

ME, MY WILL, AND I.
KANT'S *REPUBLICAN* CONCEPTION
OF FREEDOM OF THE WILL
AND FREEDOM OF THE AGENT

PAULINE KLEINGELD*

ABSTRACT · Kant's theory of freedom, in particular his claim that natural determinism is compatible with absolute freedom, is widely regarded as puzzling and incoherent. In this paper I argue that what Kant means by 'freedom' has been widely misunderstood. Kant uses the definition of freedom found in the republican tradition of political theory, according to which freedom is opposed to dependence, slavery, and related notions – not to determinism or to coercion. Discussing Kant's accounts of freedom of the will and freedom of the agent in turn, I argue that this insight sheds new light on Kant's transcendental compatibilism and suggests novel responses to age-old objections.

KEYWORDS · Immanuel Kant; Character; Freedom; Maxim; Moral imputation; Natural determinism; Republicanism; Transcendental compatibilism; Transcendental idealism; Will.

1. INTRODUCTION

IMMANUEL KANT'S position in the free will debate is widely viewed as mysterious, inconsistent, or both. Kant famously defends the thesis of thoroughgoing natural determinism and even goes so far as to claim that we can in principle «calculate» a human being's future conduct «with certainty, as in the case of a lunar or solar eclipse» (*KpV*, AA v 99). He also assumes, however, that «every intentional action» has a «free causality as its ground» (*KpV*, AA v 100). *And* he rejects the dominant strategy for reconciling freedom and determinism, which is to define freedom in opposition to external coercion and obstruction, rather than in opposition to natural determinism. Kant insists that freedom should be conceived as «absolute spontaneity», that is, as the power to act independently of natural causes, and that this cannot be located within nature (*KpV*, AA v 96-100).

This combination of claims looks puzzling, and Kant's move of locating freedom in an unknowable 'noumenal' realm is usually not seen as very helpful, to put it mildly. Even sympathetic interpreters find it impossible fully to reconcile Kant's commitment to natural determinism with his defense of absolute freedom.¹ This verdict is reflected in the fact that Kant's position on the matter is not seen as a live philosophical option in the current free will debate.²

* pauline.kleingeld@rug.nl, University of Groningen.

¹ For details and references, see section 2. 3 below.

² This is illustrated by the fact that surveys of the free will debate either treat Kant's position as a 'mysterious' view of merely historical interest (R. KANE, *A Contemporary Introduction to Free Will*, Oxford, Oxford University Press, 2005, pp. 42-44) or do not discuss the view in any detail at all (e.g., M. GRIFFITH, *Free Will: The*

In this essay, I argue that the difficulties surrounding Kant's theory of freedom stem largely from the fact that his *notion of freedom* has been misunderstood. Given that he refuses to define freedom in opposition to coercion and obstruction, it is widely taken for granted that he defines it in opposition to natural determinism. It has gone unnoticed that Kant instead employs a third, *republican* notion of freedom, according to which freedom is defined in opposition to slavery, dependence and domination. Since Kant means something different by 'freedom' than most readers assume,¹ it is not surprising that his theory has struck many as strange and incoherent. To illustrate the advantages of a republican reading of Kant's notion of freedom, I will show how it solves some of the most notorious problems diagnosed in the literature.

I will also argue for a second interpretive thesis, namely that we should distinguish between freedom of the *will* and freedom of the *agent*. The two are often conflated, which obscures Kant's analysis of the relation between natural determinism and noumenal freedom.

I first present the common understanding of Kant's notion of freedom, sketch his transcendental idealist defense of freedom, and outline the main difficulties associated with it in the literature (section 2). I then turn to the republican notion of freedom and show that Kant uses it in his political theory (section 3). In section 4, I argue that Kant employs the very same notion of freedom in his discussion of freedom of the will, and I illustrate how, on a republican reading, several canonical objections evaporate (section 4). In section 5, I turn to the absolute freedom of the agent. I argue that a republican reading sheds new light on Kant's conception of absolute freedom and its compatibility with natural determinism (section 5).

2. KANT'S POSITION ON FREEDOM AND DETERMINISM, AND THE PROBLEMS IT IS THOUGHT TO FACE

2.1. *Kant's Critique of the 'Comparative' Notion of Freedom*

As is familiar, Kant argues both that without freedom there is no genuine moral obligation or moral responsibility and that freedom cannot exist within the realm of natural determinism.² Formulating the 'Consequence Argument' *avant la lettre*, he explains that in a deterministic world my actions are never within my control, and I am therefore never free:

For, [natural determinism] implies that every event, and consequently every action that occurs at a certain point in time, is necessary under the condition of what preceded it. Now, since the past is no longer within my control, every action that I perform must be necessary because of

Basics, New York, Routledge, 2013; M. MCKENNA, D. PEREBOOM, *Free Will: A Contemporary Introduction*, New York, Routledge, 2016; T. O'CONNOR, C. E. FRANKLIN, *Free Will*, «Stanford Encyclopedia of Philosophy» [Spring 2020 Edition], ed. by E. N. Zalta, <https://plato.stanford.edu/archives/spr2020/entries/freewill/>).

¹ There is at least one exception. After I presented this paper at a conference, I learned that Sven Nyholm has also briefly suggested that Kant's notion of freedom is republican (S. NYHOLM, *Revisiting Kant's Universal Law and Humanity Formulas*, Berlin-Boston, de Gruyter, 2015, pp. 9n., 25, 115).

² In one footnote, Kant calls 'predeterminism' preferable to 'determinism' (RGV, AA VI 49-50n.), since, on his view, a free will also needs to be 'determined', namely determined by reason. I generally use 'natural determinism' to accommodate this point.

determining grounds *that are not within my control*, that is, at the point in time at which I act I am never free.

(*KpV*, AA v 94)

Kant acknowledges that this conclusion is challenged by those who contest the very notion of freedom it presupposes. On what he calls their «comparative» (non-absolute) conception of freedom (*KpV*, AA v 96), freedom is compatible with natural determinism. These compatibilists distinguish between free and unfree doings *within* the realm of natural determinism. Free actions, on their conception, are those deterministically caused actions that result from natural causes *internal* to the agent. The contrast class is formed by cases in which the agent is moved or obstructed by *external* causes (though perhaps mediated by internal processes), such as cases of coercion, manipulation, and impediments of various sorts.

Kant argues that this comparative notion of freedom is entirely insufficient for moral responsibility and that it is not the «proper sense» of the term (*KpV*, AA v 97). He famously denounces it as «the freedom of a turnspit» (*KpV*, AA v 96-97). Surely there is a possible (comparative, non-absolute) sense of 'freedom' in accordance with which we can say that a turnspit turns around freely, namely when nothing else hinders its movement and, «once it is wound up, it carries out its movements of itself» due to its own inner mechanism (*KpV*, AA v 97). But we do not hold a turnspit morally responsible for turning on account of its turning 'freely' in this sense. In the same sense, we can say that a human being acts freely when it is moved to act by its own inner psychological mechanism (*KpV*, AA v 96). But just as we do not hold the turnspit morally responsible for turning, we should not hold a human agent morally responsible on account of his acting 'freely' *in this sense*. If human actions are the result of natural determinism, whatever the human being does or does not do is *necessary* given the laws of nature and causal antecedents. Then actions are *not within the agent's control*, and the agent is not morally responsible (*KpV*, AA v 97).

Kant's rejection of the 'comparative' understanding of freedom is also clear from a number of passages that must sound like utter provocations to the ears of traditional compatibilists. In discussing the famous 'gallows example' in the *Critique of Practical Reason*, for instance, Kant asserts that if a ruler threatens to hang a man unless he provides false testimony against an innocent person, it is *then* that the man becomes aware of his freedom (*KpV*, AA v 30) – even though this is a clear case of external coercion.

