



ARTICLE

# The Will of All in Kant’s *Groundwork*

T. A. Pendlebury

University of Chicago, Chicago, IL, USA

Email: [tapendlebury@uchicago.edu](mailto:tapendlebury@uchicago.edu)

## Abstract

In Kant’s *Groundwork* II, the Formula of Universal Law (FUL) seems to be the argumentative link between the notion of a categorical imperative and later formulae (e.g. of humanity), its function as this link dependent on its equivalence to both. Some commentators have denied this equivalence and read the section as a failure. Others have abandoned its expository development by reading later formulae into the FUL. I argue that we need do neither if we distinguish the universality of the FUL from that of the will of all and read *Groundwork* II as extracting the latter from common moral cognition.

**Keywords:** categorical imperative; formula of universal law; kingdom of ends; common moral cognition; practical judgement; action from duty; will

## 1. Introduction

Among the apparently central themes of Kant’s ethical theory is that of universality. Even those with a cursory acquaintance with his ethics will recognize his Formula of Universal Law (FUL): ‘act only according to that maxim through which you can at the same time will that it become a universal law’ (G, 4: 421).<sup>1</sup> This formula, he claims, gives expression to the categorical imperative, the single principle of morality.

It is, then, unsurprising that much criticism of Kant’s ethics takes the FUL as its principal target. My concern here is with a kind of criticism which alleges that the FUL is unfit to serve its assigned role in the argument of the *Groundwork of the Metaphysics of Morals* and especially in the *Groundwork*’s second section (henceforth ‘*Groundwork* II’). In that text, the FUL stands between Kant’s discussion of the character of categorical imperatives – principles of action insensitive to agents’ given purposes – and his introduction of such other central notions as those of humanity as an end in itself, the autonomy of the rational will, and the kingdom of ends. And it appears to suffer defective argumentative connections both to what precedes it and to what follows it. I shall describe these apparent problems in detail in §2, but here is a sketch of their basic structure. On the one hand, Kant seems to claim that an examination of the sheer notion of a categorical imperative reveals that there is a single categorical imperative and that the FUL is its expression. But it seems that this is false. It seems that FUL differs from the notion of a categorical imperative in both sense and extension. On the other hand, Kant claims that principles introduced later in *Groundwork* II, formulated in terms of the notions of humanity as an end in itself,

the autonomy of the rational will, and the kingdom of ends, are equivalent to the FUL and thus themselves expressions of the single categorical imperative. But this too seems false. It seems that the Formula of Humanity (FH) and Formula of Autonomy (FA) differ from the FUL in, again, both sense and extension.

These two problems are bad enough, but they yield a third: if the FH and FA are not equivalent to the FUL and the FUL is not equivalent to the notion of a categorical imperative, then the FUL cannot serve as the argumentative link between what precedes it and what follows it which Kant appears to intend it to be. *Groundwork* II, whose central task is apparently the procession of the three formulae from the notion of a categorical imperative, is a comprehensive failure.

Commentators have tended to respond to these apparent problems in one of two ways. First, there are those who believe them to be real and, indeed, insoluble, so that the expository development of *Groundwork* II must be declared a failure, though, they ordinarily hold, much of philosophical value can be extricated from it.<sup>2</sup> Second, there are those who believe these apparent problems to be merely so, who believe that the notion of a categorical imperative and the FUL, once properly understood in light of the FH and FA, are seen to be equivalent to each other and to the FH and FA.<sup>3</sup> But this approach seems also to require interpretative abandonment of the expository structure of *Groundwork* II, at least insofar as it does not include a demonstration of how the FUL itself and the material preceding its introduction provide for this reading.<sup>4</sup>

My aim in this article is to argue for another approach whose possibility emerges from two things: a recharacterization of the argumentative problems I have described and an application to them of three other important themes from Kant's practical philosophy. The recharacterization is in terms of three different notions of universality: the first apparently embodied in the sheer notion of a categorical imperative, the second in the FUL, and the third in the later formulae. In these terms, the problems above can be described as that of how these notions are related to one another and of how this relation is expressed by the text of *Groundwork* II in such a way as to support its argumentative structure. The three themes to be brought to bear are the idea of the *Groundwork* as proceeding from what Kant calls 'common moral cognition', the notion of a 'typic of pure practical judgement' introduced in the *Critique of Practical Reason*, and the notion of a capacity (*Vermögen, Fähigkeit*) at work in Kant's Critical philosophy.<sup>5</sup> In the end, I hope to have shown with the help of these resources that the basic argumentative development of *Groundwork* II is to be acquitted of the two charges described above, though problems about the FUL itself remain.

I begin by describing in more detail the two problems about the *Groundwork*'s interpretation introduced above and by recharacterizing them in terms of the three different notions of universality (§2). I then consider what Kant says about the relation between the second and third notions in the second *Critique* (§3). With this in hand, I turn to *Groundwork* I (§4) and II (§5) and present an account of the latter's argumentative structure in light of §§2–3.

## 2. Two argumentative problems

### 2.1 The gap in the derivation

The first problem is that of the gap in the derivation.<sup>6</sup> The gap, if there is one, is thoughtfully marked in *Groundwork* II by a paragraph break:

[W]hen I think of a categorical imperative I know at once what it contains. For since besides the law the imperative contains only the necessity of the maxim to conform with this law, whereas the law contains no condition to which it was limited, nothing is left but the universality of a law as such, with which the maxim of the action ought to conform, and it is this conformity alone that the imperative actually represents as necessary.

There is therefore only a single categorical imperative, and it is this: *act only according to that maxim through which you can at the same time will that it become a universal law.* (G, 4: 421)

The notion of the possibility of willing that my maxim become a universal law is then illustrated by the application of Kant's universalizability procedure to his four examples (G, 4: 421–4). This illustration funds our understanding of the notion: what it means for it to be possible for me to will that my maxim become a universal law is given by the procedure.<sup>7</sup> Thus, suppose I consider whether to eat this piece of cake, which no one else wants, because it looks delicious. For it to be possible for me to will that my proposed maxim become a universal law is for it to be possible for me to will this and at the same time will a world in which everyone, as a matter of course, eats pieces of cake which no one else wants because they look delicious. The universality of the universalizability test is the universality of cake-eating under circumstances relevantly similar to mine: similar, that is, according to the content of my maxim.<sup>8</sup> Call this *maxim-universality*. A maxim is universalizable if I can act on it and at the same time will that it should have maxim-universality.

Before the break, Kant seems to say that the categorical imperative expresses the necessity of the conformity of an agent's maxim to an unconditioned law. In the context, in which Kant has contrasted categorical with hypothetical imperatives (G, 4: 414f.), it is clear that the kind of condition of relevance is that of an agent's given purposes. A hypothetical imperative expresses the necessity of my doing something, conditional on my having some purpose (4: 414): should it be my purpose to bake a cake, I must preheat the oven. A categorical imperative expresses the necessity of my doing something whatever my given purposes might be: 'The categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end' (4: 414). The law to which I must, according to such an imperative, conform is universal not just in the sense of applying to all rational beings but also in being universal with respect to given purposes, in being indifferent to their content. Let us call this kind of universality *purpose-indifference*. It is important that purpose-indifference is a property of imperatives and that an imperative is 'the representation of an objective principle in so far as it is necessitating for a will' (4: 413). Suppose I tell you to make me some toast and in so doing betray no concern for your given purposes. Though I use the imperative mood, what I express is not an imperative in Kant's sense: the representation of a principle necessitating for your will. So although it may be apt to describe my instruction as insensitive to your purposes, it is not a case of purpose-indifference in the sense I have introduced.

