

Valtteri Viljanen*

Kant on Moral Agency: Beyond the Incorporation Thesis

<https://doi.org/10.1515/kant-2019-8899>

Abstract: This paper aims to discern the limits of the highly influential Incorporation Thesis to give proper weight to our sensuous side in Kant’s theory of moral action. I first examine the view of the faculties underpinning the theory, which allows me to outline the passage from natural to rational action. This enables me to designate the factors involved in actual human agency and thereby to show that, contrary to what the Incorporation Thesis may tempt one to believe, we do not always act on maxims. The result is a revised and more balanced view of how Kant sees the character of moral life.

Keywords: maxims, incorporation thesis, moral agency, faculty of desire, freedom

Introduction

“The essential point is that in characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says.”¹ So reads the passage that continues to echo dominantly in prominent interpretations of Kant and his immediate successors. It concerns knowing, but the idea has been applied to Kant’s view of morality and human agency as well; to take a recent example from the scholar often hailed as the most influential of his generation, Henry Allison: “The point is that just as a cognizer must view itself as a self-determiner with respect to its judgments, so a rational agent must regard itself as determining itself to act on the basis of self-imposed principles and therefore as standing in the logical space of *practical* reasons.”²

1 Sellars, Wilfrid: “Empiricism and the Philosophy of Mind.” In: *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Ed. Herbert Feigl and Michael Scriven. Minneapolis 1956, 298 f.

2 Allison, Henry: *Kant’s Groundwork for the Metaphysics of Morals: A Commentary*. Oxford 2011, 308. Here Allison (308n21) explicitly refers to Wilfrid Sellars and John McDowell.

*Kontakt: Dr. Valtteri Viljanen, School of History, Culture and Arts Studies, University of Turku, 20014 Turku, Finland; valvil@utu.fi

Indeed, in an earlier work Allison famously introduces what he dubs the *Incorporation Thesis*, according to which inclinations determine the will “only by being ‘incorporated into a maxim,’ that is, by being taken by the agent [...] as sufficient reasons for action.”³ The Incorporation Thesis

makes it clear that for Kant an inclination or desire does not *of itself* constitute a reason for acting. It can become one only with reference to a rule or principle of action, which dictates that we ought to pursue the satisfaction of that inclination or desire. Moreover, [...] the adoption of such a rule cannot itself be regarded as the causal consequence of the desire [...]. On the contrary, it must be conceived as an act of spontaneity on the part of the agent.⁴

Ever since its introduction, the Incorporation Thesis has remained a key fixture not only in Allison’s writings but in scholarship on Kant’s practical philosophy more generally.

It is uncontroversial (although, as we will see, rarely adequately emphasized) that the Kantian view of agency is based on the notion of the faculty of desire (*Begehrungsvermögen*), which has two special features: it is *causally efficacious* – the cause of objects – *through representations*. Insofar as it is related to understanding, Kant calls it the will (KpV, AA 05: 55), which is “eine Art von Causalität” (GMS, AA 04: 446). The will is thus not only a representational power of action but one that is *reason-*, *rule-*, or *maxim-*mediated. And precisely here we encounter what has been regarded, along Sellarsian lines, as the veritable game changer: although our faculty of desire can never remain untouched by (“impulsive”) desires or inclinations, we are not at their mercy. To take an influential example, John Rawls holds that all predispositions – including the one involving inclinational instincts – “however they may affect us, cannot determine our will unless they are incorporated into our maxims by our free power of choice. This is Kant’s principle of elective will, a basic principle of his moral psychology.”⁵ Barbara Herman contends similarly: “It is the mark of a rational agent that incen-

³ Allison, Henry: *Kant’s Theory of Freedom*. Cambridge 1990, 126. In this paper, I will follow Allison in understanding “inclination” (*Neigung*) in a wide sense, encompassing “momentary desires, instincts, passions, fears, and disinclinations [...], all of which pertain only to sensuously affected beings” (*Kant’s Theory of Freedom*, 108 f.), and there are good textual grounds for this move (see e.g. GMS, AA 04: 398, 413n). As has been pointed out (see e.g. Morrison, Iain: *Kant and the Role of Pleasure in Moral Action*. Athens, Ohio 2008, 37–39; Frierson, Patrick: *Kant’s Empirical Psychology*. Cambridge 2014, 68–70), however, Kant also defines inclination as a *habitual* desire (see esp. MS, AA 06: 212). With this noted, these classificatory questions do not affect my overall argument, and I will set them aside for the sake of clarity and brevity.

⁴ Allison: *Kant’s Theory of Freedom*, 40.

⁵ Rawls, John: *Lectures on the History of Moral Philosophy*. Ed. Barbara Herman. Cambridge, Mass. 2000, 295.

tives determine the will only as they are taken up into an agent's maxim."⁶ Here the focus is on maxims that determine the human faculty of desire: a specific kind of principle-ladenness is simply a consequence of our rational condition. I believe that Christine Korsgaard sums up very instructively how the relationship between desires (or inclinations) and reason looks from this perspective: "You may of course *choose* to act on a desire, but *that does not mean that you are impelled by it*. It means you take this desire as a reason, or, in Kant's language, you make it your maxim to satisfy this desire."⁷

Kantian maxims – subjective principles of volition – are famously a matter of prolonged debate. Apart from the Incorporation Thesis, I wish to stay as neutral about them as possible and accept all but the very first of the following features of maxims on which Rob Gressis estimates, in his recent overview article, there to be "significant agreement in the secondary literature":

1. Everyone always acts on maxims.
2. Maxims determine how we act in specific situations through the use of practical rules.
3. Everyone has a highest maxim that affects the other maxims she adopts.
4. Maxims can be linguistically formulated and then evaluated for their moral worth (and the deontic status of their accompanying actions) by the FUL [Formula of Universal Law].
5. Because maxims determine how we act, the deontic status and moral worth of our actions, and whether we are good or evil people, maxims express our character.
6. Maxims, or rather whatever kind of maxim it is that people adopt whenever they act, describe what we intend to do or what we see as good rather than prescribe courses of action that we may or may not live up to.⁸

The first item on the list really is agreed on to a remarkable degree. Despite and against this, I will argue that it is simply not the case that, according to Kant, we always act on maxims. Even though the Incorporation Thesis draws attention to something that is definitely integral to Kant's theory of moral agency – to the central place that *free* and *rational* choice occupies in that theory – it still con-

⁶ Herman, Barbara: *The Practice of Moral Judgment*. Cambridge, Mass. 1993, 11 f. Proceeding along Rawls-inspired lines, Andrews Reath ("Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination." In: *Kant-Studien* 80, 1989, 295) contends that "no incentives (including sensible incentives) determine the will directly except through a choice by the individual." It is striking that even Richard McCarty, who otherwise criticizes this line of interpretation and (controversially) reads Kant as a psychological determinist, nevertheless contends that "[t]o be human is to act always, and only, on maxims" (*Kant's Theory of Action*. Oxford 2009, 231).

⁷ Korsgaard, Christine: "Kant's Analysis of Obligation: The Argument of *Groundwork* I." In: *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. Ed. Paul Guyer. Lanham 1998, 63, the latter emphasis added.