Given Kant's criticism of the compatibilist 'comparative' notion of freedom and his defense of 'absolute' freedom, and on the assumption that there are only two options, the vast majority of interpreters assume that what he means by 'freedom' is, or in any case involves, the ability to choose among alternative possibilities in the empirical world.¹

¹ A small number of authors have proposed a compatibilist reading of Kant along Davidsonian lines (e.g., R. MEERBOTE, *Kant on the Nondeterminate Character of Human Actions*, in *Kant on Causality, Freedom, and Objectivity*, ed. by W. A. Harper, R. Meerbote, Minneapolis, University of Minnesota Press, 1984, pp. 138-163; H. HUDSON, *Kant's Compatibilism*, Ithaca, Cornell University Press, 1994), following Davidson's own suggestions in his essay *Mental Events* (D. DAVIDSON, *Essays on Actions and Events*, Oxford, Clarendon Press, 1980, pp. 207-227). The main problem for this reading is that it does not allow noumenal agency to have an effect in the phenomenal world, although the latter idea is clearly central to Kant's position. Consequently, this reading has not gained many adherents.

2. 2. Kant's Theory of Freedom

If absolute freedom is understood to involve genuine alternative possibilities in the phenomenal world, however, it becomes hard to see how Kant can reconcile his defense of it with his commitment to natural determinism. Yet he claims that his transcendental idealism accomplishes just this. The goal of this section is to sketch a foil against which the diagnosed difficulties, outlined in 2. 3, can be understood. Such a sketch is a rather precarious undertaking, however, given the large number of interpretive debates and the deep divisions in Kant scholarship concerning the proper understanding of Kant's transcendental idealism. I here present one possible line of interpretation, but the difficulties diagnosed in 2. 3 are not uniquely associated with the particular account I present here.¹

Kant's general transcendental idealist approach in the *Critique of Pure Reason* is guided by the thought that the mode of operation of our cognitive powers determines certain necessary features ('formal conditions') of the world as it appears to us.² To use an anachronistic analogy:³ a functioning black-and-white TV will transform any signal it receives into a two-dimensional display in black, white, and/or grey. If we know this, we also understand that whatever appears on the screen will necessarily be two-dimensional and on the white-grey-black spectrum. Analogously, if we know how our cognitive powers operate, we have a priori knowledge of the necessary features of the world as it appears to us (the world as we cognize it).

Kant argues that this guiding thought immediately forces the distinction between «things as they appear to us» (phenomena) and their underlying ground, «things as they are in themselves» (noumena). Kant calls the latter «noumena» because they can only be *thought*. Per definition, the ground of appearances is cognitively inaccessible to us, since it is precisely the thing as it is independently of our mode of cognition. We arrive at the thought of this «unknown something» by abstracting from the way we cognize (*KrV*, A 235-260 B 294-315).⁴

One of Kant's core claims is that his approach provides the philosophical underpinning of the foundational, metaphysical principles of (Newtonian) natural science. Consider the principle that every event has a cause which determines it necessarily, in accordance with natural laws (e.g., *KrV*, A 201 B 246-247). If this principle derives from our mode of cognition, then it is true a priori, of the world as we know it, that every event has a cause which determines it necessarily in accordance with natural laws. Put differently, in that case natural determinism is a necessary feature of the world as it

¹ For an overview of different interpretive positions, with special attention to their relevance to Kant's theory of freedom, see E. WATKINS, *Kant and the Metaphysics of Causality*, Cambridge, Cambridge University Press, 2005, ch. 5. For a discussion of the methodology of transcendental argument, see G. GAVA, *Kant, the Third Antinomy and Transcendental Arguments*, «Pacific Philosophical Quarterly», c. 2, 2019, pp. 453-481.

² See the very beginning of the «Introduction» to the *Critique of Pure Reason* (B version).

³ It is important, here and below when I return to this analogy, only to focus on the point for which the analogy is used. There are many respects in which the cases are *not* analogous. To mention an obvious disanalogy: in the TV case, the 'ground' of the image on my screen can be identified and described (actors in a studio, computer-generated imagery, etc.).

⁴ For Kant's statements that phenomena, as such, must have «grounds» in noumena, see *KrV*, A 537 B 565, A 538 B 566, A 540 B 568, A 545 B 573, A547-548 B 575-576; see also *GMS*, AA IV 453.

appears to us. And in the *Critique of Pure Reason*, Kant provides a lengthy argument in support of the claim that this is indeed the case. This motivates his claim that we can in principle 'calculate' the future behavior of human agents.

Kant also points out, however, that his approach makes it *not contradictory* to combine the claim that the natural world is thoroughly deterministic with the idea that human beings have absolute freedom. As he argues in his discussion of the «Third Antinomy», *knowledge* of noumena is per definition impossible, but it is not impossible to *conceive* or *think* of noumena in a certain way. He claims that this makes room for the idea of absolute freedom (*KrV*, A 538-541 B 566-569). To return to the earlier analogy: if I know that whatever image appears on my TV screen must be black, white or grey, due to my TV's mode of operation, this does not make it contradictory for me to conceive of the ground of an image on my screen (say, a flower) as having color, even if, given merely my TV set and my knowledge of its operation, this ground remains entirely inaccessible to me.

Kant subsequently takes the further step of arguing, still in his discussion of the «Third Antinomy», that we may legitimately *assume* that we have absolute freedom. We happen to have a strong ground for assuming that we do, namely our consciousness of moral obligation.¹ Here we believe ourselves to be unconditionally obligated to act in accordance with moral laws even if doing so runs counter to all of our sensible desires. On the basis of our consciousness of moral obligation we judge that we *can* act in accordance with moral laws, which means that we assume that we can act independently of our sensible desires, that is, freely. Given that his transcendental idealism has room for this assumption, Kant argues, we may indeed believe we have absolute freedom.²

This belief clearly includes the idea that we can act on the basis of something other than sensible desires, and this, Kant writes, «we call reason» (*KrV*, A 547 B 575). He writes:

Now that this reason has causality, or that we at least represent to ourselves something of the sort in it, is clear from the *imperatives* that we give our executive powers as rules in everything practical. The *ought* expresses a type of necessity and a connection with grounds that does not occur anywhere in the whole of nature.³

(*KrV*, A 547 B 575)

¹ Although I cannot discuss this here, I believe that Kant defends a similar view in the *Groundwork* and the *Critique of Practical Reason* and that there is more continuity in his argument than is sometimes thought. See P. KLEINGELD, *Moral Consciousness and the Fact of Reason*, in *A Critical Guide to Kant's Critique of Practical Reason*, ed. by J. Timmermann, A. Reath, Cambridge, Cambridge University Press, 2010, pp. 55-72 for Kant's argument in the *Critique of Practical Reason*; see S. TENENBAUM, *The Idea of Freedom and Moral Cognition in Groundwork III*, «Philosophy and Phenomenological Research», LXXXIV, 3, 2012, pp. 555-589 and O. WARE, *Kant's Deductions of Morality and Freedom*, «Canadian Journal of Philosophy», XLVII, 1, 2017, pp. 116-147 for the continuities in Kant's position.

² On the epistemic status of this 'assumption', which is subjectively certain (although it lacks objective justification) and hence qualifies as belief (holding-to-be-true) but not as theoretical knowledge, see *KrV*, B xxviii-xxx; A 820-831 B 848-859, *KpV*, AA v 142-146, *WDO*, AA VIII 137, *MS*, AA VI 354; P. KLEINGELD, *The Conative Character of Reason in Kant's Philosophy*, «Journal of the History of Philosophy», xxxvi, 1, 1998, pp. 77-97; M. WILLASCHKE, *Freedom As a Postulate*, in *Kant on Persons and Agency*, ed. by E. Watkins, Cambridge, Cambridge University Press, 2018, pp. 102-119.

³ I bracket the issue of whether Kant is speaking here of both moral and prudential imperatives or of moral imperatives alone.

Furthermore, Kant writes that we sometimes regard actual human actions as having been performed on the basis of «grounds of reason»:

At times ... we find, or at least believe we find, that the ideas of reason have actually proved their causality in regard to the actions of human beings as appearances, and that these actions have occurred not because they were determined by empirical causes, no, but because they were determined by grounds of reason.