There appears to be a gap in the derivation because Kant appears to proceed as follows. First, he defines categorical imperatives as purpose-indifferent imperatives. He

then – this is meant to be the derivation – claims that the one and only purpose-indifferent imperative is this: act only on maxims with maxim-universality. But the notion of maxim-universality is not the same as that of purpose-indifference, notwithstanding the use of the word ‘universal’ in the definition of each. Thus, he must give some indication of how to confirm or disconfirm a candidate principle’s claim to be a purposive-indifferent imperative and then show that this criterion applies favourably to the FUL and *only* to the FUL. Consider a competitor to the FUL: the candidate principle ‘do whatever Lytton Strachey would do in your situation’. It may seem ludicrous to suppose that this principle might be an imperative, but at this point in the *Groundwork*, Kant appears to have said nothing to rule out that possibility.<sup>9,10</sup> I said above that not every unqualified instruction expressed in the imperative mood is a categorical imperative. But Kant has not, at this stage, explained how we are to distinguish those which are not imperatives from those which are.<sup>11,12,13</sup>

## 2.2 The procession of the formulae

The problem of the connection of the FUL to later formulations of the categorical imperative concerns a much longer stretch of text. Henry Allison puts it in the following way. He argues – and I agree – that *Groundwork* II proceeds by way of ever-deeper characterizations of rational agency (Allison 2011: 182). He identifies the procession’s beginning as what he calls ‘intra-subjective universalizability’, the compatibility of an agent’s maxim with the same maxim considered as a universal law, and its conclusion as what he calls ‘inter-subjective universalizability’, which requires the endorsability of an agent’s maxim by all rational agents (Allison 2011: 196). The latter universalizability is immediately recognizable in the idea of the kingdom of ends: ‘a whole of all ends (of rational beings as ends in themselves, as well as the ends of its own that each of them may set for itself) in systematic connection’ (G, 4: 433). Call this *systematic universality*. Allison says that there is a ‘great problem’ in understanding whether such a procession is possible.

This is unsurprising. The requirement of the endorsability of one’s maxim by others seems to yield a considerably more confined space of permission than the requirements embodied in the universalizability procedure alone. Consider, for example, Christine Korsgaard’s cases of killing for revenge or out of hatred. She remarks that ‘[t]hese grim kinds of cases are managed without difficulty when using the Formula of Humanity, but it will be difficult to find any contradiction of the sort needed’ in applying the FUL (Korsgaard 1996: 100).<sup>14</sup> This is because the universality of killing for revenge or out of hatred would not make it any more difficult for me to kill out of revenge or out of hatred (as long as killing is all I will to do: as long as, for example, I do not will to kill and live to kill another day, since it is likely that, in the world of the universalized maxim, someone else will kill me before too long) and because I can consistently will the universality of killing for revenge and will never to be killed (as long as I will never to do anything avengeable).

The problem of the procession of the formulae is, therefore, that though some moral content can be identified in the FUL, it appears that later formulations have more. How, then, can they be not just equivalent to the FUL, but deeper expressions of what it expresses?

### 3. The type of systematic universality

The problem of the procession of the formulae is that of how the requirement to act only on maxims whose universality I can will – in the sense of maxim-universality – can be understood to be equivalent to the requirement that my action has systematic universality. It is natural for the reader of *Groundwork* II to want to be led from the FUL to systematic universality, since the FH and the FA follow the FUL in the text. But I want, for the moment, to set this matter aside and consider something that occurs in the second *Critique* which suggests that whatever the *Groundwork's* expository order might be, in the proper order of understanding, it is systematic universality that comes before the FUL. Later, I shall explain how I think this material should be brought to bear upon the interpretation of *Groundwork* II.

#### 3.1 The will of all

Consider a remark Kant makes in the discussion of the third theorem of the second *Critique*. In this remark, he distinguishes between the sense in which the wish or demand (*Verlangen*) for happiness is universal and the sense in which an action performed in observance of the moral law is universal. He says that, on account of this distinction, the latter cannot be explained in terms of the former:

It is, therefore, strange that intelligent men could have thought of passing off the desire for happiness as a universal *practical law* on the ground that the desire, and so too the *maxim* by which each makes this desire the determining ground of his will, is universal. For whereas elsewhere a universal law of nature makes everything harmonious, here, if one wanted to give the maxim the universality of a law, the most extreme opposite of harmony would follow, the worst conflict, and the complete annihilation of the maxim itself and its purpose. For then the will of all has not one and the same object but each has his own (his own welfare), which can indeed happen to accord with the purposes of others who are likewise pursuing their own but which is far from sufficing for a law because the exceptions that one is warranted in making upon occasion are endless and cannot be determinately embraced in a universal rule. (*CPrR*, 5: 28)

The wish for happiness is universal: everyone wants it.<sup>15</sup> But I wish for my happiness, and you wish for yours, and so we do not wish for ‘one and the same object’. The universality of the wish for happiness is maxim-universality. What Kant says in this passage entails that if the wish for happiness is not limited by another maxim, there can be, at most, a merely accidental harmony among wills. (It might happen that the best way for each of us to get what she wants is to allow, and perhaps even to help, others to get what they want.) The maxim whose universal adoption would make for a genuine harmony among wills would be one whose universal adoption would mean that all individual wills have ‘one and the same object’. If it is meaningful – as Kant implies here – to contrast such a maxim with the wish for happiness in the case of accidental harmony, then for all individual wills to have ‘one and the same object’ must be for them to stand in non-accidental practical agreement with one another, in the sense of Kant’s idea of a kingdom of ends:

By a *kingdom*, however, I understand the systematic union of several rational beings through common laws. Now, since laws determine ends according to their universal validity, it is possible – if one abstracts from the personal differences among rational beings, and likewise from all content of their private ends – to conceive a whole of all ends (of rational beings as ends in themselves, as well as the ends of its own that each of them may set for itself) in systematic connection, i.e. a kingdom of ends, which is possible according to the above principles. (G, 4: 433)

This condition is thus one characterized by what I have called systematic universality. In this condition, because the ends of each individual will are systematically unified with the ends of every other, there is a practical coherence, and thereby a volitional unity, among wills. There is a sense in which the will of all, having one and the same object, is one will. I shall henceforth use Kant's expression 'the will of all' in this sense.

Kant suggests that this notion of universality is the fundamental notion of universality in his ethics. For though the kingdom of ends is 'of course only an ideal' (G, 4: 433), nonetheless, '[m]orality . . . consists [*besteht*] in the relation of all action to the legislation through which alone a kingdom of ends is possible' (4: 434; emphasis and translation mine).<sup>16</sup> I want to understand this in the following way. Systematic universality is, on Kant's account, the perfection of the finite rational will.<sup>17</sup> 'Perfection' here means the perfect exercise of a capacity.<sup>18</sup> Morality consists in the perfect exercise of the rational will because the latter is a capacity whose principle of exercise is the moral law; our most fundamental understanding of a capacity is our understanding of what it is a capacity to do; and what a capacity is a capacity to do is the same as its perfect exercise. We may formulate the principle of the finite rational will in these terms: 'constitute a unified will with every other finite rational being'.

### 3.2 The topic of pure practical reason

Kant thinks, however, that there is a problem with formulations like that: they are not tractable in application to actions – or, equivalently, in application to the adoption of maxims – by an individual in a particular circumstance of action. According to his doctrine in the second *Critique*, it is on account of this problem that we need the FUL. The FUL is, according to Kant, the solution to a problem of tractability in the application of moral principles and concepts to actions: that is, in practical judgement.

In the section devoted to this topic in the second chapter of the second *Critique's* Analytic, the source of the intractability is identified as the distinction between freedom and nature:

[A] practical rule of pure reason . . . brings with it necessity with respect to the existence of an action and is thus a practical law, not a natural law through empirical grounds of determination but a law of freedom in accordance with which the will is to be determinable independently of anything empirical (merely through the representation of a law in general and its form); however, all cases of possible actions that occur can be only empirical, that is, belong to experience and nature; hence, it seems absurd to want to find in the sensible

world a case which, though as such it stands only under the law of nature, yet admits of the application to it of a law of freedom and to which there could be applied the supersensible idea of the morally good, which is to be exhibited in it *in concreto*. (CPrR, 5: 68–9)

The specific character of the problem depends upon why it should seem ‘absurd’ to find a natural object to which we may have grounds to apply the concept of the morally good. One obvious reading is that to have grounds to apply a concept to a natural object is ordinarily to recognize the applicability to it of other concepts. I may, perhaps, judge that something is a lump of gold on the strength of my recognition of its being yellow, malleable, and soluble in *aqua regia*. But because the concept of the morally good belongs to an entirely different order than do theoretical concepts, it does not enjoy such connections to such concepts as would underwrite its application.

There seems, however, to be another dimension to the absurdity, because Kant describes the concept of the morally good as an idea. An idea is a representation that surpasses (*übersteigt*) everything possible in the sensible world (A320/B377). It represents a perfection of which the sensible always falls short. Thus, nothing in the sensible world satisfies the descriptive content of any idea, including that of the morally good. But the satisfaction of a concept’s descriptive content by an object is, at least ordinarily, what makes that concept applicable to that object. There is, therefore, a problem about how the concept of the morally good might be applied to any particular proposed action.

