that for such a being, although it is under the moral law, the moral law is *not* a CI: "Duty does not apply to the supreme head in the realm of ends" (434.18). And this point is yet again repeated in direct relation to the expression of 'laws of autonomy' when Kant writes (recall): "The will whose maxims necessarily harmonize with the laws of autonomy is a *holy*, absolutely good will. The dependence of a will which is not absolutely good on the principle of autonomy (moral necessitation) is *obligation*. Thus the latter cannot be referred to a holy being" (439.30). So when Kant says that "the concept of freedom is the key to the definition of the autonomy of the will," this cannot mean that the concept of freedom is the key to the definition of the autonomy of the *finite* will only. It is exactly this difference between finite (unholy) and infinite (holy) beings that makes section 1 so confusing; and

infer from one to another? Here too it shows that Kant's use of terms such as 'analytic,' 'reciprocal,' 'inference' (even 'deduction') in this practical context is of very limited help.

it is this difference that makes Kant later say that, although ‘the concept of freedom is the key to the *definition* of the autonomy of the will,’ “freedom and the will giving its own laws are both autonomy, hence reciprocal concepts, of which, however, just for this reason, *one cannot be used to define* [erklären] the other and provide the ground for it” (450.23). This tension brings up the ‘suspicion of a circle’ that consists in erroneously believing that once freedom is presupposed, the validity of the CI has been demonstrated.

For reasons of space I will not discuss Kant’s basic concepts of transcendental and practical freedom, though it will be necessary to briefly address Kant’s argument *why* freedom, autonomy, and morality are so closely related. My main interest is to understand the meaning of the thesis of analyticity which I find above all in these sentences:

[A1] Thus a free will and a will under moral laws are the same.

[A2] [a] Thus if freedom of the will is presupposed, [b] then morality, together with its principle, follows from this by mere analysis of its concept. [A3] [a] Nonetheless, the latter is always a synthetic proposition: [b] an absolutely good will is that whose maxim can always contain itself considered as universal law, [c] for through analysis of the concept of an absolutely good will that quality of the maxim cannot be found.

Let’s analyze section 1 in several steps.

1. *The problem.* For now I ignore the logical relation of [A1] to the preceding paragraphs (‘thus’) and reformulate it in a way that, I take it, is uncontroversial:

[A1]* A free will is a will under moral laws.

But what does it mean to be ‘under’ moral laws? Of course, Kant holds human beings to be ‘under’ the Categorical Imperative such that the CI is valid for them, i.e., obligating and thus necessitating. However, note that Kant expressly says that a holy being “would thus stand *just as much under* objective laws (of the good)” (414.1; my emphasis), but yet not under laws as necessitating imperatives,²³ and that a holy, absolutely good will is a will “whose maxims *necessarily* harmonize with the *laws* of autonomy” (439.28; my emphasis). With these two meanings of ‘under’ in mind – ‘under’ the non-imperative moral law, ‘under’ the CI – and just looking at [A1]*, it could mean either:

²³ Cf. 433.26.

[A1]** The free will of a human being considered as a member both of the sensible and the intelligible world is a will under the CI.

or

[A1]*** The free will of a holy being and the free will of a human being considered as a member of the intelligible world is a will under the non-imperative moral law.

I propose [A1]*** as the correct reading, i.e., I propose to understand [A1] as a formulation of the thesis of analyticity. On my interpretation, [A1] has two elements: descriptivity and analyticity. With regard to perfectly rational and free beings, or beings considered just as members of the intelligible world, the moral law is descriptive, not prescriptive; *and* [A1] is not a synthetic, but an analytic proposition. If it should turn out that my second claim – [A1] is an analytic proposition – is false, it could still be the case that [A1] is non-imperative. So in that case my claim that [A1] states a conjunctive thesis of analyticity would be falsified; however, as for the overall interpretation of *GMS* III and its structure this would be of no great consequence because in this regard it is only important to see that section 4 (not sections 2–3) makes the crucial, concluding deductive step by demonstrating the validity of the CI.

2. *The argument for the thesis of analyticity.* Clearly, [A2] needs to explain further what [A1] means. Ignoring the proviso expressed in [A2a] – that freedom of the will for now is just ‘presupposed’ – [A2] says this:

[A2]* Morality follows together with its principle from freedom of the will by mere analysis of its concept.

Obviously, the ‘analysis’ mentioned in [A2b] is just what Kant has provided in the first two paragraphs of section 1; there he speaks of a negative and positive “*definition* of freedom” (446.13; my emphasis) which at the same time is “the key to the definition of autonomy of the will” (446.5). We must now turn, albeit very briefly, to this ‘definition’ and thus to the argument for the thesis of analyticity.