(*KrV*, A 550 B 578)

Kant here claims – stressing in the first clause of both quoted passages that this is not a matter of theoretical knowledge – that, in moral contexts, we regard reason as a cause of actions.

Kant claims that his transcendental idealism can accommodate this thought by relating moral obligation and imputation to the *noumenal ground* of the actions, namely to the agent as a thing in itself. He emphasizes the difference between the *causal explanation* and the *moral imputation* of a phenomenal action. When we give a causal explanation, we explain the phenomenal action in terms of the *natural laws* and antecedent conditions that caused it to happen. When we impute an action to an agent and evaluate it morally, by contrast, we judge the noumenal agent in light of *moral laws* and assume that he can act independently of natural causes.

As an example, Kant mentions the case of a malicious lie. He asserts that we can *explain the agent's lie* in terms of *natural laws* and empirical causal antecedents, such as his natural psychological tendencies, upbringing, bad company, and so on (*KrV*, A 554-555 B 582-583). When we nevertheless hold him morally responsible for the lie and blame him, we do something entirely different: we *judge the agent* in light of *moral laws* and presuppose that he «could have and ought to have» behaved differently:

This blame is based on a law of reason, whereby one regards reason as a cause that, regardless of all the empirical conditions just mentioned, *could have and ought to have determined* the conduct of the human being *differently*.

(*KrV*, A 555 B 583, emphasis added; cf. *KpV*, AA v 95-96)

We judge that his lie «stood directly under the power of reason» and was «determined independently of sensibility» (*KrV*, A 556-557 B 584-585).

2.3. Common Diagnoses of Important Difficulties

To many readers, Kant's position as sketched above seems strained, if not downright incoherent. In the liar example, Kant seems to assume the existence of alternative empirical possibilities – to lie or not to lie. Hence Kant is taken to assume that we are «free to act in ways other than we are causally determined to act», that is, to employ a «contra-causal» notion of freedom.¹ But this seems incompatible with the natural determinism he defends.

Most interpreters hold that freedom, according to Kant, consists in an *ability to choose*. The relevant choice can be conceived as the choice between moral and im-

¹ H. E. ALLISON, *Kant's Theory of Freedom*, Cambridge, Cambridge University Press, 1990, pp. 43, 52.

moral actions,¹ moral and immoral *maxims*, or as the choice of one's *most fundamental maxim* – either the maxim of duty or the maxim of self-love – which guides one's adoption of lower-level maxims.² Other interpreters defend the view that the relevant ability should be conceived as *asymmetrical*. On their view, freedom, as independence from sensibility, consists not in the ability to choose between moral and immoral options, but in the capacity to choose to do what morality demands. This thus restricts the range of options to morally *right* actions or maxims.³

Despite these differences concerning interpretive details, there is wide agreement in the literature that Kant's reconciliation of freedom with natural determinism is unsuccessful. If the thesis of determinism is true, then all my future behavior is already predetermined. This makes it hard to see how I could ever avoid future moral failures, and if I cannot, then it will not only be *untrue* to say, later, that I 'could have and ought to have' behaved differently but also *unjust* to blame *me* for doing what I did. Conversely, if two or more alternative empirical actions are within my control (say, to lie or not to lie), then it is hard to see how this can be squared with Kant's commitment to natural determinism and the *predictability* of my actions. Furthermore, given natural determinism, it seems that the ability to act otherwise would require, rather implausibly, an ability to change the laws of nature and/or the course of history. After all, different phenomenal options would require different causal antecedents, since otherwise the chosen action would lack a sufficient causal explanation.⁴

Equally puzzling, in the eyes of many, is the fact that Kant explicitly *denies* that freedom consists in the capacity to choose whether or not to act morally (MS, VI 226). It would seem that when Kant blames a liar for lying, he must be assuming that his freedom consists precisely in his capacity to choose whether or not to lie. So it seems unclear why he would deny this.

Within Kant scholarship, some have addressed these problems by arguing that Kant's commitment to determinism is weaker than it seems. For example, it has been argued that Kant's discussion of the «Second Analogy» in the *Critique of Pure Reason* does not rule out that some phenomena (viz., human actions) have free causes.⁵ Others have argued that Kant considers the principle of natural causal determinism to be merely regulative,⁶ or that his defense of determinism leaves open alternative

¹ E.g., M. KOHL, *Kant on Determinism and the Categorical Imperative*, «Ethics», CXXV, 2, 2015, pp. 331-356; B. VILHAUER, *Immanuel Kant*, in *The Routledge Companion to Free Will*, ed. by K. Timpe, M. Griffith, N. Levy, New York, Routledge, 2017, pp. 343-355.

² R. BADER, *Kant on Freedom and Practical Irrationality*, in *The Idea of Freedom: New Essays on the Interpretation and Significance of Kant's Theory of Freedom*, ed. by D. Heide, E. Tiffany, Oxford, Oxford University Press, forthcoming.

³ A. W. WOOD, *Kant's Compatibilism*, in *Self and Nature in Kant's Philosophy*, ed. by A. W. Wood, Ithaca, Cornell University Press, 1984, pp. 73-101 (77-82); J. TIMMERMANN, *Kant's Groundwork of the Metaphysics of Morals: A Commentary*, Cambridge, Cambridge University Press, 2007, pp. 164-167.

⁴ Cf. J. TIMMERMANN, *Sittengesetz und Freiheit: Untersuchungen zu Immanuel Kants Theorie des freien Willens*, Berlin, de Gruyter, 2003, p. 113; WOOD, *Kant's Compatibilism*, p. 92; WATKINS, *Kant and the Metaphysics of Causality*, ch. 5; VILHAUER, *Immanuel Kant*.

⁵ M. WOLFF, *Kant über Freiheit und Determinismus*, in *Kants Metaphysik der Sitten in der Diskussion*, ed. by B. Tuschling, W. Euler, Berlin, Duncker & Humblot, 2013, pp. 27-42 (27-30).

⁶ J. BOJANOWSKI, *Ist Kant ein Kompatibilist?*, in *Sind wir Bürger zweier Welten? Freiheit und moralische Verantwortung im transzendentalen Idealismus*, ed. by M. Brandhorst, A. Hahmann, B. Ludwig, Hamburg, Meiner, 2012, pp. 59-76 (74).

possibilities for the future.¹ On these lines of interpretation, however, it is impossible fully to integrate Kant's repeated statements that it is in principle possible to calculate future human behavior (*KpV*, AA v 99, and similar claims elsewhere, e.g., *KrV*, A 539 B 567, A 549-550 B 577-578).

A further notorious difficulty is the fact that Kant identifies «a free will» with «a will under moral laws» (*GMS*, IV 446-447). This passage is often read as meaning that only a morally good will is a free will, or that a free will is restricted to choosing what is good, and then the statement leads to an obvious objection. It seems to entail that agents who act impermissibly do not have a free will. But if they do not, they cannot be morally responsible for their bad actions. This classic objection is known as the Reinhold–Sidgwick objection (although, ironically, Reinhold in fact presents it as an objection made by *others* and aims to defend Kant against it).² In response, Henry Allison and Jochen Bojanowski, among others, have pointed out that Kant is not philosophically committed to the problematic position attributed to him, and that he speaks of a will *under* moral laws rather than a good will.³ Nevertheless, the wording of the passage at issue remains cryptic.

Given that these difficulties constitute serious problems, it is not surprising that leading commentators view Kant's theory of freedom as a «mystery»⁴ or as «utterly unacceptable»,⁵ with some concluding that «[u]ltimately, even the most charitable interpretation of Kant's attempt to reconcile the closed causal system of natural determinism and free will is bound to fail». Nor is it surprising that Kant's theory of freedom does not have any champions in the current free will debate.⁷

3. KANT'S REPUBLICAN NOTION OF FREEDOM

What has gone unnoticed in the literature is that Kant in fact employs a *republican* notion of freedom. Rather than opposing freedom to natural determinism or to external coercion, he opposes it to *dependence, slavery, domination, despotism* and related notions. Positively speaking, and in line with core ideas of the republican tradition in political theory, he sees freedom as *independence, being one's own master, having control, and having autonomy*. This reading of Kant's notion of freedom sheds new light on his conception of freedom of the will and freedom of agents, as I argue in subsequent

¹ L. ALLAIS, *Manifest Reality: Kant's Idealism and his Realism*, Oxford, Oxford University Press, 2015, pp. 303-308.