Kant’s solution to this problem is the ‘typic of pure practical judgement’, a canon for the power of practical judgement, the power of the application of practical concepts, like that of the morally good, to proposed actions:

The rule of judgement under laws of pure practical reason is this: ask yourself whether, if the action you propose were to take place by a law of the nature of which you were yourself a part, you could indeed regard it as possible through your will. (CPrR, 5: 69)

Idiomatic differences notwithstanding, this is immediately recognizable as a variation on the FUL of the *Groundwork*. Indeed, just as in the *Groundwork*, our understanding of the formula is funded by an illustration that immediately follows its appearance:

Everyone does, in fact, appraise actions as morally good or evil by this rule. Thus one says: if *everyone* permitted himself to deceive when he believed it to be to his advantage, or considered himself authorized to shorten his life as soon as he was thoroughly weary of it, or looked with complete indifference on the need of others, and if you belonged to such an order of things, would you be in it with the assent of your will? . . . If the maxim of the action is not so constituted that it can stand the test as to the form of a law of nature in general, then it is morally impossible. (CPrR, 5: 69–70)

And near the end of *Groundwork* II, he says, in characterizing the FUL in contradistinction to other formulae, that



in moral *judging* it is better always to proceed by the strict method, and make the foundation the universal formula of the categorical imperative: *act according to the maxim that can make itself at the same time a universal law.* (G, 4: 436–7)

### 3.3 The priority of systematic universality to maxim-universality

Kant warns that we should not reckon ‘among concepts themselves that which belongs only to the *typic* of concepts’ (CPrR, 5: 70). The concept in question is that of the morally good; the *typic* is the FUL. He thereby suggests that the FUL, taken alone, is not a perspicuous expression of the content of that concept. It is merely the principle by which that concept is to be applied to objects. Systematic universality, on the other hand, as expressed by the idea of the kingdom of ends, is, as we have seen, that in which morality ‘consists’ (*besteht*). Systematic universality is the content of the concept of the morally good.

We have reason to doubt that the FUL really is a *typic* for the concept of the morally good. This is the same as the reason, discussed in §2.2, we have to doubt that the FUL is equivalent to the later formulae. Note, however, that what appears to show up in a reading of the *Groundwork* as a problem about a movement from the FUL to those principles and ideas here shows up as a problem about the movement from them to it. Part of what I am about to propose is that, appearances to the contrary, the latter is also the way in which the problem shows up in the *Groundwork*. This is because, on the reading I am about to offer, *Groundwork* II does not proceed from the sheer notion of the categorical imperative to the FUL, then from the FUL to the later formulae, in such a way as would depend upon a dubious equivalence between purpose-indifference, maxim-universality, and systematic universality. Rather, *Groundwork* II proceeds from a philosophically undeveloped understanding of systematic universality to a philosophically developed one: a procession that may be secure even if the FUL’s status as a *typic* of such universality is dubious. The difference between the *Groundwork* and the second *Critique*, according to this reading, is that in the former the FUL appears against the background of the philosophically undeveloped understanding and in the latter against the background of the philosophically developed one. In both cases, it belongs not ‘among concepts themselves’ but ‘only’ to their *typic*.

The significance of this claim depends on the specific character of the problem to which the *typic* is meant to be the solution. Suppose, for example, that when Kant says that without such a *typic* it is impossible to apply to a natural object a concept belonging to the order of freedom, he means that the terms in which the *typic* are formulated must not include those of the rational will *qua* rational (for such terms belong to the order of freedom exclusively). Then there is a sense in which notions figuring in the later formulae cannot be read into the FUL. For in that case, the FUL is intended to supply a criterion, formulable in natural terms alone, for the applicability of a concept belonging to the order of freedom. To read such a concept into it is to understand it as *not* formulable in natural terms alone.

It seems to me that this is part of what Kant means. But I do not wish to argue *this* point. I wish, rather, in what follows, to argue that even if this is what he means, and even if we have reason to doubt that the FUL really is a *typic* for the concept of the



morally good, understood in terms of systematic universality, nonetheless, we may judge the exposition of *Groundwork* II to be in good argumentative standing.

This leaves open, of course, that Kant is even better off than I suggest in this article. Perhaps the most promising plan of rescue for the FUL's status as a typic involves attention to what he means by 'nature' (here and in the Formula of the Law of Nature as expression of the FUL). Stephen Engstrom, for example, suggests that 'a universal law of nature can be a universal law of our *rational* nature ... and accordingly the consideration of whether a maxim can be willed as a universal law of nature can be understood as the attempt to conceive of it as a law that all rational beings necessarily follow out of their shared recognition of its validity' (Engstrom 2009: 161). Sven Nyholm argues for a similar position, appealing to Kant's distinction between nature considered *formaliter* and nature considered *materialiter* (Nyholm 2015: 290–4).

Now, it may be asked why, if there is a reading on offer according to which Kant does not have a problem, we should care how things stand even if he does. One of the aims of the interpretation of a philosophical text is to achieve precision about different problems an author has been alleged to face. To show that even if she faces one, she does not face another which has been thought to be the same as, or bound up with, the first, itself contributes to such an achievement. Since, however, I have not concealed my own sympathy with a reading on which the FUL is to be understood in natural terms, let me briefly indicate the ground of this sympathy. It is simply that in what is, in my view, the clearest application of the universalizability test, the case of the lying promise, it seems for all the world that what is imagined in the world of the universalized maxim is *not* 'shared recognition ... of validity' but rather everyone's being in the habit of making lying promises to repay a loan whenever in need of ready cash, and that Kant's point here is that once people have got wind of the practice, the next person who tries will not get away with it. '[N]o one would believe he was being promised anything, but would laugh about any such utterance, as vain pretense' (G, 4: 422). This is not the place to address the many interesting alternatives to this reading. (See, for some of these alternatives, Engstrom 2009: 196–209, Geiger 2015: 406–7, and Guyer 2019: 26–7.) I acknowledge that I may place too much emphasis on what a single passage appears to suggest against the weight of other evidence. But it is certainly not ridiculous to understand this passage to have the content and significance that I am inclined to attribute to it, as indeed many readers have done; this is enough to secure the interest of asking what else is true on the assumption that this reading is correct.

### 3.4 *The Groundwork and the second Critique*

Before I turn to the *Groundwork*, it is important for me to note, since I am about to apply to its interpretation a notion elaborated in the second *Critique*, that I do not here defend any particular position on perhaps the most vexed of the questions concerning the relation between these two books: that of whether the second *Critique*'s Fact of Reason (CPrR, 5: 30–1) constitutes a doctrinal departure from the content of the *Groundwork*.<sup>19</sup> Common to my discussion and to discussions of that question is the notion of common moral cognition, but it figures in the two cases in importantly different ways. The fact of reason is, roughly speaking, our appreciation of obligation and thereby of our freedom, such as might be described as an exercise of common

moral cognition. That the reader of Kant's practical works shares in this appreciation is meant to vouchsafe the objective reality of the concepts of moral obligation and of freedom insofar as they figure in those works. There is a question whether, in the *Groundwork*, and in particular in *Groundwork III*, Kant attempts to develop a different mode of establishing their reality, one which does not rely on any appreciation that might be described as an exercise of common moral cognition.

I am about to argue that in *Groundwork II*, no less than in *Groundwork I*, Kant relies on the reader's appreciation of what morality requires on the assumption of the objective reality of the concept of moral obligation (and, thereby, of freedom). *Groundwork II* is, I shall suggest, devoted to a philosophical elucidation of the nature of a capacity first characterized as a capacity the principle of whose operation is the moral law. This is the capacity that in common moral cognition we presuppose we have, though that cognition does not include the philosophically perspicuous elucidation of its nature which it is the task of *Groundwork II* to provide. *Groundwork II* is not, however, addressed to the question of whether we human beings really possess this capacity, as, in common moral cognition, we presuppose we do. That is a matter for *Groundwork III* and, in the second *Critique*, for the fact of reason. Thus, my topic is not Kant's appeal in the second *Critique* to common moral cognition to secure the reality of a concept – that of the capacity he describes in *Groundwork II* – but his appeal to common moral cognition to supply the raw material needed for a philosophical elucidation of the content of that concept, on the assumption of its reality.

#### 4. Love thy neighbour

*Groundwork II*, I shall suggest, proceeds from a philosophically undeveloped understanding of systematic universality to a philosophically developed one. But for the principal expression of the first of these understandings in the *Groundwork*, we must turn to *Groundwork I*. Having found it in *Groundwork I*, I shall explain why and how we can take it to be a point of departure in *Groundwork II*.