[A1] states that “*thus* a free will and a will under moral laws are the same.” Why ‘thus’? The argument is simple:

- 1 Freedom of the will is autonomy.
- 2 Autonomy is the quality of the will of being a law to itself.
- 3 Therefore: freedom of the will is the quality of the will of being a law to itself.

Since Kant obviously identifies ‘being a law to itself’ with ‘being under moral laws,’ the final conclusion that ‘a free will and a will under moral

laws are the same' is valid. But this argument alone doesn't help much for the interpretation of [A1] and [A2]. For now the question arises what it means that the will 'is a law to itself': are we to understand this in terms of [A1]** or [A1]***? The bottom line of section I seems to be this: a free will is a will that is not determined by natural causes and their laws (negative freedom); these causes are "alien" (446.9) causes, i.e., causes that do not stem from myself. Yet such a will must be determined by something, by some kind of law; there is no lawless causality. Since all causality is either natural causality or moral causality,²⁴ and natural causality is ruled out by negative freedom, there remains only moral causality, i.e., causality by the moral law that, unlike the natural determination from "something else" (446.23), is determination by oneself which, of course, is exactly what *auto-nomy* means. Note that Kant speaks of autonomy as "the quality of the will of *being* a law to itself" (447.1; my emphasis); the will, he says, "*is* in all actions a law to itself" (447.2; my emphasis). The moral law is a law of causality, it is descriptive, not prescriptive.

3. *But what does 'analysis' mean in [A2]?* The central discussion of how to read [A1] can only be settled if there is a clear understanding of how Kant in the *GMS* uses the terms 'analytic' and 'synthetic.' So far we have considered the analyticity and syntheticity of propositions and imperatives. There are, however, (at least) three basic meanings that need to be distinguished in order to avoid confusion: a methodological, a conceptual, and a propositional meaning. In a methodological context, 'analytic' (or 'synthetic') qualifies a procedure; in a conceptual context, 'analytic' refers to the analysis of concepts; in a propositional context, 'analytic' (or 'synthetic') qualifies propositions. The propositional meaning of analytic/synthetic I have explained. As for the methodological meaning, 'analytic' refers to the idea that one first tries to understand what morality is all about (assuming with common moral cognition that it's real) and then asks whether there is such a thing in the first place. This, I believe, is at least part of the meaning of 'analytic' when Kant says at the end of section 2: "This section [*GMS* II], therefore, like the first one, was merely analytical" (445.7).²⁵ Now when Kant in [A2] speaks of the "*analysis*" (*Zergliederung*) of concepts – as he does in a passage on the concept of

²⁴ Cf. *KrV* A532/B560; *Lectures on Metaphysics* LI 28:327; *NF* 27:1322.

²⁵ Cf. 392. Note, however, that in the *preface* the term 'method' refers to the 'transitions,' not to 'analytic' and 'synthetic' (so the 'method' is to make these transitions). There is no 'analytic method' or 'synthetic method' in the *Groundwork* as it is understood in the *Prolegomena* (4:276n.); if there were, then *GMS* III would follow the analytic method as well (which is not the case); see Schönecker 1997.

autonomy in *GMS* II (440.23; 440.29)²⁶ – does he primarily refer to the methodological meaning of ‘analytic’ such that this conceptual analysis is part of the analytic procedure of *GMS* I/II, or does he refer to the analysis of concepts that results in analytic propositions? Put another way: is it possible that in [A2] Kant refers to the methodological and/or²⁷ conceptual meaning of ‘analytic’ *without* thereby implying that this analysis necessarily yields analytic *propositions*? To answer this we must analyze [A1]–[A3].

4. *The third paragraph of section I ([A1]–[A3]).* [A3] has a most remarkable opening with “Nonetheless” (*Indessen*, 447.10), introducing an opposition between what has just been said and what is said now; any reading of section I must account for this. The “latter” (*das letztere*, 447.10) can only refer back to the “principle” (447.9), i.e., the principle of morality mentioned in the previous sentence [A2]. Abstracting briefly from the “Nonetheless,” we thus get:

[A3a]* The principle of morality is always a synthetic proposition.

Now if ‘synthetic’ here implies the imperativeness of the moral law, it is undeniable that, literally speaking and on the face of it, this is not and cannot be Kant’s position; the principle of morality is not ‘*always*’ an imperative. For as we have seen, Kant himself repeatedly argues that for holy beings the laws of morality (which are the laws of autonomy)²⁸ are *non-imperative*. As we have seen as well, the qualification ‘synthetic proposition’ is introduced explicitly with regard to the (possibility of) the Categorical Imperative. So the expectation should be that the ‘principle of morality’ in [A3a] is the CI. But if that is true, why then does Kant say in [A2a–b] that, if freedom of the will is presupposed, then morality together with its principle follows from this by mere *analysis* of its concept?

