



CAUSAL REFUTATIONS OF IDEALISM

BY ANDREW CHIGNELL

In the 'Refutation of Idealism' chapter of the first 'Critique', Kant argues that the conditions required for having certain kinds of mental episodes are sufficient to guarantee that there are 'objects in space' outside us. A perennially influential way of reading this compressed argument is as a kind of causal inference: in order for us to make justified judgements about the order of our inner states, those states must be caused by the successive states of objects in space outside us. Here I consider the best recent versions of this reading, and argue that each suffers from apparently fatal flaws.

In the 'Refutation of Idealism', Kant famously claims that the necessary conditions for having certain kinds of mental episodes, episodes that even the Cartesian sceptic admits to having, are sufficient to ensure that there are 'objects in space' outside us (B275).¹ An influential way of reading this notoriously compressed argument is as a kind of causal inference: in order for us to make justified judgements about the mere temporal order of our inner states, those states must be *caused* by the successive states of objects in space outside us.

This sort of causal reading of the Refutation ('causal refutation' for short) has been prominently championed in the English-speaking world by Paul Guyer, among others.² In a pair of recent publications, Georges Dicker defends an updated version of the causal refutation, one that carefully avoids many of the now common objections to it.³ Interestingly, while Dicker agrees with Guyer that the causal refutation has independent merits as an

¹ Quotations from Kant's work are from the Akademie Ausgabe, with the *Critique of Pure Reason* [1781/7] cited in the text by the standard A/B edition pagination, and other works by [volume:page]: Kant, *Gesammelte Schriften* (Berlin: Königlich-Preussischen Akademie der Wissenschaften zu Berlin (now Walter de Gruyter), 1902–). English translations here diverge only rarely from the translations in P. Guyer and A. Wood (eds), *The Cambridge Edition of the Works of Immanuel Kant* (Cambridge UP, 1992–).

² P. Guyer, *Kant and the Claims of Knowledge*, henceforth *KCK* (Cambridge UP, 1987), and 'Postulates of Empirical Thinking in General and the Refutation of Idealism', in G. Mohr and M. Willaschek (eds), *Immanuel Kant, Kritik der reinen Vernunft: Klassiker Auslegen* (Berlin: Akademie Verlag, 1998), pp. 297–324.

³ G. Dicker, 'Kant's Refutation of Idealism', henceforth *KRI*, *Noûs*, 42 (2008), pp. 80–108, and *Kant's Theory of Knowledge* (Oxford UP, 2004).

anti-sceptical argument, the two commentators differ radically about its relationship to Kant's transcendental idealism (TI). Guyer (*KCK*, pp. 335, 415) interprets (TI) as a metaphysical mind-dependence thesis about space and time, and then criticizes it as a 'degrading' reduction of empirical reality that Kant would have done better to avoid. One of his central aims is thus to show that the Refutation is independent of (TI), and that the proof Kant offers can be used *malgré lui* to demonstrate the existence of 'ontologically independent' spatiotemporal objects (*KCK*, pp. 291ff.).

Dicker, in contrast, argues that the causal refutation is a plausible anti-sceptical argument which is consistent with Kant's (TI), so long as we interpret the latter as the 'weak' doctrine that we cannot perceive or even think of things as they are 'in themselves' – i.e., independent of the spatiotemporal forms in which they appear to us.⁴ In what follows, I argue that although Dicker's new formulation of the causal refutation succeeds in advancing the discussion of Kant's perennially intriguing proof, it still faces serious textual and philosophical difficulties. I do not mean to suggest that *no* interpretation of the Refutation argument can be successful; on the contrary, in other work I hope to develop an interpretation which is largely based on *conceptual-semantic* rather than *causal* considerations, and is thus not vulnerable to the objections raised below. My conclusion here is accordingly the more limited one that the *causal* refutation, even in its most recent and sophisticated guise, faces formidable, albeit very instructive, obstacles as both interpretation and argument.

I. FIRST FORMULATION

Here is Kant's own articulation of the proof (the emphasis in bold is his, and the italicized portion is his own correction to the original text):

Theorem

The mere, but empirically determined, consciousness of my own existence proves the existence of objects in space outside me

Proof

I am conscious of my existence as determined in time. All time-determination presupposes something **persistent** in perception. *This persistent thing, however, cannot be an intuition in me. For all grounds of determination of my existence that can be encountered in me are representations, and as such require something persistent that is distinct even from them, in relation to which their change, thus my existence in*

⁴ Dicker, *Kant's Theory of Knowledge*, ch. 9; KRI, pp. 100ff.

*the time in which they change, can be determined.*⁵ Thus the perception of this persistent thing is possible only through a **thing** outside me and not through the mere **representation** of a thing outside me. Consequently, the determination of my existence in time is possible only by means of the existence of actual things that I perceive outside myself (B275–6; Bxxxix n.).

Here is a first pass at a reconstruction (cf. Dicker, KRI, p. 82):

1. I am conscious of my own existence in time, i.e., I am aware that I have experiences that occur in a specific temporal order
2. I can be aware of having experiences which occur in a specific temporal order only if there is some persisting element by reference to which I can determine their temporal order
3. No conscious state of my own can serve as this persisting frame of reference
4. Time itself cannot serve as this persisting frame of reference
5. If (2) and (3) and (4), then I can be aware of having experiences which occur in a specific temporal order only if I perceive persisting objects in space outside me by reference to which I can determine the temporal order of my experiences
6. Therefore I perceive persisting objects in space outside me by reference to which I can determine the temporal order of my experiences.

A few clarificatory notes: first, ‘experience’ here is a broad term encompassing any series of what Kant calls ‘representations’. These include perceptual representations as well as various thoughts, imaginings, judgements, and the like. Kant typically excludes ‘feelings’ (*Gefühle*) from the class of representations, since (according to him) ‘feelings’ fail to represent (cf. *Anthropology* 7:32, 153). But there is nothing in the argument to prevent it from applying to a temporally ordered series of feelings as well, so I shall assume here that Kant is really talking about any series of what we now call ‘mental states’.