*Groundwork I* begins with a motivational analysis of the notion of a good will, the concept of which 'already dwells in natural sound understanding and needs not so much to be taught as only to be clarified' (G, 4: 397).<sup>20</sup> A finite will is good insofar as it is exercised 'from duty' (*aus Pflicht*) (4: 398). To act from duty is, Kant stresses, not merely to act 'in conformity with duty' (4: 398). It is possible to act in conformity with duty but on account of an inclination towards the action in question, whether immediate or mediated. In illustrating these various motivational possibilities, Kant supposes that all will grant that duty demands various things: for example, that 'a shopkeeper not overcharge an inexperienced customer' (4: 397), that one 'preserve one's life' (4: 397), that one 'be beneficent where one can' (4: 398), and that one 'assure one's own happiness' (4: 399).

But these examples are not merely illustrative. They are, I suggest, essential to the reader's getting a grip on what Kant is talking about. To see this, set all possible examples aside, and ask: what is the concept of an action from duty? What we can say is that it is an action that is not performed on account of inclination, immediate or mediated. (We might try saying that it is an action performed with a good will, but it is meant to be the concept of duty which clarifies that of the good will, not the other way about (G, 4: 397). And even if it were the other way about, the reader would

require examples to get a grip on the notion of the good will.) So one mode in which actions are intelligible – that is, motivation by inclination, immediate or mediated – is unavailable. How else might an action be intelligible? Some philosophers have, of course, thought that this is the only way in which an action can be intelligible. But Kant, his examples at the ready, does not even leave the space for reflection at this level of generality. It is, he says, possible to do such things as to decline to overcharge inexperienced customers and to help others where one can simply *because these are the right things to do*. The reader's initial grip on the more generic notion of motivation by something other than inclination depends on her grip on the more specific notion of doing something because it is right. And her grip on that notion is essentially funded by her understanding of these examples.

Kant's examples are, I think, enough to elicit in the reader the content of common moral cognition which, as I argue below, he will bring to philosophical explicitness in *Groundwork* II. But it is instructive to consider another kind of remark which he may intend to serve a similar function, and which he makes having developed his four examples:

It is undoubtedly in this way, again, that we are to understand the passages from scripture in which we are commanded to love our neighbour, even our enemy. For, love as an inclination cannot be commanded, but beneficence from duty – even though no inclination impels us to it and, indeed, natural and unconquerable aversion opposes it – is *practical* and not *pathological* love, which lies in the will and not in the propensity of feeling, in principles of action and not in melting sympathy; and it alone can be commanded. (*G*, 4: 399)

This remark may seem to serve as an elaboration on one of the examples already given, that of being beneficent where one can. And beneficence certainly belongs to what Kant has in mind here. But remarks he makes on the scriptural injunction elsewhere suggest that he may comprehend more in it than this. Kant is recorded as saying, for example, that '[t]he love for others can be considered in its generality, so to that extent it rests on this, that our ends coincide with those of others in such a way that they are able to co-exist together according to the universal rule of duty' (*Eth-Vigil*, 27: 673). Even more strikingly, he is recorded as describing a 'universal will' which 'consists in the universal end of all men, and is called love for others, the principle of well-wishing' (27: 541). I wish to suggest that the structure of systematic universality is anticipated in this discussion of the scriptural injunction to practical love.

Now, it may seem that though 'love thy neighbor' boasts positive content, it boasts too much: that is, it is too specific to be an indication of what Kant means by 'action from duty'. There are at least two issues here. The first is that, as mentioned above, in the *Groundwork* passage Kant seems to identify practical love with beneficence from duty, and yet there are, according to Kant, duties to others other than beneficence. To be sure, in the *Metaphysik der Sitten Vigilantius* passages I have already quoted, it seems that Kant is recorded as understanding love to be more general than beneficence in this narrow sense, since he is happy to identify it with the universal will. And this is also suggested by his discussion of the scriptural injunction in the second *Critique*. There he calls the commandment 'Love God above all, and your neighbour as yourself'

the ‘kernel of all laws’ (CPrR, 5: 83). He says that only ‘practical love’ is understood in this commandment and immediately explains what he means as follows: ‘[t]o love God means, in this sense, to do what He commands *gladly*; to love one’s neighbour means to practice all duties toward him *gladly*’ (5: 83). Similarly, in the *Religion Within the Boundaries of Mere Reason*, we are told that the ‘founder of the one true church’

sums up all duties (1) into one *universal* rule (which includes the internal as well as the external moral relation of human beings), namely, Do your duty from no other incentive except the unmediated appreciation of duty itself, i.e. love God (the Legislator of all duties) above all else; (2) and into a *particular* rule, one namely that concerns the human being’s external relation to other human beings as universal duty, Love every one as yourself, i.e. promote his welfare from an unmediated goodwill, one not derived from selfish incentives. (R, 6: 160–1)

Since these rules are meant to sum up all duties, we cannot understand Kant’s gloss on the second rule in terms of welfare as restricted to beneficence in the narrow sense.

But can the common moral cognizer of *Groundwork* I be expected to anticipate this? I think so, but in a somewhat complicated way. To see it, consider that for Kant, even beneficence in the narrow sense is not principally a matter of addressing needs, such as any needy being might have. Rather, it is a matter of an orientation towards others’ ends, such as only a practically rational being can have, an orientation that is part and parcel of treating them as ends in themselves:<sup>21</sup>

Now, humanity could indeed subsist if no one contributed anything to the happiness of others while not intentionally detracting anything from it; but this is still only a negative and not positive agreement with *humanity, as an end in itself*, if everyone does not also try, as far as he can, to advance the ends of others. For if that representation is to have its *full* effect in me, the ends of a subject that is an end in itself must, as much as possible, also be *my* ends. (G, 4: 430)

Duties of beneficence are grounded in this, that others are ends in themselves. But that status is equally what grounds, for example, the duty not to make lying promises (G, 4: 429–30). To acknowledge the immediate normative significance of the ends of others, and to act in accordance with this acknowledgment, is to love thy neighbour in the sense of the second *Critique* and the *Religion*. So, according to Kant, once we understand the ground of duties of beneficence, we have everything we need for an understanding of perfect duties too: we have everything we need to understand what, according to Kant, ‘love thy neighbor’ really means. It means to treat others as ends in themselves.

In light of this observation, I want to suggest that even if Kant indeed intends the reader to understand the ‘love thy neighbor’ passage of *Groundwork* I, unlike passages on the same topic in the second *Critique* and the *Religion*, as concerning beneficence in the narrow sense, he nonetheless believes that the common moral cognizer will, upon philosophical reflection – upon the kind of reflection he undertakes in *Groundwork* II – come to understand that the ground of duties of beneficence in the narrow sense equally grounds perfect duties to others and that these are not just two kinds of duty,

one atop the other, but rather expressions of two aspects of one thing: the status of others as ends in themselves. In that case, the common moral cognizer can come to recognize that the formula 'love thy neighbor' is better suited to the expression of the command that we acknowledge that one thing than to the expression of the command that we acknowledge just one of its two aspects.

The second issue in connection with 'love thy neighbor' is that one has duties to oneself and one is not one's neighbour. Indeed, the passage from the *Religion* I have just quoted suggests that duties to oneself ('the internal . . . moral relation of human beings') are not covered by 'love thy neighbor' but are covered by the more general 'love God'. But it is possible that this is not as much of a problem as it might first appear. This is a topic that requires its own sustained treatment, but let me gesture towards what I mean. In the *Moralphilosophie Collins*, Kant is recorded as saying that 'self-regarding duties are the supreme condition and *principium* of all morality' (Eth-Collins, 27: 344) and that they are the 'condition under which' duties to others 'can be observed' (27: 341). This idea is not absent from the Critical published texts. Consider Kant's explanations of the various duties to oneself in terms of one's status as a practically rational being or in terms of one's readiness to fulfil obligations. Self-mutilation (*MM*, 6: 422), gluttony (6: 427), drunkenness (6: 427), and avarice (6: 432) all make it more difficult to act in the fulfilment of one's duties. Suicide would be 'withdrawal' from all obligations (6: 422). Then there are the prohibitions of recreational sex (6: 425), lying (6: 429), and servility (6: 435). Kant claims that these are offences against one's personality. But one's personality is precisely what constitutes one's fitness to enter into moral relationships with others.

