

teorema

Vol. XLII/2, 2023, pp. 49-64

ISSN 0210-1602

[BIBLID 0210-1602 (2023) 42:2; pp. 49-64

Can Determinists Act Under the Idea of Freedom?

Martin F. Fricke

RESUMEN

El determinismo que niega la libertad de acción es una posición filosófica común. ¿Es la acción de tales deterministas incompatible con la afirmación de Kant de que un ser que posee una voluntad racional “no puede obrar de otra suerte que *bajo la idea de la libertad*” [G 4, 448]? En mi artículo, examino el argumento que Kant da a favor de esta afirmación al inicio del Capítulo 3 de la *Fundamentación* y arguyo que equivale a la aserción de que no es posible actuar siendo consciente de estar guiado por principios inválidos. Creer en el determinismo no es lo mismo que tener tal conciencia. Examino un argumento estructuralmente similar al de Kant basado en el contraste entre deliberación y predicción y muestro que también es deficiente.

KEYWORDS: *autoconcepción, argumento transcendental dirigido a la creencia, problema de la concepción errónea, deliberación, Kant.*

ABSTRACT

Determinism which denies freedom of action is a common philosophical view. Is the action of such determinists incompatible with Kant’s claim that a rationally willed being “cannot act otherwise than *under the idea of freedom*” [G 4, 448]? In my paper, I examine Kant’s argument for this claim at the beginning of the Third Section of the *Groundwork* and argue that it amounts to the assertion that one cannot act while being aware of being guided by invalid principles. Belief in determinism does not necessarily amount to such awareness. A structurally similar argument based on the contrast between deliberation and prediction is examined and similarly found wanting.

KEYWORDS: *Self-Conception, Belief-Directed Transcendental Argument, Problem of Misconception, Deliberation, Kant.*

I. INTRODUCTION

In the Third Section of the *Groundwork*, Kant famously claims that a rationally willed being “cannot act otherwise than *under the idea of freedom*” [Kant (2012), p. 125 [G 4, 448]]¹. On the face of it, this claim is incom-

patible with the existence of certain determinists. A philosopher such as Kant's contemporary Johann Ulrich (1788), for example, claims that our actions are predetermined by a sum of factors that obtain prior to our decision to act and that are not under our control. He denies that there is any time when many different courses of action are possible and the action eventually to be realised depends on nothing but our free choice, without being able to be predicted, even in principle, from pre-existing determining factors. Not all determinists deny that we are free. Compatibilists define freedom in such a way – for example, as the ability to do what one wants to do – that it is compatible with determinism. On the other hand, Galen Strawson argues that, independently of whether determinism is true or false, it is provable that we are not free in the sense of “truly responsible” for our actions [cf. Strawson (2010), pp. 24f.]. My focus in this paper is on agents who believe that they are *not* free. I take it that a certain class of determinists such as the aforementioned Ulrich constitute paradigm examples of such agents. The important point about these freedom-denying determinists is that they *exist*, that they *act* and that they are, at least in some general sense, *rational*. As such, they seem to constitute counterexamples to Kant's claim that a rationally willed being cannot act otherwise than under the idea of freedom. Since they deny being free, they seem to act under an idea of unfreedom.

It seems that, faced with the fact that there are people who are freedom-denying determinists and yet capable of acting rationally, Kant could make one of the following three replies:

- (1) There are no real freedom-denying determinists. The people who *claim* to be such do not really believe that they are not free.
- (2) There are real freedom-denying determinists. But although they *believe* that they are not free, they do not *act* under the idea of unfreedom. When they act, they act under the idea of freedom. This is compatible with belief in determinism and denial of freedom.
- (3) There are real freedom-denying determinists. Because of their views they are incapable of rational action. For them to act, requires them to be irrational; for they have to act under the idea of unfreedom, which is impossible for a rationally willed being.