² K. L. REINHOLD, *Briefe über die Kantische Philosophie*, ed. by M. Bondeli, Basel, Schwabe, 2008 [1792], pp. 185-187 [267-269]; see also P. GUYER, *The Struggle for Freedom: Freedom of Will in Kant and Reinhold*, in *Kant on Persons and Agency*, ed. by E. Watkins, Cambridge, Cambridge University Press, 2018, pp. 120-137.

³ ALLISON, *Kant's Theory of Freedom*, pp. 39-40, 94-99, 133-136; J. BOJANOWSKI, *Kants Theorie der Freiheit: Rekonstruktion und Rehabilitierung*, Berlin, de Gruyter, 2006, chs. 7-8.

⁴ KANE, *A Contemporary Introduction to Free Will*, p. 44.

⁵ A. W. WOOD, *Kantian Ethics*, Cambridge, Cambridge University Press, 2008, p. 141.

⁶ TIMMERMANN, *Kant's Groundwork*, p. 166.

⁷ Christine Korsgaard's fictionalist account is an exception. She reads Kant as saying that the knowledge that one's choices are determined must be «simply ignored» in acting and that «the point is not that you must believe that you are free, but that you must choose *as if* you were free». She adds that «you will believe that your decision is a sham, but it makes no difference» (C. KORSGAARD, *Creating the Kingdom of Ends*, Cambridge, Cambridge University Press, 1996, pp. 162-163). It is hard to find this view in Kant's texts, though, not to mention the problem that believing my decision to be a sham also means believing my moral responsibility to be a sham.

sections. Here, I first explicate the republican notion of freedom and show that Kant uses it in his political philosophy.

The republican conception of freedom as non-domination or independence is well known by now in discussions of political freedom. Over the past few decades, Quentin Skinner¹ and Philip Pettit,² among others, have re-introduced this conception into debates in political philosophy. On the republican view, political freedom consists not in having options, and not in non-interference, but in non-domination – more specifically, in not *being under the power* of the *arbitrary* choice of *another*.

Slavery is the paradigmatic example of unfreedom. An enslaved person is subject to the jurisdiction of a 'master'. The master has unilateral power over the slave and can decide to use the latter at his discretion. The slave is unfree regardless of whether the master actively coerces him; having a so-called 'benevolent master'³ does not turn the slave into a free person. Thus, unfreedom does not so much consist in not being coerced as in having no protection against coercion. Unfreedom consists in not being independent of arbitrary coercion. In the republican tradition, this way of conceiving of unfreedom is extended to other forms of subordination. The subjects in a despotic state, who are subordinated to the arbitrary will of a despot, are unfree. Those living under the dominion of a colonial power are unfree. Oppressed groups may fight for 'liberation' and 'emancipation' – again in terminology associated with the master-slave relation.

Those within the republican tradition of political philosophy typically hold that freedom and equality are two sides of the same coin, meaning that the genuine freedom of citizens can be secured only within a community of equals. This is the republic. The republican tradition includes different views, however, as to the political institutions required to ensure freedom. Among early modern republican theorists, some, such as the Federalists, argued that freedom requires a system of checks and balances, to be instituted by public law. Others, such as Jean-Jacques Rousseau, emphasized that public laws must issue from collective legislation by the citizens themselves. All brands of republicanism agree, however, that freedom is opposed to subjection to the *arbitrary* (discretionary) power of *another* and that it requires the rule of law.

Note that the republican ideal of freedom and equality *among citizens* can be combined with the denial of citizenship status, partly or fully, to the vast majority of people affiliated with a particular republic. Children, women, enslaved persons, colonized peoples, Jews, foreign residents, certain racialized groups, and economically dependent males are some of the categories who were regularly excluded from full citizenship status in early modern republican theory and practice – with many individuals falling under several of these categories simultaneously.

Kant clearly belongs to the republican tradition of political theory, and he regards both the separation of powers and collective self-legislation as essential. In the *Metaphysics of Morals* he equates «freedom» with «independence from being compelled by

¹ Q. SKINNER, *The Idea of Negative Liberty: Philosophical and Historical Perspectives*, in *Philosophy in History: Essays in the Historiography of Philosophy*, ed. by R. Rorty, Q. Skinner, J. B. Schneewind, Cambridge, Cambridge University Press, 1984, pp. 193-221; IDEM, *Liberty before Liberalism*, Cambridge, Cambridge University Press, 1998.

² P. PETTIT, *Republicanism: A Theory of Freedom and Government*, Oxford, Oxford University Press, 1997.

³ 'So-called' since if he were really benevolent, he would not have slaves.

another's choice» (MS, AA VI 237). In the «Doctrine of Right», he argues that the only way for each and every ('active') citizen to be free in this sense – that is, the only way for citizens to avoid being subject to the power of another who rules over them at his personal discretion – is for all citizens to unite and collectively give public laws to which they are all equally subject. In order to enable the separation of powers, legislation should take place via a representative system rather than a (direct) democracy. Citizens should have the right to vote for representatives and legislate through these delegates (MS, AA VI 314). Kant holds that this political system, which he calls «a pure republic», is «the only constitution that accords with right» (MS, AA VI 340).

The freedom of the citizens, thus conceived, consists in their living under their own laws. Their freedom consists neither in the absence of laws nor in the ability to decide whether to obey the law, let alone in their being able to break the law. They *can* break the law, of course, but this is not what their political freedom consists in. Their freedom consists in their autonomy, that is, in the fact that the laws which they ought to obey are *their own* laws rather than the heteronomous dictates of a despot.

Kant only started to defend active citizen voting rights in the 1790s, but his commitment to republicanism and the associated notion of freedom are found throughout his work from the Critical period (and before). For example, it is evident in his *Feyerabend Lectures on Natural Law* (1784), which he held during the months in which he was writing the *Groundwork* (e.g., V-NR/*Feyerabend*, AA XXVII 1383), and it is found in the political analogies that Kant uses in the *Groundwork* and the *Critique of Practical Reason*, as will become clear below. During the 1780s, the *idea* of the republic served as a counterfactual normative criterion to be used by an enlightened autocrat: the ruler should give laws to the people that the people «could» adopt themselves.¹ For the purposes of this paper, it is not necessary to pursue the development of Kant's political theory. What matters here is merely that Kant's conception of political freedom clearly stands in the republican tradition.

Note, however, that Kant restricts full ('active') citizenship status, even in the *Metaphysics of Morals*, to (economically, socially, civilly) independent adult males (MS, VI 314-315). While he argues that men ought to have the opportunity to work their way up to independence and active citizenship, Kant flatly denies this right to women, with reference to the «natural superiority» of men (MS, VI 279, 314). This matches his description of the distinct «characteristics» and separate «vocation» of women in his Anthropology lectures (e.g., *Anth*, VII 303-311). In the early *Observations on the Feeling of the Beautiful and the Sublime* (1764), Kant wrote: «I hardly believe that the fair sex is capable of principles», claiming that women have a separate kind of virtue that involves «nothing of ought, nothing of must, nothing of obligation» (GSE, AA II 232). Kant does not seem to have ever completely renounced this assessment. In order to avoid representing his views as more egalitarian than they seem to have been, I will therefore follow his use of male pronouns.²

¹ WA, AA VIII 39; see P. KLEINGELD, *Moral Autonomy as Political Analogy: Self-legislation in Kant's Groundwork and the Feyerabend Lectures on Natural Law*, in *The Emergence of Autonomy*, ed. by S. Bacin, O. Sensen, Cambridge, Cambridge University Press, 2019, pp. 158-175 (169-170).