⁸ Gressis, Rob: "Recent Work on Kantian Maxims II." In: *Philosophy Compass* 5, 2010, 235.

cerns only a part of the story that is, in its entirety, more complex than what has been presented thus far. I will argue that we should carefully designate the limits of the Incorporation Thesis and go beyond it to see that Kant operates with a notion of agency that is broad enough to give proper weight to our sensuous or natural side. To accomplish these goals, I will proceed in the following, slightly uncommon, fashion. Instead of beginning directly with Kant's theory of moral agency, I start by examining the view of the faculties underpinning Kant's theory of action (section 1) to reveal the most general features of that theory. This enables me to outline the passage from natural to rational action (section 2), after which I will be in a position to discern the factors involved in actual human agency (section 3). This, in turn, shows that, contrary to what the Incorporation Thesis may tempt one to believe, we do not always act on maxims, which leads us to a revised view of how Kant sees the character of moral life (section 4). The resulting account is one that offers, I believe, a new and balanced understanding of the different elements, both conceptual and causal, that constitute Kant's theory of moral agency.

1 Faculty Background of Action

Already in the first *Critique*, Kant takes it as a given that we are, fundamentally, not only *knowers* but also *agents*. Now both of the basic aspects of being human presuppose, for Kant, certain abilities, powers, or faculties.⁹ Here he is very much in line with the tradition: from the thirteenth century onward, it was routinely held that there are two basic mental faculties: cognition and desire.¹⁰ Faculty psychology was very much in vogue in eighteenth-century Germany; launched by Wolff and developed further by Baumgarten, Mendelssohn, and Tetens, it led to a theory of three basic faculties, namely understanding, will, and feeling,¹¹ which

⁹ This fundamental aspect of Kant's thought has recently received welcome increased interest; see esp. Frierson: *Kant's Empirical Psychology*, ch. 2; Wuerth, Julian: *Kant on Mind, Action, and Ethics*. Oxford 2014, ch. 6.

¹⁰ Scheerer, Eckart: "On the Will: An Historical Perspective." In: *Volitional Action: Conation and Control*. Ed. Wayne Hershberger. Amsterdam 1989, 40.

¹¹ Hilgard, Ernest: "The Trilogy of Mind: Cognition, Affection, and Conation." In: *Journal of the History of Behavioral Sciences* 16, 1980, 108–110. Crusius was also important for Kant "in opposing their [Wolff's and Baumgarten's] attempts to reduce [the] faculties to a single essence" (Frierson: *Kant's Empirical Psychology*, 53).

Kant ultimately endorses.¹² This background makes it easier to understand why Kant is obviously not ready to find anything problematic in invoking faculties:¹³ he is strikingly happy to postulate a faculty (or faculties) and to simply move on to critical philosophical analysis of the limits and governing principles of the relevant basic aspects of being human.¹⁴

For our purposes, the key definition presented in the second *Critique* reads: “Das **Begehrungsvermögen** ist das *Vermögen desselben, durch seine Vorstellungen Ursache von der Wirklichkeit der Gegenstände dieser Vorstellungen zu sein.*” (KpV, AA 05: 09n)¹⁵ In other words, we can be, in virtue of our faculty of desire, the causes of represented objects. Thus, reduced to its very basics – and this is very important – *to act is simply to be causally efficacious* in this way. However, in his published works Kant does regrettably little to clarify the notion of faculty (*Vermögen*) that he operates with. Its connection to the concept of power or force (*Kraft*) is widely acknowledged in the literature, but detailed analyses of that connection are hard to find.¹⁶ Fortunately, Kant’s handwritten notes to his copy of Baumgarten’s *Metaphysics*, the work he used as the textbook in his lectures, contains a very informative formulation of the matter. Next to § 216 (on faculty and receptivity) we find the following remark by Kant:

Das innere Princip der Möglichkeit des Handelns ist das Vermögen[.] [...] Das Vermögen mit einem Bestimmungsgrunde desselben zusammen [...] ist Kraft. [...] Also *posita vi (per se) ponitur actio*. Allein dem Bestimmungsgrunde [...] in der Kraft kan ein anderer wiederstreiten. [...] Der Zustand einer Kraft bey diesem Wiederstreit ist Bestrebung (*conatus*) und das, was den Grund davon enthält, daß die Handlung nicht entspringt, die Hindernis. Ein jeder *conatus* ist eine Handlung, der aber etwas entgegen wirkt. Daher ist er todte Kraft. Die lebendige Kraft ist nicht größer als der *conatus*, bringt aber eine begehrte Wirkung hervor. (R 3585, Refl, AA 17: 73)¹⁷

In plainer terms, the view – with which Kant’s published writings are consistently in accord – is as follows. In itself, a faculty is basic causal efficacy of a certain

¹² Kant’s considered tripartite division is presented most explicitly in the third *Critique* (AA 05: 196–198).

¹³ KpV, AA 05: 09n expresses this attitude especially clearly. I do not, however, want to say that Kant did not have solid transcendental reasons for introducing faculties, but discussing this would take us too far afield.

¹⁴ For an instructive discussion of Kant’s unstable faculty terminology, see Wuerth: *Kant on Mind, Action, and Ethics*, 196 f., 202.

¹⁵ See also MS, AA 06: 211.

¹⁶ Béatrice Longuenesse (*Kant and the Capacity to Judge*. Princeton 1998, 7) and Richard McCarty (*Kant’s Theory of Action*, 18) offer brief, and differing, accounts of the connection.

¹⁷ See also V-Met-L2/Pöhlitz, AA 28: 565.

kind; once determined, it becomes a power that strives to produce specific effects. If opposing powers hinder it from that production (not from striving), the power is dead; if not, it is alive. In fact, this view is part and parcel of the Leibnizian variant of the traditional metaphysics of powers;¹⁸ in essence, the view that had established itself through centuries is that when there is (1) *something intrinsically causally efficacious* (broadly speaking, a power) coupled with (2) *a principle of determination*, (3) what results is *an operation* (striving to act in a certain way) that (4) produces *an effect* if nothing intervenes. Power → determination → operation → effect can be characterized as a deeply entrenched way of thinking about the metaphysical architectonics of causality, and despite some terminological differences, Kant's faculty → determination → power → effect scheme matches it seamlessly. Thus, and as we will see in greater detail in what follows, his theory of moral agency builds on the basic view of action as something that results from the faculty of desire, which has come to be determined in a specific way.¹⁹

2 From Natural to Purely Rational Action

Kant does have a story, albeit a very short one, to tell about the origin of the faculty of desire: we are living creatures, or entities endowed with vital force (*Lebenskraft*), and the faculty of desire is a manifestation of that force (KpV, AA 05: 23).²⁰ To be alive is to act in virtue of the faculty of desire according to one's representations.²¹ As this of course holds for all animate things, it is instructive to take a look at the relatively uncomplicated case of the instinctive life of non-rational animals. Their (1) faculty of desire comes to be (2) determined by feelings of pleasure or pain, which leads to (3) striving to behave in a certain way (to desire or aversion) (4) and thereby to producing the effect that satisfies a need and is thereby conducive to self-preservation.²² All this takes place as dictated by empirical laws of nature.²³

18 Leibniz's position is perhaps most clearly expressed in "Specimen Dynamicum" of 1695.

19 Here I am in general agreement with Eric Watkins, who argues that Kant endorses "a causal powers model of causality" (*Kant and the Metaphysics of Causality*. Cambridge 2005, 231).