In what follows, I will try to show that Kant's reply, at least according to the argument of the few lines of the *Groundwork* to be examined here, would be (2). Acting under the idea of freedom does not, on this view,

entail believing oneself to be free. My purpose in arguing for this claim is not primarily to elucidate Kant or to add significantly to the enormous literature on his theory of freedom, but rather to shed some light on the relation between our capacity to act and our self-conception as agents. I shall argue that this relation is rather loose. As the diversity of actually existing self-conceptions shows, we can act while adhering to all kinds of philosophies of action, even freedom-denying ones. I am *not* interested, in this paper, in the question of whether or not determinism is true or whether or not we are in fact free. By extension – and I mention this to stress that my aim is not to contribute to Kant exegesis – I am also *not* interested in examining Kant’s arguments for the *reality* of freedom.²

The structure of the paper is as follows. I begin with a general characterisation of the issue at hand as an instance of the “problem of misconception”, which, according to Quassim Cassam (1997), generally arises for belief-directed transcendental arguments. To argue the case that Kant would choose number (2) from the above options, I shall then examine the passage immediately following his claim that we “cannot act otherwise than *under the idea of freedom*” [Kant (2012), p. 125 (G 4, 448)]. I will distinguish three elements in his notion of acting under the idea of freedom: guidance by principles, authorship of principles and awareness of principles. A difficulty with his argument is the question of how to extend the approach from theoretical judgment to action. I conclude with a brief discussion of an argument analogous to Kant’s to the effect that, without the presupposition of freedom, we can only try to *predict* what we are going to do, but can no longer deliberate about this question. I shall argue that the argument is unsuccessful if taken to show that in order to deliberate we must not have a self-conception which denies our freedom.

II. AN INSTANCE OF THE PROBLEM OF MISCONCEPTION

Before going into the details of Kant’s arguments for the necessity of acting under the idea of freedom, it is useful to attempt a general characterisation of the problem raised by my title question. It seems to me that it is an instance of what, in relation to belief-directed transcendental arguments, Quassim Cassam has called a “problem of misconception” [cf. Cassam (1997), p. 128]. Belief-directed transcendental arguments have the following form:

- (1) P is uncontroversial.
- (2) Belief in S is a condition for P .
- (3) Therefore, we believe that S .

Transcendental arguments of this type are supposed to be less demanding than truth-directed transcendental arguments [Peacocke (1989), p. 4], which have the same form but aim to derive the *truth* of S from the fact that it is a condition for the uncontroversial proposition P . Against the truth-directed version of the argument it has precisely been objected that it cannot show that S is actually true, but only that we must believe that it is.

However, Cassam points out that a belief-directed transcendental argument faces the following challenge: someone who is sceptical about S might simply deny that she believes that S , no matter what the argument purports to show. In response to the belief-directed transcendental argument, she would be more inclined to doubt the premises of the argument, in particular premise (2), rather than accept that she in fact believes that S . This is the problem of misconception. The sceptic seems to believe something that, by the lights of the transcendental argument, she cannot believe. She seems to have a misconception. To maintain the argument as valid and sound it would have to be shown that the sceptic believes a proposition that she claims not to believe.

Now, at least on a certain reading, Kant's claim that a rationally willed being "cannot act otherwise than *under the idea of freedom*" [Kant (2012), p. 125 (G 4, 448)], likewise, makes a belief-directed transcendental argument of the following form:

- (1) We have a rational will and we act.
- (2) It is a condition for a rationally willed being's acting that it believes itself to be free.
- (3) Therefore, we believe ourselves to be free.

In this reconstruction, Kant's claim argues that we must believe something because it is a condition of us being rational and capable of action. Now, if there are persons who fulfil this condition but claim *not* to have the belief in question, then the argument faces a problem of misconception, as described by Cassam. Persons who deny that they are free, such as a considerable subgroup of determinists, seem to fit this bill. My title question means to draw attention to this problem of misconception.

In what follows, I shall attempt to unpack Kant's claim and show that he does not argue that it is a condition of a rationally willed being's acting that it believes itself to be free. His claim is weaker: it is a condition of a rationally willed being's acting that it is *not aware* of being guided in any particular action it is realising by an invalid principle. It might be possible to construct a problem of misconception for this version of the argument as well. But it would have to involve an awareness of being guided in a particular action by an invalid principle. I shall argue that being a determinist, even a freedom-denying determinist, who acts, does not *per se* involve having such awareness.