Secondly, the argument is about the ‘temporal order’ of experiences, and so might refer either to a successive series, or to two or more simultaneous states. In the Third Analogy, Kant defends a principle about the necessary interactivity of nature which might be thought to ground a kind of causal refutation of sceptical idealism on the basis of our knowledge that some of our mental states are simultaneous. This is a complicated proposal which

⁵ In the new Preface, written just before the B-edition went to press, Kant asks the reader to insert the italicized sentences here in lieu of the second sentence of the ‘Proof’ (apparently it was too late for him to change the actual body of the text). That sentence reads ‘This persistent, however, cannot be something in me, since my own existence in time can first be determined only through this persistent thing’ (B275).

has not yet been fully explored in the literature. One reason to avoid bringing the complication in here is that in the B-edition Kant explicitly says that the Third Analogy principle only applies to objects *in space* (B256).⁶ In the context of the Refutation, however, we start with the temporal order of our *mental states* in order to *conclude* that there are objects in space. So it may seem question-begging to rely on the Third Analogy principle as a premise. In any case, I propose to follow the vast majority of commentators in focusing on our knowledge of internal *successions* of experiences.

Thirdly, Kant's writings leave it unclear whether the 'awareness' in (1) and (2) is supposed to be the mere passive consciousness of altering representations, or the justified judgement *that* one's representations have altered in a specific time-order, or some other alternative. There is now a near-consensus among commentators, however, that some sort of *judgemental* awareness is required if the second premise is to have any plausibility.⁷ Thus (1) is read as the claim that I can 'correctly determine' (i.e., have a justified judgement or knowledge) that a series of mental states occurred in a specific temporal order, and (2) is read as stating a necessary condition for having this ability.

There is less agreement, finally, about whether the starting-point of the Refutation is the mnemonic ability to know that two or more *past* experiences occurred in a specific temporal order, or whether reference to a present mental state can be included. Dicker follows Jonathan Bennett among others in claiming that the focus must be exclusively on remembered states.⁸ In other words, Dicker and Bennett claim that the necessary condition in (2) will only be plausible if we take 'experiences that occur in a specific temporal order' to refer to a series of *past* experiences, rather than to any experiences we are having in the specious present. In support of this, Dicker (KRI, pp. 85–6) cites his intuition that it would be 'absurd' to suggest that we cannot tell the difference, just by introspecting, between a present mental state and a past mental state. The claim, I take it, is that worrying that one's memories will *consistently* seem more 'present' to one than one's actual present state is about as sensible as worrying, with Descartes' madmen, that one's head is a gourd or that it is made of glass. Some sceptical scenarios simply do not gain traction against common sense.

⁶ See E. Watkins, *Kant and the Metaphysics of Causality* (Cambridge UP, 2005), ch. 3.

⁷ See M. Gochnauer, 'Kant's Refutation of Idealism', *Journal of the History of Philosophy*, 12 (1974), pp. 195–206, at p. 198; H. Allison, *Kant's Transcendental Idealism*, 2nd edn (Yale UP, 2004), p. 289; Guyer, *KCK*, pp. 293ff.; Dicker, KRI, p. 82.

⁸ J. Bennett, *Kant's Analytic* (Cambridge UP, 1966), p. 205. Guyer disagrees with Dicker here (*KCK*, ch. 13), as do Allison, *Kant's Transcendental Idealism*, pp. 290–1; L. Caranti, *Kant and the Scandal of Philosophy* (Toronto UP, 2007), pp. 135ff.; J. Vogel, 'The Problem of Self-Knowledge in Kant's "Refutation of Idealism"', *Philosophy and Phenomenological Research*, 53 (1993), pp. 875–87.

Kant himself does not talk about memory in the Refutation chapter, and so it is hard to know whether he would view this kind of memory scepticism as absurd. *Contra* Hume, it does seem that we *occasionally* have memorial ideas which are phenomenologically more vivacious than, say, the repetitive real-time impressions of the inside of the subway car as it monotonously travels from one stop to another. But it is not clear, even in such a case, that we could not tell, just by introspecting, which of the two episodes is past and which is present. Whether it is possible *consistently* to be unable to tell the difference between past and present mental episodes just by introspecting is even more difficult to say. As a result, I propose to set this issue aside and accept the interpretation of (1) and (2) as focused wholly on states of the past. (Later, however, I shall raise a problem that arises as a result of this move.)

II. SECOND FORMULATION

Here is a new and longer formulation of the argument,⁹ one that takes into account the issues just discussed:

- 1'. I can correctly determine the order in time of my own past experiences
- 2'. When I remember two or more past experiences of mine, my recollection of those experiences does not itself reveal the order in which they occurred
- 3'. If (2'), then I cannot correctly determine the order of two or more past experiences of mine just by recollecting them
- 4'. I cannot correctly determine the order of two or more past experiences of mine just by recollecting them [from (2') and (3')]
- 5'. If I cannot correctly determine the order of two or more past experiences of mine just by recollecting them, then I can correctly determine the order of two or more past experiences of mine only if I know that at least some of my experiences are caused by successive objective states of affairs that I perceive
- 6'. I can correctly determine the order of two or more past experiences of mine only if I know that at least some of my experiences are caused by successive objective states of affairs that I perceive [from (4') and (5')]
- 7'. I know that at least some of my past experiences are caused by successive objective states of affairs that I perceive [from (1') and (6')].

Clearly (1'), (2') and (5') are the controversial premises here. I shall evaluate each in turn before moving to a third and final formulation of the argument.

⁹ Cf. Dicker, KRI, pp. 86, 92. I have made adjustments to Dicker's formulations here for the sake of readability, and also retained the focus on memorial states throughout.