² See P. KLEINGELD, *On Dealing with Kant's Sexism and Racism*, «SGIR Review», II, 2, 2019, pp. 3-22 for further references and discussion.

4. REPUBLICAN FREEDOM OF THE WILL

The new prominence of the republican notion of freedom in current political theory has not yet led to its reintroduction into the free will debate. In the early modern free will debate, however, the republican understanding of freedom was quite common, and Kant is a case in point.¹ There are striking terminological similarities between his discussion of political freedom, on the one hand, and his discussion of freedom of the will, on the other.²

I discuss freedom of the will and freedom of the agent in turn, although these topics are of course intimately related. Kant conceives of the will as an agent's capacity to act on the basis of reasoning. The issue of freedom of the will concerns the determination of the *content* of an agent's will, that is, its end or object. For Kant, the question here is whether reason alone can determine the object of the will, *independently of sensible desire*. The issue of freedom of the agent concerns the *causality* of the agent in determining his will. The question here is whether an agent can act *independently of natural determinism*.

Attributions of moral responsibility, according to Kant, presuppose both the freedom of the will and the absolute freedom of the agent in acting. I aim to show that in each case Kant uses a republican notion of freedom. Kant also phrases the distinction in terms of that between «practical freedom» and «transcendental freedom». He explicates practical freedom in terms of freedom of the will (*KrV*, A 534 B 562; A 800-804 B 828-832; *KpV*, AA v 167), and he equates transcendental and absolute freedom (*KpV*, AA v 98). Because Kant's discussion of 'practical freedom' is associated with exegetical debates concerning the relation between the «Dialectic» and the «Canon» of the first *Critique*, which I do not have the space to address here, I bracket this term in this paper. I discuss transcendental or absolute freedom in section 5.

4.1. *Freedom of the Will*

Imputing an action *to me* presupposes that my will is not determined by my sensible desires in accordance with the «mechanism of nature», Kant writes. For if it were, «my will would not be my own but the will of nature» (*V-NR/Feyerabend*, AA xxvii 1322). I could not be morally responsible for willing what I will, since my will would be determined by sensible desires resulting from natural causal processes that extend infinitely into the past before I even existed. And if I am not responsible for the content of my will, then I cannot be responsible for performing the corresponding actions. Thus, the moral imputation of actions to an agent presupposes that the agent's will is free. But what does Kant mean by freedom of the will?

¹ For an important contemporaneous example, see Rousseau's statement that «moral freedom... alone makes man truly the master of himself; for the impulsion of mere appetite is slavery, and obedience to the law one has prescribed to oneself is freedom» (J. J. ROUSSEAU, *The Social Contract*, in *The Social Contract and Other Later Political Writings*, ed. by V. Gourevitch, Cambridge, Cambridge University Press, 1997 [1762], book 1, ch. 8.3).

² The republican context of Kant's notion of freedom also illuminates other aspects of his moral theory, such as his prohibition of using persons 'merely as a means'. Kant, like many others in the republican tradition, uses this locution to express the opposite of freedom (e.g., *V-MP-L1-Pölitz*, AA xxviii 268).

Kant describes freedom of the will in terms of the republican contrast between independence and dependence, freedom and slavery, being one's own master and being subservient to another. He consistently conceives of the 'freedom' of the will, formulated negatively, as its 'independence' from sensible impulses. In the *Critique of Pure Reason*, for example, Kant defines it as the «independence of the will from being compelled by impulses of sensibility» (*KrV*, A 534 B 562; also A 802 B 830).

Note the remarkable similarity between this description of freedom of the will and Kant's description of freedom in the «Doctrine of Right», quoted above, as «independence from being compelled by another's choice». In both cases, he defines freedom as «independence» (*Unabhängigkeit*) from «being compelled» (*Nötigung*) by *another* (by another person, or by sensibility, respectively).

Note also that Kant does not contrast freedom of the will to coercion as such. His definition makes freedom of the will compatible with citizens of a republic being coerced by the laws they have given themselves (*MS*, AA VI 231) and with the «self-coercion» of the virtuous agent (*MS*, AA VI 231, 379, 394). Nor does Kant define freedom of the will in terms of a denial of natural determinism or sensible desires, but rather in terms of the will's *independence* from the latter – and saying that *x* is independent of *y* does not imply the denial of *y*. Nor, finally, does Kant contrast freedom to the absence of choice. The relevant contrast is between freedom of the will and the will's being subordinated to sensible desire.

The will, on Kant's definition of it, is the capacity to act on the basis of reasoning. This explains why Kant sometimes *equates* the will and practical reason (e.g., *GMS*, AA IV 412) and at other times speaks of reason *determining* the will, namely when the capacity is exercised and what is willed is indeed determined by reason. In the *Groundwork*, he defines the will as «the capacity to act on the basis of the representation of laws, that is, of principles» (*GMS*, AA IV 412, cf. 427). This definition includes reasoning from moral and prudential principles. To mention one of Kant's examples of the latter: the prospect of surgery feels extremely disagreeable, but on the basis of practical reasoning someone may decide to undergo it nevertheless (*KpV*, AA V 61). In this example, the agent reasons from the principle of promoting his long-term health. He uses practical reason *in the service* of his desire for long-term health.

The idea that the will is *free* should be understood in contrast to the idea that practical reason *merely serves* as a means for the satisfaction of desires. Negatively formulated, it is the idea that the will can be determined *independently* of sensible desires and impulses. Positively formulated, it is the idea that practical reason has principles of its own (viz., moral principles) that can determine the will. Of course, the idea that the will *can* be determined by moral principles does not imply that it *will* be. The human will is affected by sensible incentives, and moral principles hence present themselves as imperatives that state how humans *ought* to act. As Kant puts it in the *Critique of Pure Reason*:

The [will] that can be determined independently of sensible impulses, thus through motives that can be represented only by reason is called the *free will* [*freie Willkür*].

(*KrV*, A 802 B 830)

In the *Groundwork*, Kant uses the notion of 'autonomy' as a property of the will to express the positive understanding of freedom. He asks: «What, then, can freedom of

the will be other than autonomy, that is, the property of the will to be a law to itself?» (AA IV 447), and he claims that «freedom and the will's own legislation are both autonomy and hence reciprocal concepts» (GMS, AA IV 450). Kant also expresses this idea in the *Critique of Practical Reason*, again in terms of the republican notion of freedom as independence:

[the will's] *independence* [from desired objects] is freedom in the *negative* sense; whereas [the will's] *own legislation* is freedom in the *positive* sense, [*i.e.*, the legislation] of pure and, as such, practical reason.

(KpV, AA v 33)

The reason for believing that the will is indeed free in this sense, according to Kant, is our consciousness of moral obligation, as mentioned in 2. 2 above. The belief that morality unconditionally requires certain ways of acting, even if they run counter to all of the agent's sensible desires, gives us grounds for believing (without knowing) that these moral imperatives stem from pure practical reason itself – that is, for believing that the will «is a law to itself» rather than merely serving sensible desire.

In light of Kant's identification of the will with practical reason, his defense of the *freedom* of the will turns out to be the counterpoint to Hume's assertion that «reason is and ought only to be the slave of the passions».¹ Kant asserts that we are justified in assuming that practical reason is *free* rather than the *slave* of inclinations.²

Hence it is not surprising that we do indeed find statements where Kant explicitly denies what Hume asserts. In the section «On Freedom» in the *Lectures on Metaphysics Mrongovius* (1782-1783), for example, Kant reportedly said:

If reason discerns what is really good, or specifies the ends, then it looks after its own interest and is the master/mistress.³ If it merely devises a good means for the sake of the end that arises from inclination, then it merely looks after the interest of inclination and is the slave.

(V-MP/Mron, AA XXIX 899)

And a few pages later:

If in the use of our will we have regard only for our empirical happiness, then ... inclination reigns, determines the end, and reason is the slave who must provide the means. But morality says that ... reason alone must be the master (*Herr*).