My suggestion, in light of this, is the following. It is not that duties to oneself are indirect, as though preserving one's personality and developing one's capacities are merely instrumentally directed towards moral service towards others.<sup>22</sup> Indeed, there would seem to be an incoherence in the idea of preserving one's personality solely to be of any kind of service to others. Rather, it is the good of the finite rational being to live with other finite rational beings in the kingdom of ends, in which each stands in reciprocal moral relationships with all others. Standing in such relationships is in part a matter of helping each other to realize ends, thereby to be happy. But Kant thinks that there are some things one must do, and some things one must not do, in order to realize and preserve one's nature (or standing, if you like) as the kind of being who can enter into such relationships: as the kind of being who fulfils duties to others, who is expected and, indeed, counted on to fulfil these duties, and who expects and counts on the fulfilment by others of their duties towards her. These conditions are articulated in the various duties to oneself. (This is to say nothing, of course, about whether Kant has succeeded in identifying these conditions.) This is, to be sure, a hypothesis, but I submit that it fits Kant's arguments for the particular duties to self in the system of the *Metaphysics of Morals*, is true to the remarks from the *Moralphilosophie Collins*, and does not involve the error of thinking these duties indirect.

To return, finally, to the formula 'love thy neighbor': if what I have said is on the right track, then we can say that duties to oneself express conditions on the possibility of practically loving relationships (in the general sense of the formula). Not every being capable of entering into such relationships stands in readiness to do so: but because to be in such relationships is the good of such a being, she must fulfil the

conditions that constitute this readiness. To fulfil one's duties to oneself is to make and keep oneself fit for love: fit, that is, to love one's neighbours and to be loved by them. Like the true ground of beneficence, this is not understood, in all its articulation, in common moral cognition. But again, Kant is confident that philosophical reflection on the formula will expose and clarify it.

And even if none of what I have said about 'love thy neighbor' in connection with perfect duties to others and duties to oneself is tenable, still Kant can rely on his four examples, in which duties of all three kinds are accounted for. In what follows, I shall, accordingly, use 'love thy neighbor' and similar expressions to capture what Kant elicits from common moral cognition in *Groundwork I*; however, much of this content is to be comprehended in his paragraph on the scriptural injunction as such.

## 5. The capacity whose exercise is action from duty

### 5.1 Categorical imperatives

I have suggested that philosophically undeveloped anticipations of systematic universality fund the reader's understanding of the concept of duty in *Groundwork I*. In *Groundwork II*, Kant wishes to 'follow and present distinctly the practical faculty of reason, from its general rules of determination to the point where the concept of duty arises from it' (G, 4: 412). There are two ways in which such a task might be understood. Allison gives expression to the first of these in saying that *Groundwork II* 'starts not with our ordinary, pre-philosophical moral convictions, but with the abstract philosophical concept of a rational agent as such' (Allison 2011: 35; emphasis mine). Thus, Allison suggests, the two sections make different and independently intelligible approaches to the concept of duty: *Groundwork I* begins with our ordinary, pre-philosophical moral convictions and *Groundwork II* with an abstract philosophical concept of a rational agent as such. *Groundwork II* makes, in this sense, a 'fresh start' (Allison 2011: 35).

But Kant does not say, in the quoted characterization of his task, that the 'general rules of determination' of practical reason must be, or even seem to be, intelligible independently of the concept of duty which 'arises from it'. To learn something's source – to witness it 'arising' from this source – does not require that one first forget what the thing is. And not only is there no indication in *Groundwork II* to the effect that we should abandon the understanding of duty which rests upon common moral cognition: there are explicit indications to the contrary at the end of *Groundwork I*. There, Kant writes that it is 'common human reason' itself – whose cognition has been elucidated in *Groundwork I* – which must 'take a step into the field of *practical philosophy*, in order to receive there intelligence and distinct instruction regarding the source of this principle and its correct determination' (G, 4: 405; emphasis in the original). Common human reason cannot be denied entry to the classroom in which it is to receive instruction.<sup>23</sup> This is not to say that *Groundwork II* is not of a character importantly different from that of *Groundwork I*. As Jeanine Grenberg puts it, the 'rational tools' in the field of practical philosophy are distinctively philosophical: they are, in this sense, not the tools of the common moral cognizer (Grenberg 2013: 91). But that to which the tools are applied is what, as depicted by *Groundwork I*, common moral cognition yields.<sup>24</sup>

The other and, I think, better understanding of Kant's task is that he wishes to supply a description of action from duty as the exercise of a rational capacity and thereby wishes to elucidate the structure of the capacity, which he calls 'practical reason', whose exercise it is.<sup>25</sup> This is in keeping with the general description of the method given at the end of the *Groundwork's* Preface:

In this work, I have adopted the method that is, I believe, most fitting if one wants to take one's route analytically from common cognition to the determination of its supreme principle and in turn synthetically from the examination of this principle and its sources back to common cognition, in which we find it used. (G, 4: 392)

The capacity in question is the source of the principle and is from the beginning identified as this source. (Kant here announces no intention to ground the principle of common moral cognition in a capacity identifiable and characterizable in entirely independent terms.)

In *Groundwork* I, having funded the reader's understanding of the notion of duty with his examples and his discussion of practical love, he argues for two 'propositions', one of which – the 'third proposition' – is that 'duty is the necessity of an action from respect for law' (G, 4: 400).<sup>26</sup> It is with formulations in terms of law or laws that he begins his elucidation of practical reason in *Groundwork* II:

Only a rational being has the capacity to act *in accordance with the representation* of laws, that is, in accordance with principles, or has a *will*. Since *reason* is required for the derivation of actions from laws, the will is nothing other than practical reason. (G, 4: 412)<sup>27</sup>

He then tells us that for the finite rational will, principles are expressed as imperatives because this kind of will does not inevitably (*unausbleiblich*) satisfy them (G, 4: 412–13). And, finally, he says that

all imperatives command either *hypothetically* or *categorically*. The former represent the practical necessity of a possible action as a means to achieving something else that one wills (or that it is at least possible for one to will). The categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end. (G, 4: 414)

Thus does the notion of a categorical imperative emerge from the investigation of the concept of a capacity to act from duty.

Note that the notion of a categorical imperative, as it appears here, has something important in common with the notion of action from duty as it appears in *Groundwork* I. If we set aside all material for understanding made available by common moral cognition, all we can find in this is a negative notion: a categorical imperative is one which is not qualified by an (actual or possible) end. This is what I meant in §§1–2 by 'the sheer notion of a categorical imperative': this notion is exhausted by purpose-indifference. If we suppose that the sheer notion of a categorical imperative is all we can understand in Kant's description of a categorical imperative, then we cannot make



sense of why, following a discussion of hypothetical imperatives, he straightaway declares that the categorical imperative ‘may be called the imperative of **morality**’ (G, 4: 416).

Of course, we might make sense of this by declaring Kant’s exposition sloppy or his argument unsuccessful. But if it is wrong to set aside all material for understanding made available by common moral cognition, then it is wrong to make such declarations (at least on the ground that purpose-indifference is not morality). The notion of a categorical imperative at work here is not just the notion of an imperative that is not qualified by an (actual or possible) end. It is, as we might put it, the notion of an imperative that is not qualified by an (actual or possible) end, *like the injunction to practical love*: what we lean on in common moral cognition to understand action from duty in *Groundwork* I we lean on here to understand what a categorical imperative is. This characterization is more philosophically developed than any available in *Groundwork* I because it is formulated in part in terms of Kant’s account of finite practical reason. But its development is not complete, nor will it be until the end of *Groundwork* II, when it is explicitly expressed in terms of systematic universality.<sup>28</sup>

## 5.2 The *FUL* again

With all of this in hand, I can supply the promised sketch of the expository structure of *Groundwork* II. My principal claim is, as I have said, that Kant proceeds from a philosophically undeveloped understanding of systematic universality – as expressed, for example, by the words ‘love thy neighbor’ – to a philosophically developed one in the idea of the kingdom of ends. I have also said that the topic of *Groundwork* II is that of the capacity whose exercise is action from duty, initially understood in terms of the philosophically undeveloped understanding of systematic universality. Earlier, in §3.1, I said that the idea of the kingdom of ends is a philosophically perspicuous representation of the perfect exercise of this capacity and that we understand what a capacity is in terms of its perfect exercise. So much for the great sweep of *Groundwork* II, as proceeding from a philosophically undeveloped understanding of an act to a philosophically developed understanding of that act as the perfect exercise of a capacity. But this article began with a statement of some serious problems for the *FUL*, and it is time to turn to them.