II.A. *Regarding (1')*

1'. I can correctly determine the order in time of my own past experiences.

Again 'correctly determining' the time-order of past experiences involves making justified judgements that, for example, my experience of walking home occurred after my experience of swimming in the sea. (1') does not assert that we often *do* bother to make such judgements, but rather that most of us *could* do so, and that if and when we do the judgements are justified and (*ceteris paribus*) true.¹⁰

What sort of sceptic will grant premise (1')? Not Bertrand Russell's memory sceptic who insists that for all we know, we might have come into existence five minutes ago with a collection of false memories. Nor a memory sceptic who allows that we have had a lengthy series of past experiences but insists that for all we know, we are *now* just having *quasi*-memories of them, massively scrambling their order and content.

What about the Cartesian sceptic of the first two *Meditations*? He is Kant's explicit target in the Refutation, of course, but would even he accept (1')? The question is surprisingly difficult to answer. In the first *Meditation*, the meditator says that he seeks to doubt *everything* he has assumed up to now and, in the second *Meditation*, that he is trying to 'believe that my memory tells me lies, and that none of the things it reports ever happened'.¹¹ It was presumably such passages that led later philosophers like Reid to construe Descartes' sceptic as challenging the deliverances of memory as well as those of sense-perception.¹²

The meditator is hardly a reliable memory sceptic, however, and there are other texts that push in the opposite direction. In the third *Meditation*, he baldly states it as a fact that 'I exist now and remember in addition having existed formerly' in order to generate his clear and distinct ideas of duration and number. In the fifth *Meditation* he relies on 'remembering that I clearly and distinctly perceived' the truth of a proposition in the past, all without first establishing the reliability of memory (AT VII, p. 70). So it remains unclear whether the meditator of the early *Meditations* would accept (1'). What *is* clear is that Kant's argument would have greater anti-sceptical force if it could mollify one or more of the memory sceptics just described. Dicker seems to recognize this at the end of his paper (KRI, p. 104), and

¹⁰ Dicker admits that some children may not have the reflective capacities required to satisfy (1'), but he does not think this undermines the argument's anti-sceptical force for adults who do: KRI, pp. 96–7.

¹¹ Descartes, *Meditations on First Philosophy* [1641], in C. Adam and P. Tannery (eds), *Oeuvres de Descartes* (Paris: Vrin/CNRS, 1974–86), Vol. VII, p. 24.

¹² Reid, *Inquiry into the Human Mind on the Principles of Common Sense* [1764] (Edinburgh UP, 2001), p. 208.

revises the first premise once again so that it says merely that ‘I *seem* to be able to determine correctly the order in time of experiences of mine’. I shall argue (in §III below) that this move to seeming-statements in an attempt to handle memory scepticism fatally weakens the argument as a whole.

II.B. *Regarding (2’)*

2’. When I remember two or more past experiences of mine, my recollection of those experiences does not itself reveal the order in which they occurred.

I have already said something about why this premise focuses exclusively on recollections. There are still threats to it, however: a main one arises from the possibility of ‘Bennett-like’ episodes; another stems from the possibility of ‘clock-like’ features.

‘Bennett-like episodes’ are recollections which contain sufficient information for the subject to determine the temporal place of one past state relative to some other past state.¹³ An example is your state of thinking-about-the-first-half-of-the-game-while-watching-the-second-half. The complex content of that state contains in itself the information that your experience of the first half of the game must have occurred before your experience of the second half. Given that you are presently and veridically recollecting that you had that state, it seems you can now justifiably judge that the experience of the first half came before the experience of the second half.

There are many varieties of Bennett-like episodes, but (as Bennett himself concedes) comparatively few of our states in fact contain time-determining information of this sort. So even if *some* inner states wear their relative temporal order on their sleeves in a Bennett-like fashion, the Refutation argument will still apply with respect to the vast majority of states that do not. It is crucial to note, however, that this argument adverts to a contingent psychological fact: it certainly seems *possible* for someone’s mental life to include very many Bennett-like episodes, just as it seems possible for someone’s mental life to include hardly any Bennett-like episodes at all. This means that (2’) cannot be put forward as a necessary truth: if it is true at all, it is only contingently true. That said, I shall grant from now on that the ‘past experiences’ referred to in the premises of our argument do not have the kind of content characteristic of a Bennett-like episode.

Would the Cartesian sceptic accept (2’)? If it were meant to be a necessary truth, then I have already shown that the answer is ‘No’. Moreover, even setting Bennett-like episodes aside, there is no reason why there could not be a persistent tone that slowly rises in pitch over the course of my life,

¹³ Bennett, *Kant’s Analytic*, pp. 228–9.

or a ‘voice in my head’ quietly counting off each second from the moment I am born until the moment that I die, or a series of changing geometric shapes that float slowly up from the bottom of my visual field to the top of it over the course of my life.¹⁴ Each of these scenarios introduces a ‘clock-like’ feature into my mental states that allows them to wear their relative temporal order on their sleeves, so to speak, without involving a Bennett-like episode.

Causal refuters typically acknowledge this point by noting that although there *could* have been such ‘tell-tale clocks or stamped-on dates’ in our experience (Guyer calls them ‘digital timers’), in actual fact there are none (Dicker, KRI, p. 88; Guyer, *KCK*, p. 244). In other words, causal refuters admit that (2′) is not a necessary truth, but rather a merely contingent truth about our psychology. (Of course, if someone *does* experience a digital timer ticking off seconds in the corner of his visual field, or an aroma that subtly changes throughout his life, or a voice counting off seconds in his aural field, or something similar, then *he* may not be able to accept (2′). That means that *he* may remain stuck in the sceptical predicament, at least as far as this version of the Refutation argument is concerned.) I shall argue in §III that conceding the bare contingency of (2′) in this way spells far more trouble for the causal refutation than its defenders seem to realize.