III. KANT'S ARGUMENT IN THE *GROUNDWORK*

To understand Kant's argument in detail, let me now examine the passage immediately following his claim that we "cannot act otherwise than *under the idea of freedom*" [Kant (2012), p. 125 (G 4, 448)]. Kant justifies his claim as follows:

Now, one cannot possibly think of a reason that would self-consciously receive guidance from any other quarter with regard to its judgements, since the subject would not then attribute the determination of judgement to his reason, but to an impulse. Reason must view herself as the authorship of her principles, independently of alien influences, and must consequently, as practical reason, or as the will of a rational being, by herself be viewed as free; [...]. [Kant (2012), p. 125 (G 4, 448)]

This is an argument to the effect that we, as reason-endowed beings, must view ourselves as free, where freedom means independence from alien influences and being author of one's principles. The reason that Kant gives for this claim is that if we self-consciously received guidance from any other quarter with regard to our judgments, we could not attribute the determination of our judgments to our reason, but would have to attribute it to an impulse. How are we to understand this argument? Let me distinguish three elements in the argument: guidance by principles, authorship of principles and awareness of principles. This will allow us to understand more precisely what it means, according to the lines quoted above, to act under the idea of freedom and why this is necessary for a rationally willed being.

III.1 *Guidance by Principles*

Certainly, in order to judge about some state of affairs in the world we have to receive "guidance" from the world; what hope could we

have, otherwise, that our judgments approach the truth? As contemporary philosophers express it, the direction of fit here is from the world to our judgments about it. This kind of guidance, then, – being guided in our judgments by the world – does not seem to be the one that, Kant says, must necessarily come from our *reason*. Rather, when he says “guidance”, he seems to refer to *principles* which determine our judgments. Even though we are guided by the world when trying to judge about it, in the sense of trying to find out what things are really like, it might be said that we are following principles authored by our reason; for example, the principle to suspend a perception-based judgment when we cannot, by perception, distinguish clearly what is the case.

III.2 *Authorship of Principles*

If by “guidance” in judgments Kant means “guidance by principles”, then his claim in the above quote is that we cannot judge guided by principles that our reason is not the author of. What does this mean? What would a principle be that our reason is not the author of? If Kant is right, then it should be hard to find one we actually use, because we are reason-endowed and cannot, he suggests, judge according to such principles. It would have to be a principle that is obviously irrational. The following might be an example: “Judge that the moon is made of cheese, no matter what evidence you have for the material that it is made of.”

I am the author of this principle; I have literally invented it. However, this kind of authorship – simply being the factual source – does not seem to be the one Kant has in mind. Rather, he seems to have in mind principles that we recognise as valid in some sense; the notion of authorship seems to refer to the fact that we have this insight. Even though I have authored the above principle about the moon in the sense of having invented it, I do not recognise it as valid for reasoning and, in *that* sense, reason is not “authoress”³ of this principle.⁴

III.3 *Awareness of Principles*

Kant does not actually say that we cannot judge guided by principles that our reason is not the author of. Rather, he says that we cannot *self-consciously* do so. What does it mean *self-consciously* to judge according to some principle? It seems plausible to take this phrase as indicating awareness on the part of the subject as to how she judges. On this interpretation, Kant claims that I cannot judge such that I am aware of judging

according to a specific principle *and* I am aware that the principle in question is not valid.

Why is this impossible, according to Kant? Because in this case I would attribute the determination of my judgment to an impulse and not to reason. In contemporary terms, the idea seems to be that I cannot judge while being aware of a certain type of *defeating condition* for my judgment. If I find myself (i.e. am aware of) judging according to some principle that I am aware of as invalid, then I should abandon the judgment. This is because it is unlikely that the judgment will turn out to be true, and even if it does, it will do so only by accident. If I am rational, I will not sustain such a judgment, given my awareness of the invalidity of the principle followed.