The first problem is that of the gap in the derivation. Recall the moment of derivation:

[W]hen I think of a *categorical* imperative I know at once what it contains. For since besides the law the imperative contains only the necessity of the maxim to conform with this law, whereas the law contains no condition to which it was limited, nothing is left but the universality of a law as such, with which the maxim of the action ought to conform, and it is this conformity alone that the imperative actually represents as necessary.

There is therefore only a single categorical imperative, and it is this: *act only according to that maxim through which you can at the same time will that it become a universal law.* (G, 4: 421)

The first paragraph seems to express nothing more than purpose-indifference, with no indication of how to tell whether a candidate purpose-indifferent principle really is one. The second, read in light of the procedure for assessing maxims which Kant immediately illustrates, identifies the one and only purpose-indifferent principle as the requirement that one's maxims have maxim-universality. But if we understand Kant's use of 'categorical imperative' in the way I have suggested, then the first paragraph expresses a partially developed understanding of systematic universality. It thereby expresses an indication of what the one and only purpose-indifferent principle is: constitute a unified will with every other rational being. Given that our understanding of the second paragraph is funded by the illustrative examples that follow, Kant's claim here is that a requirement formulated in terms of maxim-universality is equivalent to the requirement that one's action have systematic universality (as partially understood).

As I said in §3.3, we have reason to doubt that this claim is true. But this is not because Kant has not given us any way of telling whether a candidate purpose-indifferent principle really is one. It is because it is difficult to see how maxim-universality could be equivalent to systematic universality. When we read the second *Critique*, the problem shows up as one about the equivalence of the former to the latter understood in a philosophically developed way: this is the problem of whether the FUL is really a 'typic' for systematic universality. When we read *Groundwork* II, the problem shows up as one about the equivalence of the former to the latter understood in a relatively undeveloped way. In the end, there is only one problem here, common to the two texts, and it is not that of the gap in the derivation. I think, however, that this is a serious problem, and, as I suggested in §3.3, believe that for this reason, it is wrong to read later formulations and principles back into the FUL.<sup>29,30</sup>

What of the problem of the procession of the formulae? First, if what I have just said is right, then there is indeed a serious problem about the equivalence of the FUL to the other formulations, a problem which is ultimately about whether the FUL is the principle of practical judgement: the principle according to which the bearer of practical reason self-consciously applies its concepts to proposals for action here and now. But, second, if what I have said since the beginning of §4 is right, this need not vitiate the expository sweep of *Groundwork* II, from a philosophically undeveloped notion of systematic universality to a philosophically developed one.

It need not: but does it, in fact, not? Can the FUL's problems be sequestered? An important indication that they can is that, though the FUL appears in the text before the FH and FA, Kant's exposition does not proceed directly from the FUL to the FH. Here is, in outline, his approach to the FH. First, he asks:

is it a necessary law *for all rational beings* always to appraise their actions in accordance with such maxims as they themselves could will to serve as universal laws? If there is such a law, then it must already be connected (completely a priori) with the concept of the will of a rational being as such. (G, 4: 426)

Thus, he suggests the task of discovering the *a priori* connection of the FUL to the concept of the will of a rational being as such. This necessitates a return to that

concept, the first moment of whose explicit articulation occurred at *G*, 4: 412, a moment now repeated, before Kant adds a further reflection:

The will is thought as a capacity to determine itself to acting in conformity with the *representation of certain laws*. And such a capacity can be found only in rational beings. Now, what serves the will as the objective ground of its self-determination is an end, and this, if it is given by reason alone, must hold equally for all rational beings. (*G*, 4: 427; emphasis in the original)

He then argues that there is an end given by reason alone: that is, humanity (*G*, 4: 428). The point of departure for this argument is the same as that for the stretch of text leading up to the *FUL*. And this is to say that it is not the *FUL*. Rather, it is the idea of a capacity to act in accordance with, or in conformity with, the representation of laws, where this notion is understood, in accordance with Kant's discussion in *Groundwork I*, in terms of action from duty, and where that notion is understood with the aid of common moral cognition. This suggests that, unless it makes a surprise argumentative appearance after this point – and I submit that it does not – its problems can indeed be sequestered.<sup>31</sup>

## 6. Conclusion

I began this article with the claim that the problems of the gap in the derivation and the procession of the formulae should be understood in terms of the problem of how three notions of universality figure in *Groundwork II*: purpose-indifference, as expressed by the sheer notion of a categorical imperative, a principle insensitive to an agent's given purposes; maxim-universality, the content of which notion is funded by Kant's illustrations of the universalizability procedure; and systematic universality, a practical coherence among wills whose complete actuality is depicted by the idea of a kingdom of ends, in which every individual will shares in the one will of all.

The solution I have offered is this. Purpose-indifference does not figure in *Groundwork II* in its own right; it is an abstraction from what Kant understands in his notion of a categorical imperative. Systematic universality is *the* topic of *Groundwork II*, which proceeds from the philosophically undeveloped understanding of this universality elicited in the reader by *Groundwork I* – 'love thy neighbor' – to a philosophically developed understanding of the complete actuality of such universality as the perfect exercise of practical reason, as depicted in the idea of the kingdom of ends. Maxim-universality is the notion in terms of which Kant formulates the principle of practical judgement: that is, the principle by which, as he would have it, the self-conscious possessor of this capacity exercises it in particular cases, remote from its perfect actualization in the kingdom of ends. This formulation has serious problems: we have reason to doubt that it really is the principle of practical judgement. But, as far as I can see, this leaves systematic universality, and its role in *Groundwork II*, from beginning to end, untouched.

Universality is not just an apparently central theme of Kant's practical philosophy. It is *the* central theme of that philosophy. But in saying this, I do not mean the universality of the *FUL*. For we are not to reckon 'among concepts themselves that which belongs only to the *typic* of concepts' (*CPrR*, 5: 70). The true universality, the

universality which belongs ‘among concepts themselves’, is, from beginning to end, systematic universality, that of the will of all, whereby we are many wills only insofar as we are one.

**Acknowledgements.** I extend thanks, for their generous feedback on this article, to David Finkelstein, Chandler Hatch, Jed Lewinsohn, Michael Pendlebury, and Mary Tjiattas. Thanks to Jacob Rosen for early peripatetic conversation in the Allegheny Cemetery and to Jeremy David Fix for having instigated my consideration of these topics and for many rounds of comments and discussion. And thanks to two anonymous referees who provided the opportunity for substantial improvements.

## Notes

1 Citations of Kant’s works (except of the *Critique of Pure Reason*, in citing which I supply A/B pagination) are to the Academy edition. Unless otherwise noted, translations are based on those of *The Cambridge Edition of the Works of Immanuel Kant* (Kant 1992–2016), ed. Paul Guyer and Allen W Wood. I use the following abbreviations: A/B = (1781/1787) *Kritik der reinen Vernunft* (*Critique of Pure Reason*); Eth-Collins = (1784–5) *Moralphilosophie* Collins (Moral Philosophy Collins); G = (1785) *Grundlegung zur Metaphysik der Sitten* (*Groundwork of the Metaphysics of Morals*); CPrR = (1788) *Kritik der praktischen Vernunft* (*Critique of Practical Reason*); FI = (1789) *Erste Einleitung in die Kritik der Urteilskraft* (*First Introduction to the Critique of the Power of Judgement*); CPJ = (1790) *Kritik der Urteilskraft* (*Critique of the Power of Judgement*); R = (1792) *Die Religion innerhalb der Grenzen der bloßen Vernunft* (*Religion Within the Boundaries of Mere Reason*); Eth-Vigil = (1793) *Metaphysik der Sitten Vigilantius* (*Metaphysics of Morals Vigilantius*); MM = (1797) *Metaphysik der Sitten* (*Metaphysics of Morals*); Refl = *Reflexionen*.

2 See, for example, Wood (1999).

3 See, for example, Engstrom (2009), O’Neill (1989), and Reath (2006). Reath (2015: 211) puts the general idea as follows: ‘a procedure that applies the bare idea of universalizability to maxims would not generate interesting moral content. (Perhaps deceptive promising but not much else would be ruled out.) So one might argue that in order to address the ‘content problem’ ... FUL must be understood either to implicitly contain, or to be applied in conjunction with, ideas expressed by the other formulas of the categorical imperative, such as that of persons as agents with autonomy who are ends in themselves, the ideal of moral community expressed by the realm of ends, etc.’

4 I offer some elaboration of this criticism at the end of §3.3.

5 In this I take myself to contribute to what Karl Schafer has called a ‘capacities-first’ approach to Kant’s Critical philosophy (Schafer 2019, 2021).