20 Kant (KpV, AA 05: 162) says that we do not know how we come to be provided with vital force.

21 See KpV, AA 05: 09n; MS, AA 06: 211.

22 This reconstruction is based on GMS, AA 04: 395 f.; MS, AA 06: 211 f.; and Anth, AA 07: 231; cf. Frierson: *Kant's Empirical Psychology*, 58. For the complex interplay between the basic faculties, see Frierson: *Kant's Empirical Psychology*, 55 f.; Wuerth: *Kant on Mind, Action, and Ethics*, 202.

23 See e.g. GMS, AA 04: 427.

But despite the fact that the human being is immersed in nature, inhabiting a tiny part of an infinite universe, she finds herself endowed with something that transcends it: reason. On Kant's view, reason is even a frightful capacity: strikingly disconnected from and even unsuitable for bringing about happiness but capable of opening a domain beyond nature, namely that of the ought, of how the world should be.²⁴ Having revealed in the first *Critique*, once and for all by his own lights, the boundaries and nature of knowledge as cognition of objects, Kant feels that he has landed on secure ground from which to launch an investigation into what can be philosophically said about things of which no knowledge (in the strict objectual sense) is attainable, or, more familiarly put, about matters not of knowledge but of action. And precisely here we encounter a striking turn of the tables: the critical project has given practical reason the much-needed "Ruhe und Sicherheit" (GMS, AA 04: 457) to find out that the spontaneity of reason with which it rises above understanding and sensibility is no more a weakness – a source of dialectical dead ends – but a strength: no faculty other than reason can give us access to the world beyond appearances, or the intellectual world. But what can be found in that world? Even after the first *Critique*, we are still without an account of at least two things that are as plain as they are difficult to grasp: that there is moral right and wrong, and that our actions can be free.²⁵ Both have to do with the fact that we are not only knowers but agents, and, for Kant, both can find a home only in the intellectual world of reason. In critical practical philosophy there is thus little option but to take a rationalist turn, which for Kant, "a rationalist by training as well as by temperament,"²⁶ is hardly an unpleasant prospect.

It is thus first and foremost reason that takes us from activity as mere behavior to action in the strong sense – to normative agency. In terms of the faculty background discussed above, Kant (KpV, AA 05: 22) makes a (completely traditional) distinction *within* the faculty of desire: it has two aspects or expressions,²⁷ the lower and the higher. The *former* is the human counterpart of what we have already encountered; it is a mere (*bloß*) faculty of desire (GMS, AA 04: 459), externally and therefore mainly passively determined (but not thereby causally inefficacious) by the agreeableness of objects, ultimately aiming at happiness.²⁸

²⁴ See GMS, AA 04: 395 f.

²⁵ Of course, Kant also approves of the traditional division of philosophy into logic, physics, and ethics (GMS, AA 04: 387), so no system of philosophy can be complete without ethics.

²⁶ Korsgaard, Christine: *Creating the Kingdom of Ends*. Cambridge 1996, 45.

²⁷ See KrV, A 649–652/B 677–680.

²⁸ See KpV, AA 05: 22–24. Happiness is, for Kant, the full satisfaction of one's inclinations and as such empirical (GMS, AA 04: 405, 418).

In brief, determined by a sensible representation of a pleasurable object, desire prompts action to bring about the object. The *latter* is the famous Kantian will: in itself active, rational, and pure, a faculty “sich zum Handeln als Intelligenz [...] zu bestimmen” (GMS, AA 04: 459), a desire of practical reason²⁹ – our access to the noumenal world beyond sense experience, this being the aspect of human action that most clearly distinguishes it from animal behavior. Fortunately for us, Kant is “modern” enough to understand the trepidation we may feel about taking such a non-empirical turn – as attached to sense experience as we are – and so he spills much ink to alleviate our worries.

We are, then, agents of will; but what can we find out about the nature of this supersensible power?³⁰ “Ein jedes Ding der Natur wirkt nach Gesetzen” (GMS, AA 04: 412), but what about a realm beyond sensible nature? It is, for Kant, a governing assumption – deeply integrated but still by no means concealed – that there cannot be a proper domain of reality without laws governing it,³¹ and the very notion of a law requires strict universal legitimacy.³² Kant does not say *why* law-governedness is so important; arguably the idea is that without laws, a domain does not have the structure and order required of a proper domain; a lawless domain is an unthinkable non-reality. Within the practical domain of human action this means that the source of efficacy, our faculty of desire, must also fall under universal laws, be they laws of the noumenal or the phenomenal kind.

We are of course quite familiar with the laws of nature that govern the phenomenal world. But what about the laws of rational willing? Will is, as we have seen, “eine Art von Causalität” (GMS, AA 04: 446), and so Kant states:

Da der Begriff einer Causalität den von *Gesetzen* bei sich führt, nach welchen durch etwas, was wir Ursache nennen, etwas anderes, nämlich die Folge, gesetzt werden muß: so ist die Freiheit, ob sie zwar nicht eine Eigenschaft des Willens nach Naturgesetzen ist, darum doch nicht gar gesetzlos, sondern muß vielmehr eine Causalität nach unwandelbaren Gesetzen, aber von besonderer Art sein; denn sonst wäre ein freier Wille ein Unding. (GMS, AA 04: 446)

²⁹ See Allen Wood’s *Kantian Ethics* (Cambridge 2008, 35 f.) for an argument that all action, rational or not, involves desire.

³⁰ It should be noted that to be supersensible (*übersinnlich*) is to be separated from being *supernatural* (V-Mo/Mron II, AA 29: 603).

³¹ Cf. GMS, AA 04: 433: “Ich verstehe aber unter einem *Reiche* die systematische Verbindung verschiedener vernünftiger Wesen durch gemeinschaftliche Gesetze.” See also KpV, AA 05: 43; KU, AA 05: 268; RGV, AA 06: 190.

³² See e.g. GMS, AA 04: 387; MS, AA 06: 214 f.

Kant's key move in finding the law of the rational will is to focus on the very idea of *what it is to be a law*: it follows from the universality requirement of laws that any law of the noumenal realm of the will must be such that it can be *universally* willed (a non-universal law is no law to begin with); Kant is especially explicit about this in the second *Critique* (AA 05: 34 f.). This is the rationale behind the unique categorical imperative, "*handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daß sie ein allgemeines Gesetz werde*" (GMS, AA 04: 421). In fact, what it formulates is a meta-principle: a principle for testing principles of action to find those that are universal. As such, it can reveal the laws that structure the realm of the purely rational will so that we can latch on to those laws and act accordingly.³³ Whatever its (objectual) effects, purely rational action results from the faculty of desire as will determined by these kinds of universal principles alone.³⁴

3 Determiners of Actual Human Action

Kant has a special name for the kind of will that is necessarily determined by pure practical reason alone: the holy will. It is the purest and as such in fact the paradigmatic case of a will determined to action; it reveals what can be termed the purified core of the engine of moral action.³⁵ All natural inclinations notwithstanding, a human being can form a view of what it would be like to act according to the laws of reason alone. Being, as such, out of our reach does not prevent the holy will from functioning as a regulative idea which we should strive to approximate to the fullest extent possible.³⁶ According to Kant, only here can be found the standard (*Maßstab*) for what counts as good and evil in the first place;³⁷ it is only *after* we have discovered the principles forming the bedrock of rational agency that we should seek their meaning for sensual beings such as we are *qua* human beings. And indeed, in actual human existence our faculty of desire, affected as it is by the unswerving sensible impulses, cannot remain the purely rational will; this results, most importantly, in feelings and inclinations. In the 1780s, Kant appears to be basically hostile towards inclinations: "Die Nei-

³³ See esp. GMS, AA 04: 435.