The absence of such circumstances is what Kant seems to have in mind when he uses the phrase “under the idea of freedom”. I am judging under the idea of freedom in so far as I am not aware of judging according to a principle that I am aware of as invalid. This is necessary for a rational person, he seems to argue, because awareness of judging according to an invalid principle would mean awareness of a defeating condition for my judgment and should therefore lead me to abandon the judgment.⁵

In the above quote, Kant also asserts that reason must “by herself be *viewed* as free” [Kant (2012), p. 125 (G 4, 448) my emphasis]. This might suggest that reason must *believe* or *assume* herself to be free or must have a positive conception of what freedom is and ascribe it to herself. Note, however, that the argument I have developed over the preceding lines does not support the claim that such a self-conception is necessary for a rational person to be able to make judgments. The argument I have developed only permits a negative conclusion: we must *not be aware* of certain defeating conditions in our judgments. No claim about a belief, an assumption or a self-conception to the effect that we are positively capable of freely making judgments in some sense has been argued for.

III.4 *The Problematic Transition from Judging to Acting*

The problem, at this point, is that our reconstruction of Kant’s argument so far is only about freedom in *judging*. But Kant’s famous claim, which is meant to be supported by it, is about *acting* under the idea of freedom. How do we get from freedom in judging to freedom in acting? This is a question that has vexed Kant interpreters for a long time⁶ and I will not pretend to answer it here. We might think that judging is a form of acting and, thus, if judging must be done under the idea of freedom, Kant has shown that at least some actions must be done under the idea of freedom. There is some plausibility to the idea that all judgments are actions be-

cause, we might say, they are some kind of mental operations performed in the mind. However, it is less clear that all actions are judgments. Alternatively, it might be thought that ordinary physical actions are the (distinct) *effects* of practical judgments, instead of being identical with them.

Does Kant identify acting (or perhaps deciding to act) with judging? I shall not investigate this question here. Suffice it to say that if he does, the above argument immediately applies to actions as well: If I am aware of acting (or deciding to act) according to a principle that I am aware of as invalid, then I will attribute the determination of my action (or decision) to an impulse. In consequence, the argument says, I should abandon my action (or decision). But is this true?

The answer is unclear, in part because we have not yet explained what the validity of a principle consists in. It is plausible to say that a principle followed in theoretical judgments should (at least) not make it probable that the judgments turn out to be false. Principles that do, could be described as invalid for judging. Theoretical judgments aim at truth. So, becoming aware of judging according to a principle that makes it likely the judgment turns out false, plausibly has the consequence that the subject abandons the judgment. Awareness of the fact that the principle followed makes the falsity of the judgment probable is awareness of a defeating condition for the judgment.

But what does it mean to say that a principle of *action* (or decision) is invalid? One possible answer is that invalid principles of action are self-defeating in the sense that following them leads one to act in such a way that one necessarily fails to achieve what one intends to achieve. Following the principle “To satisfy your hunger, don’t eat”, will lead to many things, but not, normally, to the satisfaction of one’s hunger. It is plausible to say that if a subject becomes aware of acting according to this principle and she is aware of the self-defeating character of the principle, then she will abandon the action. In this case, then, it is plausible that Kant’s argument is correct: Becoming aware of acting according to an invalid, i.e. self-defeating, principle should normally lead one to abandon the action. A rational person cannot act except not being aware of acting according to such a principle. This, according to this reconstruction of the argument, is what acting under the idea of freedom amounts to.

This being a discussion about a thesis from the *Groundwork of the Metaphysics of Morals*, it has to be mentioned that a principle of action might also be considered invalid because a *universal* following of the principle would lead to an impossibility of acting on it. In other words, if everyone followed the principle (as though it was a universal law), then