(V-MP/Mron, AA XXIX 901)

Kant here describes the *free* will as a will that is independent from sensibility and subject to laws of reason rather than inclination. He does not seem to reverse Hume's image by claiming that reason ought to enslave sensibility. Reason should be the 'master', but this master is described as an enlightened 'autocrat' who rules in accord-

¹ D. HUME, *A Treatise of Human Nature*, ed. by D. F. Norton, M. J. Norton, Oxford, Oxford University Press, 2007 [1739-1740], 2.3.3.4.

² The term 'passion' in translations of Kant's work is reserved for *Leidenschaft*, which means something like an obsession. The counterpart to Hume's 'passions', in Kant's work, are terms like 'sensible impulses', 'desires', or 'inclinations'.

³ Both translations are possible here; Kant uses the feminine '*domina*' (and '*serva*') because the German word '*Vernunft*' is feminine.

ance with moral principles, rather than as a capricious despot (*MS*, AA VI 383; cf. *V-Mo/Mron*, AA XXVII 1510).

In short: the vocabulary we find in Kant's discussion of the 'freedom' of the will is the republican terminology of independence versus dependence, being one's own master versus being enslaved, autonomy versus heteronomy, living under one's own laws versus living under the rule of a despot. A free will is a will that is a law to itself rather than being subject to the «despotism of desires» (*KU*, AA V 432). And given Kant's conception of the will as the capacity to act on the basis of reasoning, a free will is a will that is subject to principles of practical reason itself.

This also explains why the cases in which Kant *exempts* human beings from moral responsibility are not cases of external coercion or lack of alternative possibilities. Rather, they are cases in which human beings are not in full possession of their rational capacities. Kant's own examples are very young children and those who suffer from severe psychiatric illnesses (*Anth*, AA VII 213-214, *V-MP-L1/Pölitz*, AA XXVIII 182, 254). Their will is not free but subordinated to something else: impulses and inclinations, or a delusion resistant to rational correction.

4.2. Further Advantages of the Republican Reading

Providing a complete discussion of Kant's account of freedom of the will lies beyond the scope of this article, but two examples may serve as initial illustrations of how the republican reading can help solve (or dissolve) the notorious interpretive problems mentioned in section 2.3 above.

First, recall Kant's cryptic statement that a free will is a will under moral laws, which gave rise to the so-called Reinhold-Sidgwick objection and the problem of responsibility for bad actions. In light of the analysis above, the statement turns out simply to express the core republican conception of freedom as living under one's own laws rather than under heteronomous dictates. If moral laws are the will's *own* laws, as Kant indeed argues, then a will that is subject to (or 'under') moral laws is indeed a free will. Moral laws are not alien impositions on the will; rather, the will «is a law to itself» (*GMS*, AA IV 440, 447); it is *free* in the republican sense of the term. But this does not imply that a free will is always a *good* will. Being *under* a law means that one *ought* to obey it, not that one *will*.

Thus understood, Kant's statement does not imply that agents bear no responsibility for immoral actions. Consider the analogical case of the political freedom of citizens in a republic. Their political freedom consists in their living under their own laws. They live *under* their own laws, though: they are subject to them, meaning that they ought to obey them. When they do not obey them, their crimes can be imputed to them. Analogously, a free will is a will under moral laws, and an evil will is no less free than a good will. There is no reason to think that evil cannot be imputed, and the Reinhold-Sidgwick problem does not emerge. Thus, the republican background of Kant's notion of freedom illuminates not only the fact that Kant expresses freedom of the will in terms of its 'independence' from sensible desire, but also his otherwise enigmatic identification of a will 'under moral laws' with a 'free' will.

In substance, this response to the Reinhold-Sidgwick objection is in line with what Henry Allison, drawing on a sentence in the *Religion* (*RGV*, AA VI 23-24), has influen-

tially called Kant's «Incorporation Thesis».¹ This is the thesis that «an incentive can determine the will only insofar as it has been incorporated [by the agent] into a maxim».² Put differently, it is the thesis that sensible desires do not by themselves (necessarily, automatically) determine the will. Allison rightly presents this thesis as a feature of Kant's general account of rational agency, but it is important to recognize that the sentence on which he draws is actually Kant's description of the *freedom of the will*, that is, its independence from sensible desire: «Freedom of the will has the characteristic, entirely peculiar to it, that [the will] cannot be determined by an incentive to an action except *insofar as the human being has incorporated [the incentive] into his maxim*» (RGV, AA VI 23-24).

Second, the fact that Kant uses a republican notion of freedom also explains his statement that freedom of the will should not be defined in terms of the *choice* of whether to act morally or immorally (MS, AA VI 226). Again, compare the analogical case of political freedom. The freedom of citizens in a republic does not consist in the choice of whether to obey or disobey the laws. Citizens *can* disobey the laws, of course, and many do. But this is not what their political freedom consists in: it consists in the fact that they live under their own laws. By analogy, freedom of the will does not consist in the capacity to decide whether or not to obey the laws of morality. Human agents have this capacity, of course, and many do disobey. But the freedom of their will consists in its independence from sensibility and its living 'under' laws of reason (MS, AA VI 226-227).

These two examples indicate that a 'republican' reading of Kant's notion of freedom of the will can be hermeneutically fruitful. Both statements, which have long puzzled Kant's readers, now make straightforward sense.

4.3. A Brief Note on the Word 'Will'

I should add a brief explanation of why I use 'will' throughout this essay rather than using different translations for Kant's terms *Wille* and *Willkür*. The reason is that Kant goes back and forth in using these terms and does not clearly distinguish them until the *Metaphysics of Morals*.

In the *Critique of Pure Reason*, Kant uses both *Wille* and *Willkür*; moreover, he sometimes uses *Vernunft* where he later uses *Wille* in the *Groundwork*. In the *Groundwork*, he uses almost exclusively the notion of *Wille* (dozens of times), using *Willkür* only twice (GMS, AA IV 428, 451), and not, it seems, to mark a distinction in meaning. In the *Critique of Practical Reason*, Kant uses both *Wille* and *Willkür* without consistent difference in meaning. For example, he goes back and forth several times between «determining ground *der Willkür*» and «determining ground *des Willens*» at v 22-24. In the *Religion*, he uses *Willkür* to refer to the human will and *Wille* to refer to the will of God. Exceptions here are the expressions 'good will', 'evil will', and 'general will' (in the political sense), for which he uses *Wille*.

In the *Metaphysics of Morals*, Kant explicitly distinguishes between *Wille* and *Willkür*. He writes that *Wille* is the legislative faculty or practical reason itself, insofar as it can serve as the determining ground of *Willkür*; *Willkür* is the elective will or the faculty of choice, including the choice of maxims (MS, AA VI 213, 226).

¹ ALLISON, *Kant's Theory of Freedom*, p. 5.

² *Ibidem*, p. 189.

This distinction between *Wille* and *Willkür* does not match Kant's usage in the earlier texts, as the overview above makes clear. To add one salient example: the *Groundwork's* doctrine of freedom as the 'autonomy' of the will, which he there phrased in terms of the *Wille's* being «a law to itself» (GMS, AA IV 440), cannot be articulated in the terminology of the *Metaphysics of Morals*. Indeed, he there speaks of the *Wille's* being a law to the *Willkür* (MS, AA VI 213), stops speaking of the «autonomy» of the will, and claims that the *Wille* «can be called neither free nor unfree» (MS, AA VI 226).

Because I focus mostly on Kant's publications of the 1780s, given the interest in Kant's reconciliation of freedom and natural determinism, I use 'will' throughout. This avoids unnecessary confusion and preserves the ambiguities of Kant's own terminology during this decade.

5. NATURAL DETERMINISM AND REPUBLICAN FREEDOM OF THE AGENT

Imputing an action *to me* presupposes not only that my will is free in the sense explained above but also that I can act freely. It requires my 'absolute' or 'transcendental' freedom, for if my phenomenal actions were merely the result of natural causal deterministic processes, then they would not be under my control. In that case, I would be no more morally responsible for my actions than the turnspit is morally responsible for turning (see 2.2).¹ This raises the question of *whether and how* my absolute freedom is compatible with the natural causal determinism of the phenomenal world in which my actions appear. As explained above in section 2.3, Kant's answer is generally viewed as inconsistent.