6 This name for the problem appears in the title of Allison (1991).

7 Some commentators do not take the content of the FUL to be given by the instructions for the universalizability test. For example, Allison (2011: 140) argues that the FUL merely expresses in the imperative mood the notion of a finite will: of a ‘capacity to act according to representation of laws’ (G, 4: 412). In effect, Allison says, it shows us what it is for a finite rational being to exercise this capacity in directing such a being to such an exercise; its import is exhausted by its function of expressing the contrast between it, a prescription, and the more generic notion of conformity to universal law, which applies also to the species of rational will for which digression from the law is impossible. But my aim is not to show that the FUL itself must be understood as I understand it: as expressing not merely a prescription of universal conformity to law but also what is given by the instructions for the universalizability test. My topic is that which I understand the FUL to express, which every reader admits is somehow expressed in Kant’s practical philosophy, perhaps by the so-called Formula of the Law of Nature (FLN), perhaps by his instructions for the universalizability test.

8 This is an example of what is often called the test of ‘contradiction in conception’; another kind of contradiction meant to be prohibited by the FUL is the so-called ‘contradiction in the will’ (G, 4: 422–3). Though the character of the contradiction is different – in the former case, it is of the proposed maxim with itself universalized; in the latter case, it is of the universalized maxim with some materially different determination of the will – the kind of universality, my topic here, is the same. For a discussion of the character of the contradiction, see O’Neill (1989), Korsgaard (1996: 101–2), Kleingeld (2017, 2021), and Timmermann (2018).

9 This is not how Kant's readers usually frame the problem, but it captures much of what underlies their dissatisfaction with his text at this point. Thus, for example, Allen Wood's comments on the derivation can be understood as a specification of this problem. He writes that 'it does not follow from the mere concept of a categorical imperative that *the will of a rational being* – what a rational being wills or can consistently will – has any role to play in determining the content of universal laws' (Wood 1999: 81). It is apparently compatible with the mere concept of a categorical imperative that the content of universal laws – of the purpose-indifferent laws which govern rational action – need not admit of systematization only under the principle of what a rational being can consistently will, but admits also of systematization under, say, the principle of what Lytton Strachey would do.

10 Indeed, it might be argued on Kant's behalf that this principle, though it masquerades as insensitive to given purposes, is not really insensitive to them. Allison says this of the principle of rational egoism: '[s]uch a policy is deemed reasonable in the first place only because of certain presupposed ends' (Allison 1991: 12). Perhaps that can be shown, but Kant does not attempt any such demonstration at the point at which the 'gap in the derivation' is thought to appear.

11 This suggestion is in some respects similar to Thomas Hill's claim that the derivation of the categorical imperative is Kant's (in his view failed) attempt to move from a conception of practical reason as 'enabl[ing] us to reach conclusions on some basis other than that they get us what we most want' to the identification of the FUL as expressive of this alternative basis (Hill 1992: 121–2). Like Hill's, this reading takes the material before the paragraph break at G, 4: 421, to express something negative, though what is negated is different: for Hill, the point is that the imperative does something other than guide us in the fulfilment of 'what we most want'; according to this reading, the point is that it does something other than guide us in the fulfilment of some purpose we happen to have.

12 It might be suggested that it is a mistake to read Kant's characterizations of what the categorical imperative 'would be' (at, e.g. G, 4: 414) as intelligible independently of the FUL. But Kant, in the passage in which the gap has been thought to occur, proceeds from a characterization of a categorical imperative which he has already introduced to the conclusion that there is only one thing satisfying that characterization and which might be expressed by the FUL. This is not to say that, according to Kant, there might be, or might have been, other categorical imperatives. Indeed, he says that 'there is only a single categorical imperative' (G, 4: 421). But it is to say that he does not regard it as obvious from the outset that the FUL is the expression of the only thing satisfying his initial characterization of categorical imperatives as such: for if he did, he would not seek to extract the former from the latter.

13 To be sure, different apparent problems have fallen under the title of the gap. Wood, as we have seen (n. 9), objects that in the notion of conformity to universal law as such, the notion of universality might be taken in any number of ways; it need not implicate the notion at work in the FUL. The problem of the gap is often, however, associated with Aune (1979), in whose book I find a different problem. Aune does not seem to doubt that the notion of conformity to universal law as such has moral content; his problem is that this notion does not tell us what the universal laws are, whereas the FUL does (or purports to do), so that the latter is of 'practical import' in a way in which the former is not (pp. 29–30). Since I shall argue below that the notion of universal law indeed has (notwithstanding the appearances at issue in this section) moral content, I shall eventually agree with Aune on this point. But I do not think that Aune explains why he is happy to allow that what precedes the FUL has moral content, whereas I shall insist on identifying such content's expository source.

14 I assume for the purposes of this article that the later formulations are equivalent.

15 'Happiness' is 'the general name for subjective determining grounds' (CPrR, 5: 25), rather than the name of some particular material end. What I am about to say about the will of all suggests that, in a kingdom of ends, each individual will is orientated towards her own happiness – the systematic satisfaction of her own ends, explicitly represented (G, 4: 418; CJ, 5: 430) – but in a way which is qualified or limited by every other will. See, for discussion of the internal connection of happiness to the individual will as such, Engstrom (2009: 81f.), and, for discussion of happiness as something other than a 'first order end that you pursue directly', Hills (2009).

16 This agrees with such characterizations as these, dated to the 1770s: '[m]orality consists in the relation of free actions with the laws (conditions) of the universal will, either of humanity or of human beings' (Refl 6950, 19: 212); '[m]oral philosophy is the science of ends insofar as they are determined through pure reason. Or of the unity of all ends . . . of rational beings' (Refl 6820, 19: 172).

- 17 The will is, according to Kant, a desiderative capacity (*Begehrungsvermögen*): ‘a being’s capacity to be by means of its representations the cause of the actuality of the objects of these representations’ (*CPrR*, 5: 8n.; see also *Fl*, 20: 206; *CJ*, 5: 177). It is, in particular, a rational desiderative capacity. (For discussion, see Fix and Pendlebury (forthcoming).) The will is, therefore, a capacity to act; its perfect exercise is completely successful action, where success is measured by its own principle. It may be that the principle of the rational will should be so understood as to embrace the Universal Principle of Right (*MM*, 6: 230). (For discussion, see, e.g. Ripstein 2009; Willaschek 2009; Guyer 2016; Herman 2021.) If so, I mean here only its perfection as a capacity for internal freedom or the aspect of its perfection which corresponds to this.
- 18 If the kingdom of ends is the perfection of the rational will, then the perfection of an individual rational will depends on the maxims adopted by other rational wills. This is not to say that you are at fault if you do everything in your power to actualize the kingdom of ends but do not enjoy the cooperation of others. In that case, you are not at fault, though your will is exercised imperfectly.
- 19 For an historical survey of discussion of this question, see Ware (2021: 71–99).
- 20 For discussion of this motivational analysis, see Korsgaard (1996: 55f.).
- 21 For discussion, see Herman (2010, 2021: 134–45). This is not to say that ends are not in part grounded in needs: see *CPrR*, 5: 25, for a discussion of the significance of the neediness of finite rational beings.
- 22 For discussion, see Timmermann (2006: 506–9). Unlike Timmermann, I am suggesting that duties to oneself can ‘be established on the grounds of our moral interaction with others’ (Timmermann 2006: 509), though not in the terms of the notion of indirect duty.
- 23 In this article, I focus on one lesson which common human reason may hope to learn from the instruction constituted by the metaphysics of morals: it may hope to gain a philosophically articulated understanding of the nature of the capacity which, in common moral cognition, we presuppose we possess. It should be noted that, in Kant’s view, such instruction is by no means limited to this lesson. For example, with philosophical articulation comes an opportunity to address common moral errors, as noted by Sticker (2016: 87). And Kant not only acknowledges that there are common moral errors but seeks to account for an especially important kind of moral error in terms of his notion of a *natural dialectic* of human reason (*G*, 4: 405). Because he believes that moral philosophy, of the kind developed in *Groundwork II*, is a necessary response to such a dialectic, the story I tell of the relation of *Groundwork II* to common moral cognition is incomplete. (See Callanan [2019] for discussion of moral philosophy as response to the natural dialectic and the emergence of Kant’s conception of the necessity of such a response in his engagement with the thought of Rousseau.) My aim is to tell only as much of the story as is required to address the argumentative problems described in §2.
- 24 Grenberg seems to suggest that there would be no need for the application of philosophical tools were common moral cognition not ‘conflicted’ or ‘corrupted’: were the common moral agent not given, that is, to ‘lose her ability to distinguish categorical demands from hypothetical ones’ (Grenberg 2013: 90). Here I disagree. To be sure, the reason common human reason is ‘impelled’ (*angetrieben*) to exit its sphere is its tendency towards conflict and corruption. But it does not follow that, were there no such tendency, philosophical tools would not be required for the perspicuous characterization of the capacity whose exercise is action from duty. (Kant may think that in that case there would be no need, grounded in practical life itself, for such a characterization, but that is another matter.)
- 25 This may be compared to Reath’s description of ‘a Kantian view’ which ‘does not attempt to derive morality from a morally neutral starting point’ but whose ‘general structure is that it ties the content of a moral conception to a more general set of ideals – of the person, of agency, and of rationality – which, while applying widely and providing some kind of independent perspective on morality, need not be empty of moral content’ (Reath 2006: 68). I am, however, suggesting something more determinate than this: the Kantian view finds the ground of the common understanding’s moral conception in the structure of the capacity whose everyday self-consciousness that moral conception is. (For discussion of an analogous structure in Kant’s theoretical philosophy, see Pendlebury [2022].)
- 26 See Allison (2011: 121–35) for discussion of the problem of the three propositions.
- 27 It is controversial what sort of thing Kant means by ‘laws’ and ‘principles’ here. (See, e.g. Timmermann [2006: 60] and Allison [2011: 152–3].) On the reading I am giving, he means us to have in mind the ‘third proposition’ of *Groundwork I*, in which ‘action from respect for law’ is identified with duty. Thus, the sense of ‘laws’ here is not so generic as to permit, say, the possibility that a rational being has the capacity to act from hypothetical but not from categorical imperatives.