individual actions following the principle could no longer achieve what they were intended to achieve. As the prime example for this kind of invalidity says, if everyone acted on the principle “Break your promises if this is convenient for achieving some other end”, then it would no longer be possible to achieve any end by falsely promising, because no-one would believe your promises anyway.⁷ If we understand the invalidity of principles of action in this way, Kant’s argument about the necessity of acting under the idea of freedom might seem less plausible. Suppose I become aware of acting according to a principle that cannot be universalised. Does this awareness lead me to abandon my action? It seems clear that I might abandon my action, or I might not. Whether or not I do, depends on my (Kantian) morality. Clearly, there are people who do evil actions whose evilness lies precisely in the fact that they perform them although being aware that they are not universalisable. Is there a *rational* requirement on me to abandon such an action? This might be disputed, but Kant certainly thinks there is. So another way of reconstructing Kant’s argument about freedom is to say that if a (rational) person becomes aware of acting according to a non-universalisable principle, then she will abandon her action. Such principles constitute defeating conditions for this person’s actions. In other words, a rationally willed person cannot act except not being aware of acting on a principle that is not universalisable and this is what it means to act under the idea of freedom.⁸

III.5 *Determinists Acting Under the Idea of Freedom*

No matter how we understand the invalidity of principles of action in our reconstruction of Kant’s argument, it seems to me that now we are in a position to explain how he should reply to the title question of this article. He should say that determinists who deny freedom can still act under the idea of freedom and this does not make them irrational (reply (2) in the Introduction).

Freedom-denying determinists believe that they are unfree. But this general belief does not necessarily amount to awareness of an invalid principle guiding them in their actions. Determinism is a general theory about the existence of certain factors that supposedly determine our actions and judgments. But although it is a theory about our actions, it is not, *qua* determinism, a theory containing principles that could guide actions or decisions. A determinist might, coherently or not, infer such principles from her theory; for example, about whether and how criminals should be punished for their crimes. But even a detailed theory which explains how actions and decisions are determined does not as

such contain principles of action, and belief in the theory does not amount to awareness of such principles. This means that someone who believes in determinism and denies freedom, might nevertheless act without being aware of being guided by invalid principles. In other words, despite her beliefs, she might still act under the idea of freedom.⁹

Kant does not need to show that freedom-denying determinists do not really believe what they say they believe (reply (1) in the Introduction). The determinists' beliefs about how our actions (or judgments) are determined, are irrelevant to the claim that when they act, they cannot – as rational agents – be aware of acting or judging according to an invalid principle. It is also unnecessary for Kant to denounce freedom-denying determinists as irrational because they act under the idea of unfreedom (reply (3) in the Introduction). They do not; for believing in determinism does not necessarily entail being aware of acting (or judging) according to an invalid principle. And not being so aware is sufficient for counting as acting (and judging) under the idea of freedom.

This must be a welcome result for anyone who wishes to defend Kant's theory of freedom (or, at any rate, the few lines from the *Groundwork* I have examined here), since it would be rather embarrassing if the theory could not accommodate acting determinists. People have held a great variety of self-conceptions throughout history and continue to do so today. But although they are thus divided by their philosophical views, they seem to have in common a capacity for rational action. Whatever the exact details of Kant's views, it is surely desirable that they be compatible with this fact.

IV. DELIBERATION *VERSUS* PREDICTION

It is possible to construct an argument structurally similar to Kant's based on the distinction between deliberation and prediction. Deliberation about what to do, it might be said, only makes sense if what I am going to do has not yet been determined. It requires that the choice is up to me, not decided prior to my deliberation. That this is so, it might be added, comes out if we assume that my future action is predetermined by factors prior to my deliberation. In this case, it might be argued, it does not make sense to deliberate about what to do. There is nothing to be deliberated about; at best, I could try to *predict* what I am going to do. This is because I cannot determine by deliberating and coming to a decision what I am going to do. I can only try to find out what, in fact, I am going to do anyway. In summary, deliberation presupposes that the de-

liberator takes herself to be free, while belief in determinism condemns her to mere attempts at predicting her own actions.

Again, we seem to be dealing with a belief-directed transcendental argument:

- (1) We deliberate about what to do.
- (2) It is a condition for deliberating that one believes oneself to be free.
- (3) Therefore, we believe ourselves to be free.