[A3a] ends with a colon, followed by a sentence [A3b] that quite obviously is intended to spell out the principle of morality (‘the latter’ just mentioned in [A3a]) as a synthetic proposition:

[A3b]* An absolutely good will is that will whose maxim can always contain itself considered as universal law.

²⁶ Also compare the idea of a “*development* of the generally accepted concept of morality” in 445.2 (my emphasis).

²⁷ I say ‘and/or’ inasmuch as the conceptual analysis belongs to the analytical procedure (though it’s not identical with it).

²⁸ Cf. once more 439.

To be brief, let's call this "quality of the maxim" (447.14) – 'can always contain itself considered as universal law' – the *moral* quality. With regard to this moral quality, [A3c] then provides the reason (*denn*, 447.12) why the principle of morality is always a synthetic proposition:

[A3c]* Through analysis of the concept of an absolutely good will the moral quality of the maxim cannot be found.

Now isn't this confusing? If [A1] and [A2] really formulate the thesis of analyticity, why then does [A3] suddenly speak of the very same moral law as a *synthetic* proposition? [A1] and [A2] seem to state that the "analysis" (447.9) of the concept of a free will yields the moral principle, [A3], however, seems to claim that the "analysis of the concept of an absolutely [*schlechterdings*] good will" (447.12) does *not* yield the moral principle.²⁹

But let's look more closely into this. The 'absolutely good will' referred to in [A3a] shows up in [A3b] as well. As we have seen, Kant describes the 'holy will' as an "absolutely good will" (439.29). However, it's important to see that the 'absolutely good will' in [A3] *cannot* be what Kant elsewhere calls a holy will; for the 'analysis' of such a holy will *does* yield that its maxims always *have* that moral quality – recall that from a 'perfect will' the morality of any volition (action) *can* be 'analytically derived' (420). So what is the 'absolutely good will' in [A3]? The absolute goodness of a holy will consists in the goodness of its "subjective constitution" (414.4) due to which *all* its maxims are moral and cannot be otherwise. The goodness of the will of a sensuous-rational being has no such 'constitution.' Its will is 'absolutely good' *when* and *inasmuch* as it has a maxim that has the moral quality. After all, this is Kant's famous claim in chapter 1 of the *GMS*: the only thing that is "good without limitation" (393.6) is a "good will" (393.7). And Kant even *calls* the good will of a being whose 'constitution' *is* such that it is, unlike holy beings, "under certain subjective limitations and hindrances" (397.7) an '*absolutely* good will,' provided its will is determined by a maxim that can be a universal law: "That *will* is *absolutely good* which cannot be evil, hence whose maxim, if it is made into a universal law, can never conflict with itself" (437.6). In this passage and context Kant clearly does not talk about holy beings. For he continues: "This principle is therefore also its supreme law: 'Act always in accordance with that maxim whose universality as law you can at the same time will'; this is the single condition under which a will can never be in conflict with

²⁹ From Stattler 1788 [1968]: 239–40, through Ross 1954: 71–72, to Korsgaard 1989: 339, and Guyer 2009: 179–82, [A1–3] have caused much confusion and perplexity in the literature; for many more examples in the literature see Schönecker 1999: 168–71.

itself, and such an imperative is categorical" (437.9). The point is not that such a will *could* 'never conflict with itself'; it can, but it *does* not as long as its maxim can be a universal law. So Kant refers to the will as 'absolutely good' only inasmuch as its (particular) *maxim* can be universalized: "The absolutely good will, whose principle must be a categorical *imperative*" (444.28; my emphasis).³⁰ It is true that the moral principle formulated in [A2b] sounds descriptive rather than prescriptive.³¹ But the formulation in the passage just quoted (437) is descriptive as well: 'That will *is* absolutely good which *cannot* be evil, hence whose maxim, if it *is* made into a universal law, *can* never conflict with itself' and yet Kant clearly refers to the CI ('and such an imperative is categorical').

So what's Kant point? To say in [A3a] that the principle of morality is '*always*' a synthetic proposition cannot literally be true by Kant's own book. For as we have seen in our analysis of 420, Kant argues that if a perfect will is presupposed, the volition of the moral action can be 'derived analytically' which is to say that the moral law is analytic; and this is what Kant means by saying in [A2] that 'morality *follows* together with its principle from freedom of the will by mere *analysis* of its concept.' The 'always' only makes sense, and does make great sense indeed, if we read it as emphasizing that for *sensuous-rational beings*, i.e., for beings that are members both of the world of sense and understanding, the moral law is synthetic, i.e., an imperative. So first Kant states:

[A3a]** For sensuous-rational beings, the principle of morality is always a synthetic proposition.