In developing his account of the absolute or transcendental freedom of the agent, Kant again uses the republican notion of freedom as independence. He describes absolute freedom negatively as the power to «produce something independently of those natural causes» (KrV, A 534 B 562). He describes it positively as «absolute self-activity» (KrV, A 418 B 446), as the «power (*Vermögen*) to begin a state (*Zustand*) of oneself (*von selbst*)» (KrV, A 533 B 561), or as a «spontaneity that can start to act of itself, without needing to be preceded by another cause that in turn determines it to action according to the law of causal connection» (KrV, A 533 B 561). These and similar descriptions elsewhere are couched in terms of the familiar republican contrast between independence and dependence, between the power to act by oneself and subjection to the power of another, although here the 'other' is not another human being but natural causality. But it is not immediately clear what this free activity amounts to.

Conceptually speaking, the republican notion of freedom as independence makes absolute freedom compatible with natural determinism: saying that one thing is independent of another does not entail denying the other. It says nothing about *how* they can go together, however, and this is the harder question. It might seem that

¹ Pereboom argues that Kant is mistaken about this because even if one is causally determined to do bad things, and even if it is therefore «false that one *ought* not to do so», one's actions are «still morally wrong» (D. PEREBOOM, *Kant on Transcendental Freedom*, «Philosophy and Phenomenological Research», LXXIII, 3, 2006, pp. 537-567 (562). Pereboom leaves unclear, however, in what (Kantian) sense one's actions can be morally wrong if one bears no moral responsibility whatsoever.

human beings simply *cannot* act independently from phenomenal natural determinism if their future actions can be calculated.

The republican background of Kant's notion of absolute freedom makes it easier, however, to recognize the internal structure of his transcendental idealist version of compatibilism. As long as one conceives of absolute freedom as the power to intervene in phenomenal causal chains or choose between alternative phenomenal options, the idea of freedom does indeed have problematic implications. But Kant does not describe absolute freedom in that way. Instead, he describes it as the noumenal agent's power to *adopt his maxims independently of natural determinism*, and he describes these maxims as *grounding the entire series* of an agent's phenomenal actions. This suggests a solution to the remaining problems mentioned in section 2.3, or so I will argue.

I presuppose the background provided above, especially concerning the guiding thought of transcendental idealism and concerning Kant's claim that morality gives us reason for believing (without knowing) that we have absolute freedom. I first focus on Kant's description of the relation between phenomenal actions, empirical regularities, and noumenal character in the context of moral responsibility attributions. This serves to clarify what Kant *means* by absolute freedom (5.1). I then argue that this analysis sheds new light on his reconciliation of absolute freedom and natural determinism (5.2).

5.1. *Phenomenal Actions, Empirical Character, and Noumenal Ground*

In discussing the Third Antinomy in the *Critique of Pure Reason*, Kant presents a three-fold analysis of human agency. He distinguishes between an agent's (1) *phenomenal actions*, (2) *empirical character*, and (3) *noumenal character*.

Kant writes that a human being's phenomenal actions make it possible to discern his «empirical character». This term refers to the regularity in his behavior. Kant also calls empirical character the «law of [his] causality» and states that it can be «cognized (*erkannt*) on the basis of experience», that is, on the basis of patterns in the person's phenomenal actions (*KrV*, A 540 B 568).¹ Knowledge of the agent's empirical character, in combination with other empirical laws and conditions, makes it possible to explain his past behavior and to predict – in principle, though not necessarily in practice² – his future behavior (*KrV*, A 539 B 567, A 549-550 B 577-578, *KpV*, AA v 99).

The agent's assumed noumenal character cannot be cognized (by definition), nor can it play any role in the natural causal explanation of actions (*KrV*, A 546 B 574). But, Kant suggests, when we impute an action to an agent and evaluate it morally, we conceive of the agent as acting on the basis of certain «subjective principles of his will» (*KrV*, A 549 B 577), and we evaluate these assumed underlying principles. In the *Groundwork* and elsewhere, Kant calls these principles «maxims» (cf. *GMS*, AA IV 400n., 421n.).

¹ Allison claims that Kant understands empirical character as an agent's set of beliefs, desires, and intentions, including «reasons» as «empirical causes» (ALLISON, *Kant's Theory of Freedom*, pp. 5, 31-33, 49), but this claim does not seem to be supported by the texts.

² That complete and accurate prediction is impossible in practice is obvious, and not only because of our complex physiology and psychology. Human behavior is influenced by hard-to-predict factors such as the weather, flying insects, and other humans who are in turn similarly influenced by hard-to-predict factors.

He there clarifies that they are ‘subjective’ in the sense that they are the principles «in accordance with which the subject *acts*», as distinct from the ‘objective’ principles that state how he *ought* to act (421n.). The agent’s maxims constitute his noumenal character, or, as Kant also calls it, his «moral character» (*Anth*, AA VII 285). Character in this moral sense «refers to that property of the will, by which the subject binds himself to determinate practical principles that it has prescribed to itself unalterably through its own reason» (*Anth*, AA VII 292). Moral character can be good or evil, depending on the moral quality of one’s maxims. For example, the ruthless Roman tyrant Sulla had (evil) character in this moral sense (*Anth*, AA VII 293).

Since we do not have cognitive access to other people’s maxims (or to our own, for that matter), we are always fallible in attributing a particular moral character to an agent (*KrV*, A 551n. B 579n.), and we typically need more than one data point. A single instance of honesty may leave open whether the agent is acting on the maxim of honesty, or on the maxim of egoism (say, out of fear of being caught lying). Thus, we generally attribute specific action principles to agents on the basis of *patterns* we observe in their actions, that is, on the basis of their empirical character (*KrV*, A 540 B 568; A 549 B 577; *RGV*, AA VI 20; see also section 2).

Rather than targeting the agent’s phenomenal behavior as such, therefore, moral imputation and evaluation target the agent qua noumenal ground of this behavior who is responsible for his assumed *underlying maxims*. For example, we would not blame a person if we believed that his behavior resulted, say, from involuntary spasms. When we blame someone, we do so on the assumption that the action was intentional and performed on a morally impermissible maxim.

Moreover, Kant argues that when we impute phenomenal actions to an agent, we assume that *he* is responsible for adopting the maxims that underlie those actions. We assume that in adopting his action principles, *he* gave himself the particular moral character that underlies the corresponding patterns in his empirical behavior. This assumption presupposes that he was *free* in adopting these maxims. We assume, in Kant’s terminology quoted above, that he adopted them by himself, of his own accord, independently of natural determinism, *qua* noumenon.¹

Kant articulates a similar account in the *Critique of Practical Reason*. When we blame others for their phenomenal actions, he writes, we presuppose that they have freely adopted the reprehensible action principles that we attribute to them on the basis of empirical patterns of conduct. This is how he describes what is presupposed when we hold someone responsible (*verantwortlich*, *KpV*, AA v 100) for a string of bad actions:

This [viz., the blame] could not happen if we did not presuppose that everything that arises from his will (as every action intentionally performed undoubtedly does) has as its ground a free causality, which ... expresses its character in its appearances (the actions); these actions, on account of the uniformity of conduct, make knowable a natural interconnection that does not, however, make the vicious constitution of the will necessary but is instead the consequence of the evil and unchangeable principles voluntarily adopted, which make him only more reprehensible and deserving of punishment.

(*KpV*, AA v 100)

¹ For a discussion of Kant’s conception of the ‘action’ of substances, including the action of noumenal agents, as not necessarily being an event in time, see WATKINS, *Kant and the Metaphysics of Causality*, ch. 4.