³⁴ Should one wonder how principles (or laws) and faculties square with each other, the answer is, I think, that in Kant's view, human (and, more generally, animate) causality is about faculties being determined by principles.

³⁵ See KpV, AA 05: 119, 122; MS, AA 06: 383, 409.

³⁶ See KpV, AA 05: 32; KrV, A 569 f./B 597 f.

³⁷ See GMS, AA 04: 390, 408, 442.

gungen selber aber als Quellen des Bedürfnisses haben so wenig einen absoluten Werth, um sie selbst zu wünschen, daß vielmehr, gänzlich davon frei zu sein, der allgemeine Wunsch eines jeden vernünftigen Wesens sein muß” (GMS, AA 04: 428). Even though his attitude seems to have become considerably more lenient by the 1790s,³⁸ there is no denying that the human agent is faced with a complicated situation.

I would like to make two brief remarks before moving on to introduce the crucial tripartite distinction at play in Kant’s theory of moral motivation. First, in the real world of sensible “friction,” our reason is not purely rational and moral; it is also prudential, used to find the means to attain the ends we consider conducive to our happiness.³⁹ This does *not* deprive us of freedom, for also taking up a prudential end – or adopting a prudential maxim – is always up to us.⁴⁰ Second, it should be kept in mind that both inclinations and maxims concern *one and the same* causally efficacious faculty (that of desire). Whether entitled to it or not, Kant treats the faculty of desire as one package (albeit with different aspects) and is keen to emphasize the unity of the acting human subject.⁴¹

We can now move on to the fundamental distinction triad. Kant does not offer it himself, but I would argue that it underpins so many of his discussions on moral motivation that it would not make sense to try to list the relevant occurrences; the division into the three original predispositions to good in *Religion* (AA 06: 26–28) comes closest to an explicit formulation of the tripartite distinction. Given the faculty → determination → action → effect scheme, the crucial question concerns what *determines* the faculty of desire.⁴² If it were pure reason alone, we would be holy beings, but we are not, and there are three possible determiners of our faculty of desire: (1) maxims that comply with the moral law, (2) maxims

³⁸ See RGV, AA 06: 58.

³⁹ The imperatives operative here are of course hypothetical, not categorical.

⁴⁰ See Allison: *Kant’s Theory of Freedom*, 35; Morrisson: *Kant and the Role of Pleasure in Moral Action*, ch. 3. Kant considers it a blatant contradiction to think that someone could *make* us take something as an end; at most, we can be made to act as a means to someone else’s end (MS, AA 06: 381).

⁴¹ See e. g. GMS, AA 04: 456. Here it is instructive to consider Kant’s similarly unificatory view of *character*: even though the noumenal aspect of character is more fundamental than the phenomenal, the formation of character can – and should – be crucially influenced by empirical causes, especially by education (see Frierson: *Kant’s Empirical Psychology*, 78–80), and Kant sees nothing problematic in this.

⁴² The question thus concerns, to use Kant’s favorite expression, “der Bestimmungsgrund des Willens” (GMS, AA 04: 401). KpV (AA 05: 30) declares that free will must also be determinable. For discussions on determination, see Watkins: *Kant and the Metaphysics of Causality*, 354–360; Timmermann, Jens: “Acting from Duty: Inclination, Reason and Moral Worth.” In: *Kant’s Groundwork of the Metaphysics of Morals: A Critical Guide*. Ed. Jens Timmermann. Cambridge 2009, 47.

(ultimately of happiness) that comply with our inclinations,⁴³ and (3) inclinations themselves.⁴⁴

As we have seen, the first two types of determiners are widely emphasized in the literature, whereas the third type, namely inclinations, are treated as determining us only if they are incorporated into the maxims of the second kind. Let us thus consider reasons for thinking that it is important for Kant to allow inclinations to directly determine our faculty of desire. Before considering textual evidence, I would like to note that the idea that we can be directly determined to act by inclinations has considerable *prima facie* plausibility; what, one might ask, could be more quotidian than rationally endorsing a principle (e. g. concerning healthy nourishment) and still acting impulsively against it (e. g. by eating a fattening piece of cake)? In fact, I think it is much more contentious to deny the existence of this phenomenon than to affirm it. But of course, Kant could still be committed to precisely such a denial. To show that he is not, I would like to discuss a select group of passages that indicate that he acknowledges that we can be determined to act directly from inclination or instinct, after which I will critically examine the famous passage on which the Incorporation Thesis rests.

Let us begin with a passage from the second *Critique*, in which Kant declares in plain terms that inclinations can by themselves determine the will (or faculty of desire), and in which no mention is made of maxims:

Und darauf beruht der Unterschied zwischen dem Bewußtsein, *pflichtmäßig* und *aus Pflicht*, d. i. aus Achtung fürs Gesetz, gehandelt zu haben, davon das erstere (die Legalität) *auch möglich ist, wenn Neigungen blos die Bestimmungsgründe des Willens gewesen wären*, das zweite aber (die *Moralität*), der moralische Werth, lediglich darin gesetzt werden muß, daß die Handlung aus Pflicht, d. i. blos um des Gesetzes Willen, geschehe. (KpV, AA 05: 81, the penultimate emphasis added)

In his lectures on moral philosophy, Kant says that we can act from inclinations *instead of* maxims, which indicates that not all actions are maxim-mediated:

Wir wollen die Handlungen des Menschen [...] erwegen [...]. Sie entspringen aus Antrieben und aus Neigungen oder aus Maximen und Principien. (V-Mo/Collins, AA 27: 345)

⁴³ See e. g. GMS, AA 04: 406, 420n, 427; KpV, AA 05: 36, 61 f.

⁴⁴ Here a qualification (to what I state in note 3 above) is in order: it is difficult to see how a subclass of inclination, namely a *habitual* desire, could determine our faculty of desire other than by being incorporated into a maxim. I would like to thank an anonymous reviewer for pointing this out.

Unsere Maximen müssen wohl überlegt seyn, und *es ist ärger böses zu tun aus Maximen als aus Neigung*. (V-Mo/Collins, AA 27: 368, emphasis added)⁴⁵

In his *Anthropology*, Kant endorses the idea that some people are prone to act in certain ways naturally or instinctively, with no principles (and thus maxims) involved, while for others acting from principles is relatively easy:

Dagegen, um von ihm sagen zu können: “Er hat ein gut *Herz*”, ob dieses zwar auch zur Sinnesart gehört, will schon mehr sagen. Es ist ein Antrieb zum Praktisch-Guten, wenn es gleich nicht nach Grundsätzen verübt wird [...]. (Anth, AA 07: 286)

Mit einer ganz gewöhnlichen Dosis der Vernunft, aber zugleich diesem Phlegma von der Natur ausgestattet, ohne zu glänzen, und doch *von Grundsätzen, nicht vom Instinct ausgehend*, hat der Kaltblütige nichts zu bereuen. (Anth, AA 07: 290, emphasis added)⁴⁶

Finally, Kant’s criticism of heteronomous moral systems is based on the notion of motives that *cannot* be “principled.” It is difficult to see what such non-principled motives could be other than inclinations:

45 As Manfred Kuehn (“Collins: Kant’s Proto-Critical Position.” In: *Kant’s “Lectures on Ethics”: A Critical Guide*. Ed. Lara Denis and Oliver Sensen. Cambridge 2015, 51f.) points out, notes of Kant’s lectures were widely copied and sold, and this no doubt applies to the Collins lecture notes as well. Collins obviously bought a ready-made set of notes (most probably stemming from 1775), copied them himself, making only minor additions. Given that the lecture notes are dated to 1784, it is well-grounded to claim, along the lines of J. B. Schneewind (“Introduction.” In: Immanuel Kant, *Lectures on Ethics*. Ed. Peter Heath and J. B. Schneewind. Cambridge 1997, xvi), that Kant’s lectures on ethics remained essentially the same between 1775 and 1784. This means that the Collins lecture notes, even though not composed from scratch in 1784, represent very well what Kant taught on the topic during that period.