It might be objected that even if we accept that belief in predetermination requires a predictive stance, this does not mean that deliberation requires belief in freedom. It only means that the deliberator cannot believe herself to be predetermined. So the conclusion of the argument would not be that a certain belief (in freedom) is necessary for deliberation, but only that the *absence* of a certain belief (namely in determination) is necessary. Either way, the argument is clearly vulnerable to the problem of misconception. Freedom-denying determinists seem to be capable of deliberation, although they claim to believe that their actions are predetermined and that they are not free. By the lights of the above argument, they have a misconception of themselves.

Unless we wish to deny that freedom-denying determinists really believe what they say they believe, the problem of misconception seems to indicate that there is something wrong with the argument. I think the problem lies with the second premise. It is not true that one has to believe oneself to be free in order to deliberate. It is not even necessary *not to believe* oneself to be determined. The reason is that these general self-conceptions are too far removed from the activity of deliberation to be able to influence it, in the sense of “defeating” or “disabling” it. Deliberation is an activity that is common to all humans who are capable of considering reasons for or against their own actions. It is so basic to our existence that it is unclear how any general self-conception could ever impede it. In Kant’s original argument, a theoretical judgment is defeated by awareness of the invalidity of a principle followed in making the judgment. General belief in determinism is no such defeater for deliberation. That I was predetermined to consider these specific reasons for or against some action and to come to see what course of action, on balance, they favoured does not defeat my considerations or conclusions. For deliberation and its results to be defeated the belief should imply that I cannot consider and evaluate reasons *correctly*. But it is unclear why

belief in determinism should have such an effect. The belief does not seem to play any role in the deliberation itself, nor does it invalidate the deliberation. It is just a (meta) theory about the nature of this activity.

What about the claim that deliberation does not make sense if one believes in determinism, because this means that the choice of action is not up to oneself? If deliberation can proceed despite one's belief in determinism, then it seems moot to say that there is a problem with the choice of action not being up to one. The choice is up to one's deliberation. In so far as one's deliberation produces an outcome, for example a decision to act in a specific way, there is a sense to it.

Does belief in determinism condemn the agent to take a predictive stance towards her own actions? It does not, if it is possible to deliberate despite believing in determinism. In any case, when trying to predict someone's action – be it our own or someone else's – it generally seems to be a good strategy to take the person's own deliberation and decision into account. "What have you/I decided to do?" seems a good question to start with in this case, independently of whether or not the person in question adheres to some philosophy of determinism. Prediction here depends to some extent on deliberation, rather than being able to replace it.

In general, it seems plausible to say that deliberation can be practiced under a variety of different conceptions an agent might have of herself. She might, in fact, not even have any opinion on the matter of how deliberation proceeds and what it presupposes. Or she might understand deliberation not as determining an as yet open course of action, but perhaps as the development of a predetermined battle of ideas and motives taking place in her mind. Others might want to characterise the experience of coming to a decision in terms of a more active subject, pondering, discarding, or adopting ideas.¹⁰ However, it is not at all clear that we *have to* accept a determinate philosophical theory of deliberation – let alone the *true* philosophy of deliberation – in order to be able to deliberate. There must surely be a true self-conception of the deliberator. But it is not necessary for us to know it in order to deliberate. We can deliberate as much under a self-conception of freedom as under the conception of different kinds of freedom-denying determinism.¹¹ This should come as no surprise. There are many freedom-denying determinists in the world and perhaps even more Kantian believers in freedom. It would seem rash to declare that members of one or the other of these groups cannot really deliberate or do not really believe what they say they believe.

V. CONCLUSION

Kant claims that a rational person can only act under the idea of freedom. But what does it mean to act under the idea of freedom? Does it preclude having certain determinist self-conceptions which deny freedom of the will? In this paper, I have argued that this is not so. The way Kant argues for his claim, at least in the passage immediately following it in the *Groundwork*, makes it clear that he understands acting under the idea of freedom as acting without being aware of being guided by an invalid principle. We cannot act otherwise because – this is his argument – an invalid principle in some sense defeats the purpose of the action. A rational person who becomes aware of acting on such a principle will abandon the action. As I have pointed out, the plausibility of this argument depends to some extent on the notion of an invalid principle. In any case, I have argued that if we accept Kant’s argument, acting under the idea of freedom turns out to be compatible with freedom-denying self-conceptions because beliefs about whether or not we are free do not necessarily amount to awareness of invalid action-guiding principles.