In this passage, Kant again distinguishes between (1) appearing actions, (2) the characteristic empirical patterns they display («uniformity of conduct»), and (3) the «free causality» that is thought to underlie these patterns as their (noumenal) «ground»: the agent's will, characterized in terms of his voluntarily adopted evil maxims. The agent is considered to be responsible for his bad actions by virtue of his having freely adopted the maxims that underlie them. Thus, the agent is held responsible not for his individual phenomenal actions *qua* phenomena, but for having voluntarily adopted the action principles that *underlie* them as their ground.

In the quoted passage, Kant speaks of the agent's 'free causality' even though his maxims are evil. This underscores that he does not conceive of absolute freedom as an asymmetrical capacity for adopting morally good maxims only. Moral imputation presupposes the agent's absolute freedom even when he is assumed to have made bad use of it.

It is also important to note what Kant does not say. Nowhere does he describe noumenal freedom as the freedom to alter existing natural causal chains. And nowhere does he posit the existence of alternative phenomenal possibilities at a particular point in time. Rather, he describes the presupposed «free causality» of the agent by reference to his voluntary adoption of the *maxims* that constitute his moral character. He conceives of these maxims as *grounding the entire phenomenal chain* of the agent's actions, indeed «*everything* that arises from his will». This is crucial for understanding Kant's reconciliation of absolute freedom and natural determinism, to which I now turn.

5.2. Reconciling Natural Determinism and Absolute Freedom

We are now in a position to see the structure of Kant's reconciliation of the absolute freedom of the noumenal agent with the natural determinism of his phenomenal actions. The republican notion of freedom as independence enables Kant to locate the absolute freedom of the agent entirely on the noumenal side, in a way that a 'contra-causal' understanding of freedom cannot. Recall that according to transcendental idealism, all phenomena, including appearing human actions, are subject to natural determinism, due to the mode of operation of our cognitive powers. Recall further that all phenomena – whether ships sailing down the river, beavers building dams, or humans telling lies – are thought to have 'things in themselves' as their unknowable ground. Only in the case of some phenomena, namely human actions, do we have any reason to make further assumptions about the unknowable «things in themselves» that ground them. This reason is our consciousness of moral obligation and our practice of moral imputation and evaluation (as explained above, 2. 2 and 5. 1), since they prompt us to conceive of ourselves and others as having the power to adopt maxims *independently* of natural determinism. Phenomenal actions are part of the realm of natural determinism, and as such they are never free.¹ But insofar as these actions can be regarded as stemming from the noumenal agent's adopted action principles, transcendental idealism makes it possible to impute his phenomenal actions *to him*.

¹ Appearing actions can be called 'free' at most in an indirect sense, namely insofar as they are regarded, in moral contexts, as resulting from the agent's use of his freedom (*KrV*, A 551-557 B 579-585; *V-Mo/Mron*, AA xxvii 1437; *RGV*, AA vi 170n.).

This sheds new light on Kant's suggestion in the first *Critique* that the liar could have and ought to have acted differently. Had the liar adopted the maxim of honesty, he would have been an honest man. He would have displayed a pattern of honest actions, also in the case at hand. He *ought* to have adopted that maxim. But this means just that: he ought to have *adopted that maxim*. It does not mean that he could have or ought to have meddled with existing natural causal chains. His free activity consists entirely in his adopting his maxims. Had he adopted the maxim of honesty, his empirical conduct as a whole would *thereby* have been different, given that his maxims underlie his phenomenal actions as their ground. Hence «another intelligible character would have given another empirical one» (*KrV*, A 556 B 584).

This analysis also suggests that the relevant question is *not* how the liar could ever have been honest, considering that the deterministic universe was all lined up to cause his lie. Or, more generally, the question is *not* how noumenal agents can ever choose their character if natural determinism has already charted their life course for them. On Kant's conception as reconstructed here, these questions are ill conceived. Noumena are the («thought») things as they are in themselves, that is, in abstraction from the specific mode in which we cognize them as appearances. The phenomenal series is thought to *depend* on this noumenal ground. Thus, if the underlying thing in itself had constituted itself entirely differently, then, had we cognized this different thing, it would have appeared to us differently. Given the grounding relation between noumena and phenomena, if the liar had been an honest man (if he had made the maxim of honesty part of his moral character), then the patterns in his phenomenal conduct would have been different, and the universe would not have been lined up to produce the lie.¹ Had he adopted a different maxim, «the whole sequence of his existence as a sensible being» would have been different (*KpV*, AA v 98).² As a result, the difficulties mentioned in 2.3 – concerning the possibility of avoiding future moral failure and the possibility of producing the proper causal histories of alternative phenomenal options – do not emerge.

Why did the liar adopt his bad maxim? How does Kant account for moral failure? It might seem that this question poses a new difficulty for Kant. His answer is clear, however: evil *cannot* be explained. Natural causal explanations can be given only within the realm of possible experience and in terms of natural causal laws (*GMS*, AA IV 459). The agent's adoption of his noumenal character lies outside this realm and hence beyond the reach of explanation.³

¹ For a reply, along similar lines, to the related worry that an agent's giving himself one noumenal character rather than another implies his ability to determine the phenomenal past at will, see T. ROSEFELDT, *Kants Kompatibilismus*, in *Sind wir Bürger zweier Welten? Freiheit und moralische Verantwortung im transzendentalen Idealismus*, ed. by M. Brandhorst, A. Hahmann, B. Ludwig, Hamburg, Meiner, 2012, pp. 77-109.

² The following passage offers another instructive statement of Kant's view: «In this regard, a rational being can rightly say of every unlawful action he performs that he could have omitted it, even though, as appearance, it is sufficiently determined in the past and, in that respect, is unavoidably necessary; for *this action*, with all the past that determines it, *belongs to a single phenomenon of his character, which he gives to himself*, and in accordance with which he imputes to himself, as a cause independent of all sensibility, the causality of those appearances» (*KpV*, AA v 98, emphasis added).

³ Kant addresses many other follow-up questions that lie beyond the scope of the present essay, for example how to conceive of the possibility of moral self-improvement (*RGV*, AA VI 44-53).

Kant views this not as a weakness but as a great strength of his theory. If we could *explain* an agent's noumenal character, we would be tracing it back to *other* causes, thereby rendering imputation to the agent *himself* impossible. In the *Religion*, Kant does not tire of repeating that the impossibility of explaining a person's noumenal character is a crucial feature of his account, since it preserves the possibility of moral imputation (*RGV*, AA VI 21, 25, 38, 260n.) – and this, after all, is what motivated his defense of believing in freedom in the first place.

CONCLUDING REMARKS

The fact that Kant uses the republican notion of freedom as independence sheds new light on his theory of freedom and suggests novel responses to age-old objections. It reveals what Kant means by «freedom of the will» and «absolute freedom» of the agent, and it illuminates the relation between the (phenomenal) 'me' and the (noumenal) 'I'. The term 'transcendental compatibilism' would be a fitting name for his theory. It would indicate both its rootedness in Kant's transcendental idealism and its distance from the standard versions of compatibilism.

I have focused on the *meaning* of 'freedom' in the context of Kant's account of freedom of the will and freedom of the agent. I have not provided a full discussion and analysis of his account, let alone a defense of it. It may well turn out that his theory faces a new set of difficulties when read along the lines I have suggested. Whether it does is a question well worth exploring.¹

¹ This essay is an expanded version of the 2019 Frankfurter Kant-Vorlesung, sponsored by the Institute of Philosophy of the Goethe University Frankfurt and the Società Italiana di Studi Kantiani. I thank Gabriele Gava, Stefano Bacin, and Marcus Willaschek for the honor of the invitation and the organization of the lecture. I am also grateful to them, as well as to Daphne Brandenburg, Michael Gregory, Patricia Kitcher, Sven Nyholm, Tobias Rosefeldt, and Janis Schaab for valuable comments. I thank the audience of the Frankfurter Kant-Vorlesung, as well as audiences and organizers at the Humboldt University in Berlin, the Dutch Research School of Philosophy, the 13th International Kant Congress in Oslo, University College Dublin, and the University of Munich, for helpful questions and discussion.