II.C. *Regarding (5′)*

5′. If I cannot correctly determine the order of two or more past experiences of mine just by recollecting them, then I can correctly determine the order of two or more past experiences of mine only if I know that at least some of my experiences are caused by successive objective states of affairs that I perceive.

A premise along the lines of (5′) is clearly going to be the linchpin of any causal refutation. The premise says that if I cannot discern the relative order of my past mental episodes from the information contained in my recollection of them, then my judgements about this order are only going to be correct and justified if I know (or at least justifiably judge) that the episodes are severally simultaneous with the successive states of something else. That something else, moreover, must be (5a′) perceived by me; (5b′) numerically distinct from me; (5c′) causally responsible for the episodes in question; and hence (5d′) objectively ‘in space outside me’.

In order to support (5′) textually, causal refuters place a great deal of weight on one of Kant’s unpublished reflections from 1788–90:

¹⁴ The persisting tone suggestion is from Strawson, *The Bounds of Sense* (London: Methuen, 1966). Karl Ameriks suggested the floating shapes scenario in conversation.

[Succession in inner sense] cannot be represented except by means of something which endures, with which that which is successive is simultaneous. This enduring thing, with which that which is successive is simultaneous, that is, space, cannot in turn be a representation of the mere imagination but must be a representation of sense, for otherwise that which lasts would not be in the sensibility at all.¹⁵

Guyer (*KCK*, pp. 306–8) reads this as saying that ‘successive representations in one’s experience can be [correctly] judged to be successive only if they are [correctly] judged to be severally simultaneous with the severally successive states of some enduring object[s]’ which we perceive. This is then supposed to ‘entail that those objects are objects *acting on the self* ... enduring objects are conceived as agents of the empirical succession of self-consciousness’.

An initial concern about the causal refuter’s use of this passage is that it is taken from just one unpublished *Reflexion*, and it is in tension with what Kant says in other notes as well as in the Refutation text itself. There is more to be said here, but I set this textual concern aside for present purposes.

A second concern is that the claim about causation which is putatively entailed (though not explicitly stated by Kant) makes the whole passage seem like a throwback to the anti-sceptical proof of the *Nova Dilucidatio* of 1755 (1:410–11). That argument was *explicitly* causal in nature, and this passage as well as premise (5’) as a whole seems to fit better with the transcendental realism about time that accompanied it. For if the ‘successive objective states of affairs’ that caused our past mental episodes are not themselves in time, it is hard to see how their causal agency could justify our judgements about the relative time-order of the effects.

Guyer (*KCK*, pp. 290–1, 308–9) concedes this second point, and admits that his causal refutation is an epistemological version of the pre-Critical argument and thus difficult to reconcile with Kant’s transcendental idealism about time. But, as noted above, Guyer thinks that (TI) about time is a hopeless doctrine anyway, and so he does not propose to make it fit his interpretation. Dicker, in contrast, avoids taking a stand on the issue by weakening the fifth premise once more, so that it concerns only how the causes of our inner states *seem* to us to be, and not whether those causes are *really* temporally ordered or not. Again I shall say more about this ‘seeming’ formulation of the argument in §III below, but for now I propose to consider the various components of the fifth premise, beginning with (5a’).

(5a’) *Perceived by me*

The argument for (5a’) is simply that

¹⁵ This passage (R6313, 18:614) is cited as crucially important by both Guyer, *KCK*, p. 305, and Dicker, *KRI*, p. 83.

we could not possibly correlate remembered experiences with successive states of an enduring reality if, when those states occurred, they were unperceived or unperceivable. One cannot correlate As with Bs if Bs are unperceived – if Bs do not enter into our experience (KRI, p. 93).

As a general claim, this just seems incorrect: surely there could be an argument which would justify me in *conceiving* of some set of As (e.g., certain features of physical objects) as correlated with some set of Bs (e.g., certain features of electrons), even if I never *perceive* the Bs. Moreover, an argument for the existence of external objects (as opposed to any particular states of them) would require no more than this, i.e., that we justifiably conceive of some of our perceived mental states as correlated with the objective states of some external objects, perceived or not. Clearly the idea here is that the only way we could justifiably *correlate* experiences with states of an enduring object is if we also perceive that object. But this is a substantive premise which would require defence.

(5b') *Numerically distinct from me*

The causal refutation also says that in order to make justified judgements about the time-order of our fleeting inner states, we have to be able to correlate our states with perceived objective states of affairs that are 'numerically distinct' from the self. But why should we accept the latter part of this claim? Again, why could not my own self, or some of its states, play the role of objective correlate here? We can grant straight away that what Kant calls the 'logical' or 'transcendental' self will not be able to play this role, since its concept is that of a bare apperceiver, a mere 'I think', which can be attached to any mental episode and does not persist as a distinct substance or cause throughout those alterations (A349ff.; B277).

But what about the so-called 'empirical self', i.e., the self as it appears to us in inner sense? Kant sometimes refers to this as a phenomenal or empirical substance (see citations below): why could not *I* be the item that persists through inner alterations and with whose states I correlate those alterations? To put this in the terms of the *Reflexion* just cited: why could not my empirical self be the enduring item some of whose *other* states are perceived as severally simultaneous with a given internal succession?

Many commentators try to rule out this suggestion by citing a passage from the A-edition (A107) in which Kant appears to agree with Hume about the systematic elusiveness of the self in introspection:

The consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called **inner sense** or **empirical apperception**.