Now there is reason to think that *the moral status* of what Kant calls ‘maxims’ changes somewhat from instrumental or selfish in the pre-1785 period to morally neutral in the post-1785 period (so that maxims can be both bad or good; see Kuehn: “Collins,” 62f.). However, this is a merely terminological issue in the sense that Kant makes it abundantly clear even in his pre-1785 lectures that the key point is that our actions *should* be determined by *principles* that conform to universal rules; thus, a (possible) partial change to the meaning of the term ‘maxim’ has no bearing on the question of whether all actions should be – or inevitably are – “principled” or not. I would like to thank an anonymous reviewer for pressing this point.

46 Kant also holds that inclinations can *play us*, which sounds like something that is not mediated by maxims we choose: “weil, ohne daß die Vernunft die Zügel der Regierung in ihre Hände nimmt, jene über den Menschen den Meister spielen” (MS, AA 06: 408). He once even compares our usual way of acting to the (instinctive) behavior of insects: “so hat doch das Formelle des Willens überhaupt, nach festen Grundsätzen zu handeln (nicht wie in einem Mückenschwarm bald hiehin bald dahin abzuspringen), etwas Schätzbares und Bewundernswürdiges in sich; wie es denn auch etwas Seltenes ist” (Anth, AA 07: 292; see also GMS, AA 04: 396).

[E]ine vermischte Sittenlehre, die aus Triebfedern von Gefühlen und Neigungen und zugleich aus Vernunftbegriffen zusammengesetzt ist, das Gemüth zwischen Bewegursachen, *die sich unter kein Princip bringen lassen*, die nur sehr zufällig zum Guten, öfters aber auch zum Bösen leiten können, schwankend machen muß. (GMS, AA 04: 411, emphasis added)

By contrast, the famous passage usually presented as espousing the Incorporation Thesis, “die Freiheit der Willkür ist von der ganz eigenthümlichen Beschaffenheit, daß sie durch keine Triebfeder zu einer Handlung bestimmt werden kann, *als nur sofern der Mensch sie in seine Maxime aufgenommen hat*” (RGV, AA 06: 23f.), only concerns our *power of choice* (*Willkür*), not the whole faculty of desire, or even its rational part (*Wille*).⁴⁷ I will shortly explain the absolutely central role that the power of choice plays in Kant’s theory of moral agency; for present purposes, it suffices to note that the fact that we have *a power to choose between maxims* in no way implies that *the faculty of desire* could not be directly determined to action by inclinations.⁴⁸ Altogether, then, the textual evidence is decidedly in favor of the thesis that the faculty of desire is directly responsive both to principles and to inclinations: it can be – and in fact always is – affected by inclinations, which may by themselves determine us to action, even though there is no denying that inclinations can also be taken up and incorporated into maxims of action.

However, it may be argued that perhaps this is not a serious problem for those who espouse the Incorporation Thesis because they are merely focusing on us insofar as we are *rational agents* in whose actions freely chosen maxims of course figure prominently.⁴⁹ I am not sure whether interpretations building

⁴⁷ Rawls – whose reading of this passage has been so influential – renders the passage in a way that obfuscates this crucial distinction: “Freedom of the will is of a wholly unique nature in that an incentive can determine the will to an action *only so far as the individual has incorporated it into his maxim*” (*Lectures on the History of Moral Philosophy*, 294). George di Giovanni (in: *Religion and Rational Theology*. Cambridge 1996, 73) correctly translates this passage, making it clear that Kant is here referring to the power of choice. The following passages from the introduction to the *Metaphysics of Morals* also make it abundantly clear that Kant’s point concerns precisely *Willkür*: “Die menschliche Willkür ist [...] eine solche, welche durch Antriebe zwar *afficirt*, aber nicht *bestimmt* wird[,] [...] Die *Freiheit* der Willkür ist jene Unabhängigkeit ihrer *Bestimmung* durch sinnliche Antriebe” (MS, AA 06: 213).

⁴⁸ The power of choice (*Willkür*) is one of the two main features of the higher faculty of desire (the other feature is the legislating will): “Von dem Willen gehen die Gesetze aus; von der Willkür die Maximen. Die letztere ist im Menschen eine freie Willkür; der Wille, der auf nichts Anderes, als bloß auf Gesetz geht, kann weder frei noch unfrei genannt werden, weil er nicht auf Handlungen, sondern unmittelbar auf die Gesetzgebung für die Maxime der Handlungen [...] geht” (MS, AA 06: 226). For a helpful discussion, see Wood: *Kantian Ethics*, 121.

⁴⁹ See e.g. Frierson: *Kant’s Empirical Psychology*, 221n10.

on the Incorporation Thesis are always very clear about this. Be that as it may, there is definitely a tendency in the literature – intertwined with the preeminence of the Incorporation Thesis – to define *human* action in terms of *rational* action so that only actions resulting from maxims qualify as *genuine* actions.⁵⁰ This is a problematic interpretative move, however, for it results in a restricted view of agency that Kant himself shows no signs of endorsing; on the contrary, he operates with a notion of agency (and action) that is broad enough to give proper weight to both rational and natural (or “animal”) elements.⁵¹ What he says is that human action is about the human faculty of desire being determined so that an effect follows – (purely) rational determination is of course to be *preferred*, but that is another matter.

Here we encounter an interpretative issue that is extremely general in nature and not easily settled by textual evidence, having to do with one’s overall take on the spirit of Kant’s view of human agency. My sense is that its rational aspect has been overemphasized, perhaps simply because so much revolves around it when Kant analyzes (especially in the *Groundwork*) his way to the very core of morality.⁵² But strong focus on this can lead to a lopsided understanding of his views on agency and moral psychology.⁵³ To take an especially pertinent example, in the *Doctrine of Virtue* Kant is quite sensitive to the fact that we are both rational and natural beings. As the former, and to put the issue in terms of ends instead of maxims, we are free to choose the ends we pursue; indeed: “Das Vermögen sich überhaupt irgend einen Zweck zu setzen ist das Charakteristische der Mensch-

50 This is seldom made explicit, but Frierson (*Kant’s Empirical Psychology*, 221 f., 237) offers an especially clear and noteworthy recent example of this tendency.