I have also examined a structurally similar argument, which says that we can only deliberate about what to do under the presupposition that we are free, because without this presupposition the correct stance to take to our action is that of prediction rather than deliberation. I have argued that this argument fails if taken to show that the deliberator cannot believe herself to lack freedom. Our capacity to deliberate about our actions is too basic to be dependent on a specific self-conception. Determinist conceptions do not constitute defeating conditions for deliberation because deliberation itself can be conceived of in a determinist framework. There exists a great variety of self-conceptions among humans, some of which deny freedom. The capacity for rational deliberation and action does not seem to vary correspondingly. Even determinists can deliberate and act rationally.

*Departamento de Humanidades y Sistemas Sociales
Escuela Nacional de Estudios Superiores, Unidad Mérida
Universidad Nacional Autónoma de México
Carretera Mérida-Tetiz Km 4.5
Municipio de Ucú, C.P. 97357, Yucatán, México
E-mail: mfcphcis@gmail.com*

ACKNOWLEDGMENTS

An earlier version of this text was discussed at the seminar “Kant y la filosofía alemana”, organised by Efraín Lazos at the National Autonomous University of Mexico (UNAM). I am very grateful to Efraín Lazos and the participants of the seminar for their thorough discussion and useful comments. I have also presented the main ideas of the paper at congresses in Kent, Mexico and Porto and benefitted from the comments of the audiences. Finally, I would like to thank two anonymous referees for **teorema** for their helpful suggestions.

NOTES

¹ Here, as elsewhere in the text, I am adding the page number according to the Academy Edition (volume 4) of the *Groundwork*, as given in Kant (2012).

² Another topic that I shall not discuss, although it is central to Kant’s theory, is his claim that we are free in so far as we are things in themselves but determined according to causal laws in so far as we are empirical appearances. For a recent discussion see Pereboom (2006).

³ Gregor and Timmermann’s translation characterises reason as “author-ess”, rather than just “author”, because “Kant is clearly using personification as a stylistic device in this passage, which is why the allegorical style of the original has been reproduced in this translation” [Kant (2012), p. 172].

⁴ There is no explicit mention, in the passage I am trying to interpret, of the “validity of principles”. However, the terminology does not seem alien to some central ideas in the *Groundwork*, such as the claim that our actions are governed by maxims (or “subjective *principle[s]* for action”), of which we can will – if they conform to the categorical imperative – that they become universal laws (or “objective *principle[s]*, *valid* for every rational being”) [cf. Kant (2012), pp. 69ff, footnote [G 4, 420f], my italics]. Reference to our capacity for discerning the validity of principles seems to me the best way of elucidating Kant’s figurative talk of reason as the “author-ess” of principles.

⁵ Awareness is a factive state: if you are aware that *p*, it must be true that *p*; if you are aware of an object *O*, *O* must exist. However, it might also be possible to run the argument of the preceding paragraphs with reference to non-factive states. If I *mistakenly* take myself to be judging according to an invalid principle, this might also lead me to abandon my judgment. If this is true, judging under the idea of freedom might not only require *not being aware* of judging according to an invalid principle, but also *not mistakenly taking oneself* to be judging according to such a principle. In what follows, I shall ignore this further complication.

⁶ Cf. for example Henrich (1998). For a recent contribution to the discussion and overview see Hiller (2016).

⁷ In the *Groundwork*, Kant discusses a more specific principle of this type: “when I believe myself to be in need of money I shall borrow money, and promise to repay it, even though I know that it will never happen” [Kant (2012),

p. 73 (G 4, 422)]. For more details on how principles of action can, individually or in universalised form, be self-defeating, compare O'Neill (1989), chapter 5: Consistency in action.