On Hume's view, of course, a subject is aware over time of impressions of thirst, quenching, dizziness, tiredness, pleasure, as well as of ideas of these, yet is unaware of any enduring self that underlies and exemplifies this series of impressions and ideas. If Kant thinks the empirical self is systematically elusive in this way, the argument goes, then it could not play the role of enduring object which the subject uses in making judgements about the temporal order of a series of perceived states.¹⁶

There are a number of reasons, both textual and philosophical, to think that this argument will not do. First, the A107 passage was deliberately removed by Kant from the B-edition of the first *Critique*. Perhaps he felt he had strayed dangerously close to the Humean 'bundle' view of the self there, or that he had failed to articulate his considered position.

Secondly, there are many passages in the *Critique* and elsewhere in which Kant appears implicitly or explicitly to reject the Humean doctrine, for instance,

... time is an *a priori* condition of all appearance in general, and indeed the immediate condition of the inner intuition (of our souls), and thereby also the mediate condition of our appearances (A34/B50–1).

Everything that is represented to a sense is to that extent always appearance, and an inner sense must therefore either not be admitted at all or else *the subject, which is the object of this sense*, can only be represented by its means as appearance, not as it would judge of itself if its intuition were mere self-activity, i.e., intellectual (B68; my italics; see also B69).

By means of external experience I am conscious of the existence of bodies as external appearances in space, in the same manner as by means of internal experiences I am conscious of the existence of my soul in time (*wie vermittelt der innern Erfahrung des Daseins meiner Seele in der Zeit bewusst*) (*Prolegomena* 4:336).

Kant's claim in these passages seems to be that we *are* aware in inner sense of the empirical subject, self, or 'soul' (*Seele*) as a phenomenal substance in which our mental states inhere. Even in the First Paralogism, where one would expect to find the strongest opposition to talk of a soul in inner sense, Kant discusses what sort of self we *can* be empirically aware of, and says that

we have no acquaintance with the subject *in itself* that grounds this I as a substratum, just as it grounds all thoughts. But meanwhile, one can quite well allow the proposition **The soul is substance** to be valid, if only one admits that this concept of ours leads no further, that it cannot teach us any of the usual conclusions of the rationalistic doctrine of the soul (A350–1; Kant's bold, my italics).

¹⁶ Dicker cites this passage at KRI, p. 81. Allison also uses it in defending his (non-causal) 'backdrop' reading of the Refutation: *Kant's Transcendental Idealism*, pp. 275ff.

A conclusive textual argument about this issue would require much more space than is available here. Still, these passages at least provide some indication that Kant's real opposition is to the idea that inner sense gives us the materials for an *a priori science* of the immaterial mind *in itself* as a substratum of all thought. This opposition does not entail commitment to the much stronger Humean elusiveness doctrine about the *empirical* self.

Thirdly, Kant lacks systematic grounds for claiming that we cannot cognize the empirical self in inner sense. In outer sense we only ever perceive states and properties of objects, and yet that sort of awareness *does* explicitly count, for Kant, as cognition of the empirical objects themselves, and not just as cognition of their states (see the passage from the *Prolegomena* quoted above, where Kant himself seems to draw this precise parallel). Indeed, the relevant category under which such experience is brought is the category of 'substance and accident' (A80/B106); there is no provision in Kant's table of categories for judgements about accidents that do not inhere in a substance. As far as I can see, then, there is no reason to think that while outer sense delivers cognition of altering states *and* a persisting substance, inner sense can only provide cognition of states.

Some commentators suggest that for Kant the diachronic differences in our mental states constitute *changes* (*Wechseln*) rather than *alterations* (*Veränderungen*), and that although alterations must occur in a relatively permanent substratum, mere changes need not do so. It is true that Kant sometimes makes a distinction between the two terms, for instance in the following *Reflexion*: 'Actually it is not a *modus* but rather the substance that alters. For that which is altered remains; the alteration (*Veränderung*) is merely the change (*Wechsel*) of its determinations' (R3579, 17:70). But the distinction would only support the Humean reading of Kant on the self if there were an independent reason to think that the empirical self cannot be cognized in inner sense, and thus that *only* changes rather than alterations are possible there. Again it is simply not clear that there is any such reason. Moreover, Kant is not at all consistent in his use of this terminology: he speaks of inner '*Veränderungen*' in the pre-Critical *Nova Dilucidatio* version of the Refutation proof (I:410–11), and he uses '*Veränderung*' *explicitly* to refer to *inner* states in passages throughout the Critical period (B48, B292, 4:544; 9:488; 15:245; 15:689; 15:854; 16:334; 17:615; 18:610; 20:338).¹⁷

Finally, even if Kant did hold the Humean elusiveness doctrine, we still would not be *required* to look to the external world to explain our justified

¹⁷ Thanks to Dina Emundts, Béatrice Longuenesse and Jonathan Vogel for a series of provocative conversations and correspondences on this point. Cf. Emundts, 'Kant über die innere Erfahrung', in U. Kern (ed.), *Was ist und was sein soll* (Berlin: Walter de Gruyter, 2007), pp. 189–205, and 'Die Paralogismen und die Widerlegung des Idealismus in Kants *Kritik der reinen Vernunft*', *Deutsche Zeitschrift für Philosophie*, 54 (2006), pp. 295–309.

judgements about inner time-order. This is because the objector's suggestion was really that we might be able to employ a series of *states* of the self, rather than the self as a whole, as the stable surrogate for time. If someone has a regular series of clock-like features in some of his states that allows him to make justified judgements about the relative time-order of his more fleeting experiences, then he will not have to turn to the external world at all. In other words, even if the empirical self as a whole is elusive and thus phenomenally insubstantial, there could still be ways of correlating a series of fleeting representations with other stable, clock-like features of our experience that effectively serve as a stand-in for time (like the slowly-changing tone or the counting voice or the upward-floating shapes).