51 Here he would appear to be committed to a line of thought already found in Aristotle: “[B]oth children and the other animals share in voluntary action, but not in choice, and acts done on the spur of the moment we describe as voluntary, but not as chosen” (*Nicomachean Ethics*. Trans. W. D. Ross. In: *The Complete Works of Aristotle II*. Ed. Jonathan Barnes. Princeton 1985, III.2, 1111b8–10; see also III.1, 1111a22–b3). Charles Kahn (“Discovering the Will: From Aristotle to Augustine.” In: *The Question of “Eclecticism”: Studies in Later Greek Philosophy*. Ed. John Dillon and A. A. Long. Berkeley 1988, 239) puts Aristotle’s point as follows: “A human being may act under the guidance of passion, that is, under the direct influence of anger or appetite; and for Aristotle such action is voluntary (*hekousion*) but not deliberate or ‘chosen.’” As Thomas Pink (“Suarez, Hobbes and the Scholastic Tradition in Action Theory.” In: *The Will and Human Action: From Antiquity to the Present Day*. Ed. Thomas Pink and M. W. F. Stone. London 2004, 135–137) shows, the Scholastics agreed that humans can act, like animals do, non-rationally but voluntarily; this line of thought is thus deeply entrenched in Western philosophy.

52 On this, see also Wuerth: *Kant on Mind, Action, and Ethics*, 198.

53 On a terminological note, when he discusses action, Kant usually opts for ‘Handlung’ or ‘handeln,’ and as the *Kant-Lexikon* (ed. Marcus Willascheck et al. Berlin 2015, 992) states, “Kants Verständnis von ‘Handlung’ ist nicht auf menschliches Tun beschränkt.”

heit (zum Unterschiede von der Thierheit)” (MS, AA 06: 392). But nature comes with its own end: “Denn *eigene Glückseligkeit* ist ein Zweck, den zwar alle Menschen (vermöge des Antriebes ihrer Natur) haben” (MS, AA 06: 386).⁵⁴ These two sides give rise to an internal conflict in us, forming an integral part of our moral lives:

Tugend ist die Stärke der Maxime des Menschen in Befolgung seiner Pflicht. – Alle Stärke wird nur durch Hindernisse erkannt, die sie überwältigen kann; bei der Tugend aber sind diese die Naturneigungen, welche mit dem sittlichen Vorsatz in Streit kommen können [...]. (MS, AA 06: 394)

On the present interpretation, the conflict here, *even for the rational agent*, is precisely where the passage says it is: between endorsed moral maxims and inclinations, not between the incorporation of moral maxims and inclination-based maxims. In fact, I think that neglecting the direct impact the inclinations can have on our faculty of desire overlooks a – perhaps even better *the* – basic aspect of the agency of animate beings in general, that of simply pursuing the objects represented as pleasurable and avoiding those represented as painful, with no maxims involved. Indeed, it seems that focusing on the way in which inclinations can be incorporated into maxims has often led to a view of agency and moral psychology in which we are in a rather robust sense detached from our inclinations, as if they could not be directly involved in determining what we do. As I have argued above, Kant has no intention of affirming such detachment.

4 Moral Action Beyond the Incorporation Thesis

If the present interpretation is on the right track, where does it leave us with Kant’s theory of moral action? If the Incorporation Thesis has been coupled with an overly rationalistic view of moral agency, how to move beyond it? Let us return to the determiner triad of (1) moral maxims, (2) prudential (“inclination-*al*”) maxims, and (3) inclinations. It can be said that both types of maxims are relevant insofar as we are considering ourselves as free and rational deliberators; moreover, they are consciously articulable, for they involve the setting of discursive (conceptually mediated) ends. However, moral maxims originate from the pure use of reason, prudential maxims from its instrumental use (reason finding

⁵⁴ For Kant’s definition of a natural end (*Naturzweck*), see MS, AA 06: 424; for his lists of natural ends, see MS, AA 06: 420 and RGV, AA 06: 26 f.

means to effects it does not itself determine). Inclinations are non-rational and purely desirous; they can conflict with moral maxims, but adopting prudential maxims can keep us from countering inclinations in the first place by subordinating moral maxims to prudential ones, which equals proper evil.⁵⁵ All of the three types of determiners are thus operative in Kant's theory of moral motivation: it is between them that the fate of the determination of our faculty of desire – that is, the faculty of action – comes to be decided.

On the basis of the account thus far, I would recommend that we adopt the following view, broader than the one provided by those who build their interpretation on the Incorporation Thesis, of Kant's theory of moral agency. It is based on the view that, for Kant, our faculty of desire is determined either by moral maxims, prudential maxims, or inclinations so that once the determination is in place (or "posited"), action necessarily results.⁵⁶ This does not take away our freedom, for it is built into the theory: two determiners out of three are within our free command. More precisely, the view is this. Some maxims are inclination-based, others nonempirical and moral, but maxims proceed from the power of choice, and human choice is free (MS, AA 06: 226). Thus, *we are free to choose which maxim to endorse*; incorporating inclinations into maxims is an absolutely central feature of our moral lives, and we can be guilty of prioritizing maxims wrongly even to the extent that we give up morality altogether by systematically subordinating moral maxims to maxims aiming at our own happiness. But in moral agency there is another major element that has been thus far neglected, namely *succeeding* in acting on moral maxims. It is *not* something we can freely choose: "zwischen der Maxime und der That ist noch ein großer Zwischenraum" (RGV, AA 06: 46), the traversing of which requires that our maxim-involving resolution prevails in the conflict with inclinations; this requires virtue. To sum up, moral action requires not only (1) a free choice of moral maxims but also (2) gaining the upper hand over the inclinations so that our faculty of desire comes to be determined in the right way, which results in the action that executes the moral law.

This interpretation fits well with, and draws some of its inspiration from, a striking aspect of Kant's terminology. As we have seen, virtue is about overcoming the conflicting natural inclinations (MS, AA 06: 394); Kant himself notes that

⁵⁵ See RGV, AA 06: 36 f., 58n.

⁵⁶ As Kant puts it in his lectures, no action can escape the principle of determinism (V-MS/Vigil, AA 27: 504), and here he is *not* talking about us *qua* phenomenal beings. The moral challenge is simply to be *self*-determined. Here it should perhaps be noted that (*pace* McCarty) I do not see our power of choice (*Willkür*) as determined by the strength of psychological incentives; as will be made clear momentarily, I consider our choice to be radically free.

the German word for virtue, *Tugend*, comes from *taugen*, ‘to be fit for’ something (MS, AA 06: 390). In fact, to be virtuous is to be fit for fighting the inclinations: virtue as our very *moral condition* is “moralische Gesinnung *im Kampfe*” (KpV, AA 05: 84) that never ends, with virtue sinking if it does not rise (MS, AA 06: 409). Moreover, not only does Kant write about the conflict (*Streit*) “den jetzt die moralische Gesinnung mit den Neigungen zu führen hat, in welchem nach einigen Niederlagen doch allmählig moralische Stärke der Seele zu erwerben ist” (KpV, AA 05: 147), but his language is at times viscerally combative: “Die Laster, als die Brut gesetzwidriger Gesinnungen sind die Ungeheuer, die er nun zu bekämpfen hat” (MS, AA 06: 405).⁵⁷ Clearly, the road to a noble character of self-government (MS, AA 06: 407) not only is a battle but *feels* like one as well. This is not the place to examine how exactly this conflictual aspect of moral agency is to be conceived;⁵⁸ for present purposes, it suffices to note that seeing it in terms of incorporating inclinations into maxims does not seem to do justice to the strikingly antagonistic terminology Kant opts for when he depicts our moral conflicts. In the same vein, the view of conflicts that Allison praises,⁵⁹ namely Andrews Reath’s account, according to which the appropriate metaphor for understanding them is that “of a struggle between two parties for something like legal authority or political legitimacy,”⁶⁰ is quite good for explicating what is at stake when we *choose which maxims to adopt*, but the metaphor squares poorly with Kant’s belligerent language because that language (especially in the *Doctrine of Virtue*) is so often not about the choice of maxims but about *acting according to the adopted maxims* despite the countering inclinations.