Then he states the CI in [A3b]*, arguing that, as opposed to the 'analysis' of the freedom of a perfect will ([A2]), the 'analysis' of an imperfect will, even if it is 'absolutely good' with regard to a given maxim, cannot show that it is good with regard to any maxim; this opposition – perfect will and analysis on the one side, imperfect will and synthesis on the other – is the only way to make sense of the 'Nonetheless.' With the 'Nonetheless' Kant sets off the syntheticity of the CI from the analyticity of the 'principle of morality.'

But could it not be that the 'analysis' mentioned in [A2] must be understood in terms of a conceptual analysis that is part of the analytic

³⁰ Cf. 426.10; 437.24; 437.32. So Allison 2011: ch. 10 is incorrect in claiming that no parallel passages can be quoted in which Kant speaks of a '*schlechtedings guten Willen*' with regard to the CI.

³¹ Allison 2011: 281 is right to criticize me for not paying attention to this point in Schönecker 1999.

procedure which does not necessarily yield analytic propositions?³² I don't think so. First, to *emphasize* that despite everything said before ('Nonetheless') the principle of morality is a synthetic *proposition* would be strange if the opposite of this synthetic proposition were not another (analytic) proposition but a conceptual analysis or even analytic *procedure*. This is not a conclusive objection, but still worth mentioning. Second, when Kant speaks of "*Zergliederung*" of *concepts*, he typically means a conceptual analysis that results in analytic propositions. This is quite evident both in the first *Critique* and in the *Prolegomena*.³³ Although conceptual analyses can be an element in analytic procedures, it is undisputable that in [A2] Kant speaks of the analysis of concepts; and if such an analysis results in analytic propositions, the 'principle' in [A2] that follows from such an analysis must be an analytic proposition. Even if it were true that for Kant not every analysis of a concept results in an analytic sentence, it seems obvious that in the case of a 'holy will' or 'perfect will,' the *very meaning* of these terms is laid out in a sentence that is analytic inasmuch as it says in other words what the subject term means. If Kant's distinction between analytic and synthetic sentences makes sense at all, then – just as the very meaning of 'bachelor' is 'unmarried man' – the very meaning of 'perfect will' is 'will that only has maxims that can become universal laws' (PW). Third, and most importantly, the very idea of the syntheticity of the moral law is inseparably related to its imperativeness. As we have learned from that footnote in 420, there can be no doubt that in [Sy2] Kant claims that if a perfect will – the perfection of which consists, among other things, in its freedom – is presupposed, its moral volition can be 'derived analytically.' The analyticity of this analytic derivation, however, *cannot* be understood in terms of the methodological meaning of 'analytic'; it's not that Kant *says* that we somehow assume that there is such a thing as a perfect being and then asks what is involved. This is indeed what he *does*; but in 420 it is clearly *not the point*. The point, rather, is that if one analyzes the concept of a will that is not under the influence of inclinations, then to such a being the moral law is non-imperative. Now Allison would agree with this. However, to say it is non-imperative *is* to say that it is analytic; that's the whole point of introducing the analytic–synthetic distinction in the first place. [Sy2b] simply would make no sense otherwise.

³² This is an essential element of Allison's interpretation.

³³ Cf. *KrV* A5ff./B9ff; and *Prol* 4:274.

I conclude by saying once more that the crucial question is *not* whether the moral law under which a free will is such that a free will and a will under moral laws are the same, is analytic or synthetic. The crucial question is whether the law in this relation and context is understood as a *non-imperative* moral law or as the CI. If the latter, there simply is no way to make sense of *GMS* III; if the moral law in [A1] is the non-imperative moral law, then it should be understood as analytic as well. The third interpretative version – the moral law in [A1] is both non-imperative and synthetic – could possibly be a way to go; but such an interpretation must make sense of the overall structure of *GMS* III. So the bottom line of my interpretation of *GMS* III is untouched even if the non-imperative moral law is synthetic rather than analytic; if, however, the moral law in [A1] is the CI, my interpretation collapses. But so would the *Groundwork* itself and with it its theory of autonomy.

KANT ON MORAL AUTONOMY

EDITED BY
OLIVER SENSEN

Tulane University



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107004863

© Cambridge University Press 2013

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2013

Printed and Bound in the United Kingdom by the MPG Books Group

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Kant on moral autonomy / edited by Oliver Sensen.
p. cm.

Includes bibliographical references and index.

ISBN 978-1-107-00486-3 (hardback)

1. Kant, Immanuel, 1724–1804. 2. Free will and determinism.

I. Sensen, Oliver.

B2799.F8K36 2013

170.92—dc23 2012020416

ISBN 978-1-107-00486-3 Hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to in
this publication, and does not guarantee that any content on such websites is,
or will remain, accurate or appropriate.