If this is correct, then the only remaining defence of the claim that the empirical self or at least some subset of its states does not serve as the stand-in for time appears to be a baldly *empirical* and *a posteriori* one. In other words, the causal refuter has to assume as a premise that inner sense does not actually deliver the sort of stable clock-like features that would allow us to perform the time-ordering task, though even if this is a fact, it is a contingent one. This observation might be able to justify the claim in (5b') that we have to go outside the self to make justified judgements about our inner states after all, but it would do so only on contingent and *a posteriori* grounds.¹⁸

(5c') *Causally responsible*

So far I have shown that there are no compelling textual or systematic reasons to hold that the empirical self and/or some of its states cannot be perceived as persistent in relation to some other series of fleeting inner alterations. If that is right, then the defence of (5') as a whole cannot depend on a Humean elusiveness doctrine about the self. But perhaps the causal refuter does not need to defend anything as controversial as the latter. For his thesis in (5c') is that the relation between our fleeting inner states and the stable series of states which enables us to determine the time-order of the former is a *causal* relation. (Some commentators have resisted this thesis in favour of the claim that mere correlations or even a pre-established harmony would suffice, but I propose to bypass that whole discussion and simply grant that genuine causal relations between the two series of states are somehow required for us to determine correctly the order of our inner

¹⁸ Again some causal refuters admit this when pressed. Dicker says that Kant is trying to show us 'what conditions must be satisfied in order for us to be able to order our past subjective experiences, *given the kind of experiences we actually and unquestionably have*, and that kind of experience does not exhibit enough regularity or stability to enable us to order our experiences simply by reference to other experiences'. Thus, he says, the argument relies on a merely 'contingent fact about our subjective experiences' (KRI, p. 98). My claim below is that this is a much bigger problem for their reading than causal refuters seem to realize.

states.) Given this, the causal refuter really only needs to show that the self and its states cannot serve as the *cause* of some other series of representations in itself. I shall call this doctrine ‘*no inner causation*’ for short.

Unfortunately, although causal refuters spend a lot of time arguing for the importance of causal relations over mere correlations, they devote almost no time to defending *no inner causation* itself. Guyer, for example, issues (*KCK*, p. 308) a blanket dismissal of the possibility of such inner causation, remarking that nothing in the empirical self could ‘stand in a causal relation to the empirical self’. This is clearly correct if we are talking about the *entire* self causing the *entire* self, but it seems incorrect if it implies that no *part* of the self could cause some other part. Surely my thought of a friend might cause in me, by way of some associative link, a recollection of the time that we spent together in Scotland last summer. My desire to enter into a thought-experiment during an epistemology class surely might lead me to form the image of, say, a brain in a vat. Indeed, inner causation of this sort is so commonplace that the first *Meditation* sceptic uses it to support the claim that for all we know, some ‘hidden faculty’ in the self may be causing all of our experiences (*AT VII*, p. 7).

Kant himself also takes the possibility of inner causation seriously: he thinks that it threatens inferences which start from the features of particular inner states and arrive at claims about the character of outer objects. Such inferences, he says, are simply ‘unreliable, since the cause of the representations that we perhaps falsely ascribe to outer things can also lie in us’ (*B276*). Instead of a blanket dismissal of the possibility of inner causation, then, it seems that the causal refuter needs a principled defence of *no inner causation*. But it is very hard to see how that would go.

Suppose the causal refuter recognizes this and allows that inner causation within the self is possible after all, but goes on to argue that the part of the self doing the causing would be unnoticed and thus unable to help with the time-ordering of the series of subjective states that we consciously encounter in inner sense. I shall call the doctrine that there are no suitably *clock-like* inner series that we can use to determine correctly the order of other inner states ‘*no clock-like inner causation*’.

As far as I can see, the unargued assumption of *no clock-like inner causation* is just as indefensible as the dogmatic acceptance of *no inner causation*: there is simply no *a priori* reason to accept this claim about the features of our inner states. Dicker ultimately concedes this, admitting (*KRI*, p. 98) that *no clock-like inner causation* can only be defended

on the ground that human experience does not exhibit enough regularity or stability for us to establish this order [of our inner states] by reference only to subjective

experiences themselves. In a possible world in which the states that constitute the empirical self exhibited enough regularity, we might indeed be able to order our more short-lived past experiences by reference to our more regular or stable ones.

Once again, then, the causal refuter ultimately bases the move to the external world on the baldly contingent fact that we do not (usually) perceive features in inner sense that could explain our ability to determine correctly the relative time-order of our other past states.

(5d') *Objectively 'in space outside me'*

The claim that the objects that cause my inner states are in space is supposed to be a consequence of the doctrines that the required surrogate for time must be both (5b') numerically distinct from me and (5a') perceived by me. In the Aesthetic Kant had already argued that we can only perceive something numerically distinct from the self through outer sense, and that the form of outer sense is space. So if we know that an object is numerically distinct from ourselves, then we know that it will appear to us in space. Given that (5d') relies on (5a') and (5b') in this way, as well as on the results of the Aesthetic, there is no need to consider it separately here.

In the light of the discussion in this section, I submit that the linchpin premise of the causal refutation, *viz* (5'), can be grounded, if at all, on the same sorts of contingent psychological considerations as were used to support (2'). But the fact that the causal refutation has to start from contingent facts about our psychology like this, facts which are only knowable *a posteriori*, should now give serious Kantians serious pause. Does Immanuel Kant, the great anti-psychologistic defender of *a priori* armchair philosophy, really mean to refute the sceptic by pointing to the brute fact that we do not happen to have anything sufficiently clock-like in our experience? Can it really be *a posteriori* knowledge of a bare contingency that 'proves' the existence of an objective order of external spatial objects, and thus saves philosophy from the age-old 'scandal' of external-world scepticism (Bxxxix n.)? Throughout his career, Kant claims that philosophical demonstrations, just like mathematical proofs, must be performed *a priori*, and that their premises must be necessary and universal truths whose concepts are clearly elucidated. It is nearly impossible, given this, to believe that his own much vaunted 'proof' of the external world is based on merely contingent psychological premises.