A broader interpretation of Kant’s theory of agency is also consistent with his view that the real grounds of our actions are largely *opaque* to us:

57 Lewis White Beck (*A Commentary on Kant’s Critique of Practical Reason*. Chicago 1960, 227n48) draws attention to what he calls Kant’s “military symbolism” (in KpV, AA 05: 84).

58 Such an examination would have to take into account Kant’s moral anthropology: “Das Gegenstück einer Metaphysik der Sitten, als das andere Glied der Eintheilung der praktischen Philosophie überhaupt, würde die moralische Anthropologie sein, welche, aber nur die subjective, hindernde sowohl als begünstigende Bedingungen der *Ausführung* der Gesetze der ersteren in der menschlichen Natur, die Erzeugung, Ausbreitung und Stärkung moralischer Grundsätze [...] und dergleichen andere sich auf Erfahrung gründende Lehren und Vorschriften enthalten würde” (MS, AA 06: 217). In general, in his published *Anthropology* as well as in his lectures on anthropology, Kant shows conspicuous sensitivity to the conflictual aspect of our moral psychology.

59 Allison: *Kant’s Theory of Freedom*, 126.

60 Reath: “Kant’s Theory of Moral Sensibility,” 296.

[S]o kann man doch in keinem Beispiel mit Gewißheit darthun, daß der Wille hier ohne andere Triebfeder, bloß durchs Gesetz, bestimmt werde, ob es gleich so scheint; denn es ist immer möglich, daß ingeheim Furcht vor Beschämung, vielleicht auch dunkle Besorgniß anderer Gefahren Einfluß auf den Willen haben möge. (GMS, AA 04: 419)

I see no reason to think that these kinds of causally potent obscure apprehensions – which might remind one of Leibnizian *petites perceptions* – necessarily involve maxims for the reason that they operate so far under the radar.⁶¹ Indeed, Kant paints a memorable picture of a cool observer, made shrewd and acute over the years, who knows how indiscernible the motives of our actions are to the human heart (GMS, AA 04: 407 f.),⁶² inclinations secretly (*ingeheim*) working against the moral law (KpV, AA 05: 86).

The present interpretation shows its strength also by offering sensible and direct solutions to a number of interpretative problems which have been noted in recent decades. First, Allison considers puzzling the notion of *degrees of freedom* to which Kant refers in his lectures on ethics.⁶³ To solve the issue, he suggests that Kant is voicing Baumgarten's view here instead of his own. But degrees of freedom appear in the *Metaphysics of Morals* as well: “Es ist ihm Pflicht: sich aus der Rohigkeit seiner Natur, aus der Thierheit (*quoad actum*), immer mehr zur Menschheit, durch die er allein fähig ist sich Zwecke zu setzen, empor zu arbeiten” (MS, AA 06: 387). “Je weniger der Mensch physisch, je mehr er dagegen moralisch [...] kann gezwungen werden, desto freier ist er” (MS, AA 06: 382n). On my interpretation, these statements are just as they should be: even though our choice is *absolutely* free, we can still be said to be *more or less* free according to

61 This of course raises the question of how implicit and unconscious the Kantian maxims may be. The second *Critique* provides one of the most sustained discussions on the nature of maxims; there, Kant says that as practical principles they are “Sätze, welche eine allgemeine Bestimmung des Willens enthalten” (KpV, AA 05: 19), and something “wir uns [...] entwerfen” (KpV, AA 05: 29); the examples he gives are such as “keine Beleidigung ungerächt zu erdulden” (KpV, AA 05: 19) and “mein Vermögen durch alle sichere Mittel zu vergrößern” (KpV, AA 05: 27). Thus here the paradigm cases of maxims are subjective but explicit and rather general principles of action. Clearly, not all of our actions are *consciously* governed by such principles (for Kant or in general). Now it may be argued that in spite of this, our actions are always governed by maxims of which we can at the very least *be made* aware, which can be articulated to us even if we are not immediately conscious of them (see e. g. Allison: *Kant's Theory of Freedom*, 90). But then again, as far as I know Kant says nothing to this effect, whereas, as I have shown above, there is plenty of textual evidence for the claim that many of our actions are not, for Kant, principled in *any* way – which would of course explain why so many of our actions are such that we are not aware of any maxims governing them.

62 See also MS, AA 06: 392f., 447; RGV, AA 06: 63; V-MS/Vigil, AA 27: 704.

63 Allison, Henry: *Essays on Kant*. Oxford 2012, 142. Allison refers to V-Mo/Collins, AA 27: 268.

the extent to which we are capable, through the struggle within, to keep in check the animal inside us – although there is little doubt that on Kant’s view, probably more often than we would like to think, it is the animal side that prevails.⁶⁴

Second, overlooking the fact that inclinations can directly determine our faculty of desire has made it difficult to determine Kant’s take on akratic action. The first propensity to evil, which Kant calls “die Gebrechlichkeit (*fragilitas*) der menschlichen Natur” (RGV, AA 06: 29), refers to weakness of the will:⁶⁵ “[I]ch nehme das Gute (das Gesetz) in die Maxime meiner Willkür auf; aber dieses, welches objectiv in der Idee (*in thesi*) eine unüberwindliche Triebfeder ist, ist subjectiv (*in hypothesi*), wenn die Maxime befolgt werden soll, die schwächere (in Vergleichung mit der Neigung)” (RGV, AA 06: 29). In other words, I endorse the (morally) better but still follow the worse. Now Allison has admitted that this propensity to evil poses an interpretative challenge to his reading, and especially to the Incorporation Thesis.⁶⁶ His attempt to meet that challenge leads to the complex interpretative move according to which *even here* we are dealing with a “freely adopted” propensity to evil, namely the first degree of *radical* evil, “the direct result of the original primacy granted to the claims of one’s sensuous nature”; because of *self-deception* one depicts “what is in reality a free evaluation on one’s part as a ‘weakness’ for which one is not responsible.”⁶⁷ But as Iain Morriison argues, “[o]ne problem with Allison’s solution is that it seems to deny the phenomenon of weakness; it denies that the agent is really committed to the moral action. [...] But, Kant’s account of the three degrees of evil [...] indicates that the individual really is committed to the (untaken) moral action in the first degree [...] and not in the second two.”⁶⁸ Moreover, and contrary to what Allison holds, I would say that succumbing to the inclinations is not a form of *radical* evil at all: only the third propensity to evil, resulting in a *reversal* of the ethical order,

⁶⁴ It should be noted that even then, given the structure of the struggle (and especially the fact that it involves, perhaps apart from certain unusual cases, free choice even when we act against that choice), the moral psychology involved is so complex that the phenomenon fundamentally differs from mere animal behavior; see e. g. V-Mo/Collins, AA 27: 345.