28 This reading discourages the search in *Groundwork* II for an argument to the effect that morality lies hidden within a capacity that is, at least initially, characterizable in non-moral terms. This does not entail that Kant nowhere attempts such an argument, though I find it difficult to find one in his Critical published works. (This is not to deny that he insists that no capacity whose principle is other than the moral law is practical reason. See, e.g. his remarks about the ‘favoured creature’ at G, 4: 395.)

29 Something similar can be said about the initial appearance of the FUL in *Groundwork* I (G, 4: 401–2). The big difference between its appearance there and its appearance in *Groundwork* II is that in the former case, no notion of a categorical imperative – and, in general, no notion defined in terms of the capacity of practical reason – makes an appearance because the only resource Kant permits himself at that stage is common moral cognition. Thus, the FUL appears against the background of a less philosophically developed understanding of systematic universality than it does in *Groundwork* II.

30 Thus, while I agree with Ido Geiger when he claims that ‘a substantive conception of morality is given and the task of the *Groundwork* is to make explicit the form of common moral reason, that is, its universality’ (Geiger 2010: 272), I have suggested that it is open to us to doubt whether the FUL succeeds in making that form – which is, I have claimed, systematic universality – explicit. It is compatible with everything I have said that Geiger is right to say that if neither the maxim assessed by the universalization test nor the FUL itself contains any moral notions, the test cannot generate morally normative results (p. 276). Indeed, it is compatible with everything I have said that, if the FUL fails to express systematic universality, this is on account of the constraint that Kant takes to be grounded in its status as a typic: that it and what it assesses be formulated in wholly natural terms (§3.3). My aim in this article has been to argue that *even if* the FUL does indeed fail to express systematic universality (as I think it well might), nonetheless, it is important that this failure is different from the commonly alleged failure of the FUL to be equivalent to purpose-indifference, and the main expository sweep of *Groundwork* II, from a philosophically undeveloped understanding of systematic universality to a philosophically developed one, is in good argumentative standing.

31 My claim is not equivalent to Wood’s suggestion that ‘[t]he FUL and FLN are merely provisional and incomplete formulations of the principle of morality, which always depend for their application on other independent rational principles’ (Wood 1999: 91). My claim is that what immediately precedes the introduction of the FUL expresses a merely partially developed understanding of systematic universality. It is that material which is, in a sense, merely provisional, for what it expresses gains in philosophical explicitness in later formulations, all of which express something of which the FUL is merely the typic.

## References

- Allison, Henry E. (1991) ‘On a presumed gap in the derivation of the categorical imperative’. *Philosophical Topics*, 19(1), 1–15.
- (2011) *Kant’s Groundwork for the Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press.
- Aune, Bruce (1979) *Kant’s Theory of Morals*. Princeton, NJ: Princeton University Press.
- Callanan, John J. (2019) ‘Kant on misology and the natural dialectic’. *Philosophers’ Imprint*, 19(47), 1–22.
- Engstrom, Stephen (2009) *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Cambridge, MA: Harvard University Press.
- Fix, Jeremy D. and Pendlebury, T. A. (forthcoming) ‘The rational faculty of desire’. In Stefano Bacin and Carla Bagnoli (eds.), *Reason, Agency, and Ethics: Kantian Themes in Contemporary Debates* (Oxford: Oxford University Press).
- Geiger, Ido (2010) ‘What is the use of the universal law formula of the categorical imperative’. *British Journal for the History of Philosophy*, 18(2), 271–95.
- (2015) ‘How are the different formulas of the categorical imperative related?’ *Kantian Review*, 20(3), 395–419.
- Grenberg, Jeanine (2013) *Kant’s Defense of Common Moral Experience: A Phenomenological Account*. Cambridge: Cambridge University Press.
- Guyer, Paul (2016) ‘The twofold morality of *Recht*: once more unto the breach’. *Kant-Studien*, 107(1), 34–63.
- (2019) *Kant on the Rationality of Morality*. Cambridge: Cambridge University Press.
- Herman, Barbara (2010) ‘The difference that ends make’. In Julian Wuerth (ed.), *Perfecting Virtue: Kantian Ethics and Virtue Ethics* (Cambridge: Cambridge University Press), 92–115.



- (2021) *The Moral Habitat*. Oxford: Oxford University Press.
- Hill, Thomas E., Jr. (1992) *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca, NY: Cornell University Press.
- Hills, Alison (2009) 'Happiness in the *groundwork*'. In Jens Timmermann (ed.), *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide* (Cambridge: Cambridge University Press), 29–44.
- Kant, Immanuel (1992–2016) Paul Guyer and Allen W. Wood (eds.), *The Cambridge Edition of the Works of Immanuel Kant* (Cambridge: Cambridge University Press).
- Kleingeld, Pauline (2017) 'Contradiction and Kant's formula of universal law'. *Kant-Studien*, 108(1), 89–115.
- (2021) 'Self-contradictions of the will: reply to Jens Timmermann'. *Kant-Studien*, 112(4), 611–22.
- Korsgaard, Christine M. (1996) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Nyholm, Sven (2015) 'Kant's universal law formula revisited' *Metaphilosophy*, 46(2), 280–99.
- O'Neill, Onora (1989) *Constructions of Reason*. Cambridge: Cambridge University Press.
- Pendlebury, T. A. (2022) 'The real problem of pure reason'. *European Journal of Philosophy*, 30, 45–63.
- Reath, Andrews (2006) *Agency and Autonomy in Kant's Moral Theory*. Oxford: Clarendon Press.
- (2015) 'The 'Kantian roots' of the original position'. In Timothy Hinton (ed.), *The Original Position* (Cambridge: Cambridge University Press), 201–23.
- Ripstein, Arthur (2009) *Force and Freedom*. Cambridge, MA: Harvard University Press.
- Schafer, Karl (2019) 'Kant: constitutivism as capacities-first philosophy'. *Philosophical Explorations*, 22(2), 177–93.
- (2021) 'Transcendental philosophy as capacities-first philosophy'. *Philosophy and Phenomenological Research*, 103(3), 661–86.
- Sticker, Martin (2016) 'Kant's criticism of common moral rational cognition'. *European Journal of Philosophy*, 25(1), 85–108.
- Timmermann, Jens (2006) 'Kantian duties to the self, explained and defended'. *Philosophy*, 81(317), 505–30.
- (2018) 'A tale of two conflicts'. *Kant-Studien*, 109(4), 581–96.
- Ware, Owen (2021) *Kant's Justification of Ethics*. Oxford: Oxford University Press.
- Willaschek, Marcus (2009) 'Right and coercion: can Kant's conception of right be derived from his moral theory?' *International Journal of Philosophical Studies*, 17(1), 49–70.
- Wood, Allen W. (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.