Yet another and quite different problem for (5') arises from the fact that we are focused here on *past* inner states. The problem is that when it is thus focused, the argument seems at most to prove that there *were* external objects which I perceived, and with which my past states were causally

correlated. The formulation of (5') above is thus misleading, since its final clause says that 'my experiences *are* caused by successive objective states of affairs that I *perceive*'. In fact, nothing in the argument guarantees that I am not an unfortunate midlife victim of a Cartesian demon's attack, so that though I can still determine the order of my *past* states by means of my memories of the perceived external objects that caused them, I am nevertheless *presently* wholly deceived about my experience of the external world. If we make the formulation less misleading by construing the conclusion as that I am just as certain of the temporal order of my experience as I am that there *were* objects in space outside me, however, then the argument offers no succour to those who are worried by present-tense sceptical arguments.

III. THIRD FORMULATION

As mentioned briefly above, the causal refutation can be reformulated yet again in an effort to start from a much weaker premise about how *we conceive* of the causes of our inner series of representations, rather than one about how those causes *in fact are*. This final formulation evades many of the objections I raised above, including those involving the perception condition, memory scepticism, the temporality of the ultimate cause of our experience, and the causation condition generally, while still not rendering the result uninteresting. Here it is in Dicker's words, with the new seeming-statements in bold:

- 1''. I **seem to** be able to determine correctly the order in time of experiences of mine
- 2''. When I remember two or more past experiences, my recollection of those experiences does not itself reveal the order in which they **seem to me** to have occurred
- 3''. If (2''), then I cannot correctly determine the order in which two or more past experiences of mine **seem to me** to have occurred just by recollecting them
- 4''. I cannot correctly determine the order in which two or more past experiences of mine **seem to me** to have occurred just by recollecting them [from (2'') and (3'')]
- 5''. If I cannot correctly determine the order in which two or more past experiences of mine **seem to me** to have occurred just by recollecting them, then I can **seem to** be able to determine correctly the order in time of two or more past experiences of mine only if I conceive some of my experiences as being caused by objective states of affairs that I perceive

- 6''. I can **seem to** be able to determine correctly the order in time of two or more past experiences of mine only if I conceive some of my experiences as being caused by successive objective states of affairs that I perceive [from (4'') and (5'')]
- 7''. I conceive some of my experiences as being caused by successive objective states of affairs that I perceive [from (1'') and (6'')].¹⁹

This is a truly ingenious variation on the causal refutation theme. But there are still at least two significant objections to it. The first has to do with its logical form, the second with its attempt to deal with contingency.

III.A. *Shifting 'seems'*

In the previous, second, formulation, the key fifth premise was articulated as follows:

- 5'. If I cannot correctly determine the order of two or more past experiences of mine just by recollecting them, then I can correctly determine the order of two or more past experiences of mine only if I know that at least some of my experiences are caused by successive objective states of affairs that I perceive.

This premise is an instance of the schema {if I cannot perform A by ϕ ing, then I can perform A only by ψ ing}. 'A', of course, is substituted for in (5') by 'correctly determining the order of two or more past experiences of mine'.

But (5'') of the present, third, formulation instantiates a completely *different* schema, namely, {if I cannot perform A* by ϕ ing, then I can perform B* only by ψ ing}. Here A* is

A* correctly determining the order in which two or more past experiences of mine seem to me to have occurred

whereas B* is something completely different, namely,

B* **seeming to** be able to determine correctly the order in time of two or more past experiences of mine.

As far as I can see, nothing in the argument justifies the unannounced shift from A* to B* in (5''). A* certainly does not *entail* B*: suppose you know that you are drunk, or that a recent blow on the head has massively damaged your memory, and as a result you lose all confidence in your ability to determine the actual order of your past experiences. Then you might very well be performing A* but not B*, and you might even openly acknowledge

¹⁹ This formulation is a direct quotation of Dicker, KRI, p. 104.

this fact. Indeed, it looks as though a connection between A* and B* holds only if we have *already* ruled out various sceptical scenarios about memory. But those are precisely the scenarios which this version of the argument is designed to rule out, and so it would beg the central question to rule them out implicitly in the course of introducing one of the premises.

This whole problem can be avoided, of course, by refraining from shifting the 'seems' in the two parts of (5'). The placement of 'seem to' in (1') and (2'), however, is not accidental. It is located at the beginning of (1') so that we do not beg the question against the memory sceptic and claim *knowledge* that we can correctly determine the temporal order of our past experiences (cf. Dicker, KRI, p. 104). So the 'seem to' has to stay where it is. (2') cannot be amended either, however; the 'seem to' would not do any work if we moved it elsewhere, and in any case the proof would then be invalid.

The premise to change in order to salvage this formulation of the proof, then, must be (3'). Clearly it should be amended to read

3''*. If (2''), then I cannot **seem to** determine correctly the order in which two or more past experiences of mine have occurred just by recollecting them.

Then to preserve validity the fifth premise would have to be changed to

5''*. If I cannot **seem to** determine correctly the order in which two or more past experiences of mine have occurred just by recollecting them, then I can **seem to** be able to determine correctly the order in time of two or more past experiences of mine only if I conceive some of my experiences as being caused by objective states of affairs that I perceive.

These changes allow the argument to avoid begging the question while still being deductively valid. But is (3''*) at all plausible? In other words, is the contingent fact that 'the temporal order of some of my states does not reveal itself in their remembered content' sufficient to support the extremely strong claim that I cannot even *seem* (to myself or anyone else) to determine that order correctly when recollecting those states? I find it hard to imagine a sceptic going along with this.