⁶⁵ See Frierson: *Kant’s Empirical Psychology*, 234n21. Kant (RGV, AA 06: 29) notes that this grade of evil “ist selbst in der Klage eines Apostels ausgedrückt: Wollen habe ich wohl, aber das Vollbringen fehlt.”

⁶⁶ Allison: *Kant’s Theory of Freedom*, 158.

⁶⁷ Allison: *Kant’s Theory of Freedom*, 159.

⁶⁸ Morriison, Iain: “On Kantian Maxims: A Reconciliation of the Incorporation Thesis and Weakness of the Will.” In: *History of Philosophy Quarterly* 22, 2005, 88n5. For similar criticism, see Nauckhoff, Josefine: “Incentives and Interests in Kant’s Moral Psychology.” In: *History of Philosophy Quarterly* 20, 2003, 60n14.

qualifies as radical.⁶⁹ By contrast, the present interpretation offers a straightforward account of the Kantian view of weakness of the will: it is simply about losing the struggle against inclinations so that we are incapable of acting according to the freely adopted maxim.

More generally, the present account offers tools for a nuanced but uncomplicated account of differences in moral culpability. It is one thing to act against the moral law from lack of moral strength – which is just want of virtue (*Untugend*) – and quite another to do so intentionally from a principled basis, which is a vice (*Laster*) (MS, AA 06: 390);⁷⁰ I would say that the latter involves a free prioritizing of a nonmoral maxim, the former “merely” a defeat in the battle against inclinations.⁷¹ Thus, there are two ways in which we can fail to act morally: either by losing to inclinations or by prioritizing nonmoral maxims. In positive terms, we should distinguish a *revolutionary change of heart* (that is, making moral maxims supreme) from a *gradual reformation* (that is, gradually becoming stronger in countering inclinations) (RGV, AA 06: 47). Even in what marks, as far as I have been able to detect, his most choice- and maxim-emphasizing moment – the beginning of *Religion II* (RGV, AA 06: 57 f.) – Kant does *not* deny that morality requires direct confrontation with inclinations, just that the *first* thing in acting morally is to dislodge the perverted maxims, whereas inclinations “erschweren nur die *Ausführung* der entgegengesetzten guten Maxime” (RGV, AA 06: 58n).

Third, Kant’s pedagogical views have been found puzzling with regard to maxims. In his lectures, Kant states not only that “[m]an muß dahin sehen, daß der Zögling aus eignen Maximen, nicht aus Gewohnheit gut handle, daß er nicht bloß das Gute thue, sondern es darum thue, weil es gut ist” (Päd, AA 09: 475) but also that “[m]an muß dahin sehen, daß das Kind sich gewöhne, nach

⁶⁹ The main purpose of the first part of the *Religion* is to explain how freely making maxims of inclination the condition of moral maxims “den Grund aller Maximen verdirbt” (RGV, AA 06: 37) and thereby reverses the moral order, which equals radical evil.

⁷⁰ See also V-MS/Vigil, AA 27: 691.

⁷¹ Since whether we win or lose that battle is not something we can choose, does this mean that the actions that ensue are not to be imputed to us? To this I would reply, largely following Frieron’s (*Kant’s Empirical Psychology*, 251 f.) insightful analysis (even though his take on the nature of human frailty, and of human agency more generally, differs from the one defended here), that they are imputable in an *indirect* sense: we can be guilty of (freely) fostering lack of virtue by not promoting, *before* the battle itself takes place, things that would enhance self-mastery and make us resistant to inclinations (in this regard, the optimal state is that of moral *apathy*; see KU, AA 05: 272; MS, AA 06: 408 f.). After all, Kant recommends what he calls *ethical gymnastics*, which “besteht also nur in der Bekämpfung der Naturtriebe, die das Maß erreicht, *über sie bei vorkommenden, der Moralität Gefahr drohenden Fällen* Meister werden zu können” (MS, AA 06: 485, emphasis added).

Maximen und nicht nach gewissen Triebfedern zu handeln” (Päd, AA 09: 480). Quite understandably, this has prompted Gressis to ask, “if people always act on maxims, then what does it mean to say that they also have to learn how to act on them?”⁷² As a solution, he recommends that we adopt Jens Timmermann’s interpretation, according to which Kant’s notion of a maxim is equivocal,⁷³ having three different senses so that “the maxims on which everyone acts (maxims in the thin sense) are not the same kinds of things as the maxims people have to learn to guide themselves by (particularly vigorous higher-order maxims).”⁷⁴ However, on the present interpretation there is no need to assume such an equivocation:⁷⁵ what Kant means is simply that children first act without maxims, from habits that arise from inclinations, and that they have to be educated so that they begin to act from maxims – and as morality is the goal of education, from good maxims.

Conclusion

I have argued that placing Kant’s views on moral agency within the framework formed by his theory of faculties and general theory of action allows us to see that the key question concerns what determines that feature of ours through which we act in the world, namely our faculty of desire. That faculty can, in turn, be determined in a way that is not in any way rational; inclinations and instinct can, and often do, determine us to act purely desirously. However, the widespread tendency to build on the so-called Incorporation Thesis has led us to neglect this and to focus on the way in which inclinations can determine us insofar as we have rationally chosen to incorporate them into our practical principles. Hence, acknowledging the direct impact inclinations can have on us requires revising our view of Kantian moral agency: free endorsement of moral maxims is certainly a crucial element of being a moral agent, but it needs to be accompanied by virtue so that we act according to the endorsed maxims against conflicting inclinations, which results in our faculty of desire’s being determined by moral principles and not by inclinations. While this broader theory certainly still leaves open questions concerning the precise nature of the conflict – a detailed analysis of the way

⁷² Gressis, Rob: “Recent Work on Kantian Maxims I: Established Approaches.” In: *Philosophy Compass* 5, 2010, 219.

⁷³ See Timmermann, Jens: “Kant’s Puzzling Ethics of Maxims.” In: *The Harvard Review of Philosophy* 8, 2000, 39–52.

⁷⁴ Gressis: “Recent Work on Kantian Maxims II,” 234.

⁷⁵ Gressis (“Recent Work on Kantian Maxims II,” 235) admits that Kant “never says that there is more than one kind of maxim.”

in which the determination of our faculty of desire comes to be decided remains to be given – it not only squares well with Kant’s combative language and his view that the real grounds of our actions are largely inscrutable to us, but also offers direct solutions to a number of problems encountered in the literature. I have focused on what Kant actually says and thinks on the topic; to conclude, however, I would not downplay the philosophical plausibility of his theory of moral agency, understood as above. Perhaps the most attractive feature of that theory is that it depicts us solidly embedded in the world, acknowledging the struggles that come with that embeddedness just as forcefully as the requirement that we rise above what is merely natural, or animalistic, in us by acting from the principles of pure practical reason.⁷⁶

76 I would like to thank Hanne Appelqvist, Carolyn Benson, John Carriero, Michael Cholbi, Barbara Herman, Veikko Launis, Peter Myrdal, Sakari Ollitero, Ville Paukkonen, Dominik Perler, Juhani Pietarinen, Arto Repo, Hannu Salmi, Jukka Sarjala, Ariane Schneck, Ninni Suni, Frans Svensson, Miira Tuominen, and Reed Winegar for their helpful remarks on earlier versions of this paper. Moreover, I would like to acknowledge that research for this article has been financially supported by the Academy of Finland (project number 275583).