⁸ As a referee has pointed out to me, it is a consequence of this explanation of acting under the idea of freedom that an agent cannot count as rational and yet act according to a principle of which she is aware that it is not universalisable; in other words, she cannot consciously act in an immoral way without being irrational. Although it might be accepted, on this conception, that a person is generally rational, her immoral acts must at least be considered as evidence of lapses of rationality and are not done under the idea of freedom. Does this mean that free evil action is impossible on this conception? – The answer to this question depends on whether a free action must be done under the idea of freedom as I have explained it. In this paper, I am not concerned with the question of what the conditions for free actions are (which might include evil actions). My interest is only in certain conditions for rational actions (which do not, on the present conception, include evil actions). The question of whether these conditions coincide merits a deeper discussion and I do not wish to pretend to have solved this venerable Kantian problem.

⁹ Christine Korsgaard confirms this when she says that the point of Kant's claim that we must act under the idea of freedom "is not that you must *believe* that you are free, but that you must choose *as if* you were free. It is important to see that this is quite consistent with believing yourself to be fully determined" [Korsgaard (1996), p. 162]. Her argument is that, even if you believe yourself to be determined, "[i]n order *to do* anything, you must simply ignore the fact that you are programmed and decide what to do – just as if you were free." [Ibid., p. 163]. It is implicit in Korsgaard's formulation that choosing or deciding requires a mindset "as if you were free", which she then describes as "regard[ing] ourselves as having free will" (ibid.). Her idea seems to be that this regard is part of a practical standpoint, which is independent from theoretical beliefs about our decisions. On my analysis, acting under the idea of freedom is not about "regarding oneself" in a special, positive way; it just means *not* being aware of acting according to an invalid principle; with less precision, we might say, in Korsgaard's terms, *not* "regarding oneself" as so acting. Formulated this way, it is clearer why such action is compatible with belief in determinism and denial of freedom. Following Korsgaard, acting under the idea of freedom while believing in (freedom-denying) determinism requires a state of mind where one believes that *p* and then acts as if not *p* (ignoring that *p*). Is this consistent, as she affirms? One way of making it consistent, is claiming that acting as if not *p* means acting under the mere *assumption* that not *p*. One can consistently assume that not *p*, say, as part of some procedure, while believing that *p*. (I am grateful to an anonymous reviewer for pointing this out to me.)

¹⁰ Galen Strawson (2010) shows how difficult it is to find a coherent place for genuine, self-determining freedom in a theory of deliberation.

¹¹ My point is not that deliberation cannot ever be defeated by a belief. A subject might believe herself to be incapable of deliberation (for whatever reason) and *thereby* become incapable of deliberating. My point, in this paper, is that neither determinism *per se*, nor a great variety of conceptions of deliberation necessarily have this defeating effect.

REFERENCES

- CASSAM, Q. (1997), *Self and World*; Oxford, Clarendon Press.
- HENRICH, D. (1998), 'The Deduction of the Moral Law: The Reasons for the Obscurity of the Final Section of Kant's *Groundwork*'; in Guyer, P. (ed.), *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, New York, Rowman & Littlefield, pp. 303-343.
- HILLER, F.R. (2016), 'The Deductions of Freedom/Morality-as-Autonomy and the Categorical Imperative in *Groundwork* III and Their Problems'; *Tópicos*, vol. 50, pp. 61-94.
- KANT, I. (2012), *Groundwork of the Metaphysics of Morals* [=G]; edited and translated by M. Gregor and J. Timmermann, Cambridge, Cambridge University Press.
- KORSGAARD, C.M. (1996), *Creating the Kingdom of Ends*; Cambridge, Cambridge University Press.
- O'NEILL, O. (1989), *Constructions of Reason*; Cambridge, Cambridge University Press.
- PEACOCKE, C. (1989), *Transcendental Arguments in the Theory of Content*; Oxford, Clarendon Press.
- PEREBOOM, D. (2006), 'Kant on Transcendental Freedom'; *Philosophy and Phenomenological Research*, vol. 73, pp. 537-567.
- STRAWSON, G. (2010) [1986], *Freedom and Belief. Revised Edition*; Oxford, Oxford University Press.
- ULRICH, J.A.H. (1788), *Eleutheriologie oder über Freiheit und Nothwendigkeit*; Jena, Crökersche Buchhandlung.