III.B. *Contingent notwithstanding*

A second objection to this final formulation of the causal refutation is that the worry about the contingency of some of the key premises remains unaddressed. For (2'') is still grounded in explicitly *a posteriori* fashion: it is neither *necessary* nor knowable *a priori* that my experiences have contents that do not reveal the order in which they seem to me to have occurred; on the contrary, they clearly could have contained clock-like features. It is hard to

see how the causal refuter could keep this contingency from bleeding through to (4'') and (7'').

Dicker anticipates the objection, and responds (KRI, p. 105) by saying that (6'') rather than (7'') is the premise which really shows us the 'force of the argument' offered by the causal refutation. This is because (6'') says that 'I *must* conceive at least some of my experiences as being caused by objective states of affairs that I perceive, where the "must" indicates that my so conceiving them is a necessary condition of my being able to do something that I unquestionably can do'.

But how does the necessity of (6'') as a whole, even granting that it *is* necessary, justify the 'must' in the passage just quoted? Given the fact that (6'') is *based* on (2'')–(5''), the most it can express is the conditional 'Necessarily, if I seem to be able to determine correctly the order in time of two or more experiences of mine, and *as a matter of contingent empirical fact I have no sufficiently clock-like states by which to measure them internally*, then I conceive at least some of my experiences as being caused by objective states of affairs that I perceive'. Once this hidden conjunct in the antecedent is made explicit, it is clear that the conclusion in (7'') can be nothing more than that *as a matter of contingent fact I conceive some of my experiences as caused by successive objective states of affairs that I perceive*. Apart from being something the sceptic will hardly need to resist, this conclusion is now an exceedingly far cry from Kant's own, according to which, necessarily, there are 'objects in space outside me' (B275).

IV. DOES WEAK TRANSCENDENTAL IDEALISM HELP?

Dicker recommends the adoption of a 'weak' reading of transcendental idealism, weak (TI), according to which Kant's claim is that we can have no knowledge, nor even any conception of, what things are like apart from the spatiotemporal way they appear to us. Combining weak (TI) with the causal refutation is supposed to deflect the contingency objection I have just been making, since weak (TI) makes it necessary in some transcendental sense that we conceive of objects in the way we in fact do. It also allows a causal refuter to avoid Guyer's outright denial of transcendental idealism about time. I shall make three brief points about this part of the argument.

First, weak (TI) is not really idealism of any sort. This makes it deeply suspect as an interpretation of the historical Kant (as Dicker himself concedes, *Kant's Theory of Knowledge*, p. 47).

Secondly, it is clear throughout the Critical philosophy that we *can* use unschematized categories to develop conceptions of things as they are in

themselves, as long as we adhere to the law of non-contradiction (Bxxiv n.; 5:55–6). Indeed, Kant says that we *must* conceive or ‘think’ (*denken*) of things-in-themselves as affecting us, at the very least, in order to explain how it is that there are appearances at all (Bxxvii). Moreover, on most interpretations of the conclusions of the Aesthetic, we have to think of things-in-themselves as non-spatial and atemporal in order to explain how we can have synthetic *a priori* knowledge of space and time. All of this is permissible as far as Kant is concerned, as long as we do not then take ourselves to have *knowledge* (*Wissen*) or *cognition* (*Erkenntnis*) of the positive properties of particular things-in-themselves. Thus weak (II) seems no more to be Kant’s transcendental idealism than it is a form of idealism, though it may still be of independent interest.

Thirdly, and most significantly, weak (II) still does not help with the objection regarding contingency which I have laid out above. For even if we were to admit that we cannot know or conceive anything about the *things-in-themselves*, we could still conceive of spatiotemporal *appearances* as being experienced differently – for instance, as being experienced along with stable clock-like features. But then the causal refutation still leaves us with a merely contingent conclusion about how we *happen* to conceive of or experience appearances. Again this is a very far cry from Kant’s stated conclusion.

V. CONCLUSION

I have argued that even the best versions of the causal reading of the Refutation of Idealism are afflicted, at the very least, with a reliance on contingent psychological evidence acquired through *a posteriori* means, and thus with a very weak conclusion about how we happen to conceive of the causes of our experiences. The reading therefore seems inadequate as an interpretation of the high-octane demonstrative *a priori* ‘proof’ which Kant inserts into the B-edition *Critique*. I have also pointed out that the focus on the past, along with the ongoing possibility of midlife attack by a Cartesian demon, means that the strongest conclusion licensed by the causal refutation is that we in fact conceive of the causes of our *past* states as external objects which we *once* perceived.

Finally, I have indicated that weak (II) is not a plausible interpretation of Kant’s (II), and that in any case its adoption would be of little help with the central obstacle to the causal refutation. For even if we grant that we cannot conceive of things-in-themselves as different from appearances (or at all), the fact that we can conceive of our *experience* as being different from what it is (e.g., as more Bennett-like or clock-like) shows that the argument still leaves

us with a contingent conclusion about how we merely happen to conceive of the grounds of our experience. This is a conclusion which the sceptic simply has no reason to resist.

The case for abandoning causal readings of Kant's Refutation in favour of some alternative is thus, I submit, a strong one. What that alternative should look like is of course an open question. There are various interpretations on offer which are not fundamentally causal in nature and thus may not be vulnerable to the sorts of objections raised here. The gist of my own positive proposal is that, for Kant, correctly determining the temporal order of one's experiences involves spatial surrogates for time, and that the conditions required for having any spatial representations at all are also sufficient for the existence of a world of objects in space. On such a reading of the Refutation, the *primary* focus is the 'semantic' content of our experiences, the fact some of them designate spatial properties, rather than their causal relations to external objects (even if some such causal relations obtain). A full articulation and defence of this 'semantic' reading of the Refutation, however, is clearly a project for another occasion.²⁰

Cornell University

²⁰ For valuable discussion of the ideas in this paper, I am grateful to Karl Ameriks, Georges Dicker, Dina Emundts, Kristen Inglis, Béatrice Longuenesse, Jonathan Vogel, Eric Watkins, and two anonymous referees.