Kantian Review (2025), **00**, 1–25
doi:10.1017/S1369415425100782

KANTIAN REVIEW
CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Cognitive Spontaneity and the Organisation of the Understanding

Claudi Brink

University of New Hampshire, Durham, NH, USA
Email: claudi.brink@unh.edu

Abstract

I offer an interpretation of Kant's doctrine of cognitive spontaneity that explains how the understanding can function outside of the efficient-causal structure of nature, without being part of what McDowell calls 'the domain of responsible freedom'. Contemporary literature is dominated by the 'cognitive agency' approach, which identifies cognitive spontaneity with a kind of freedom. Against this view, the 'cognitive processing view' banishes agential notions from its account but also reduces the understanding to mere mechanism. I argue that neither of these interpretations is obligatory, motivating a teleological but non-agential account that resists assimilation into either of the current approaches.

Keywords: spontaneity; teleology; understanding; cognition

1. Introduction

One of Immanuel Kant's most lasting contributions to the philosophy of mind is his insistence on the active role of the subject in making sensory data intelligible. The sensory manifold, which forms the basis of our experience of the world, is not, by itself, intelligibly arranged and understood by us. We cognise the world, Kant insists, only because we are able to order inputs given by the senses in accordance with a universal rational structure that is inherent in our minds. The faculty of the mind that Kant claims is responsible for creating this order and in which this universal rational form resides, he calls the 'understanding' (*Verstand*).[1] To mark the special character of the understanding's ability to make the given sensible manifold intelligible by connecting elements in rationally recognizable ways, Kant calls it 'spontaneous'. My main aim in this paper is to offer an interpretation of what this doctrine of cognitive spontaneity amounts to.[2]

If we look at contemporary literature on this subject, a consensus has formed around an interpretative approach that ties Kant's account of spontaneity to notions of cognitive *responsibility* and *agency*. John McDowell, an influential proponent of this sort of view, famously holds that 'the point of the idea that the understanding is a faculty of spontaneity [is] that conceptual capacities are capacities whose exercise is

in the domain of *responsible freedom*' (McDowell 1994: 12, my emphasis). More recently, Konstantin Pollok writes that 'spontaneity, which is required for any judgement, means that the person who judges appropriates the content of, and can be held accountable for, that judgement' (Pollok 2017: 67). And in a similar vein, Bèatrice Longuenesse claims that 'the I think is, itself, an *Aktus der Spontaneität*' (Longuenesse 2017: 104), and that '[u]sing "I" in "I think" is just the conceptual expression of a consciousness of the rational unity of an act of thinking for which I take myself to be accountable' (p. 107). In this kind of view, which has for a while had widespread appeal, it is in virtue of the spontaneity of the understanding that we can hold the thinker accountable for her cognitive activity. Accordingly, Matthew Boyle says that the questions surrounding Kant's account of cognitive spontaneity are the very same questions that arise 'in more contemporary terms, in debates about the possibility of 'cognitive' (a.k.a. "doxastic," "epistemic") agency' (Boyle 2015: 35). Of course, these authors differ in substantive ways: they disagree about what it is that makes agents responsible for their cognitive acts, how we should understand such acts, the nature of the consciousness we have of these acts, and many other important details. But all agree that if we are to make sense of the spontaneity of the understanding, it will be in agential terms. They all sign up to some version of what I will call the 'cognitive agency' view.[3]

This was not always the case. In a seminal paper, discussing the spontaneity of the understanding, Wilfrid Sellars argues that 'Kant is leaving open the possibility that the being which thinks might be something "which is not capable of imputation". It might, in other words, be an *automaton spirituale* or *cogitans*, a thinking mechanism . . . ' (Sellars 1970: 24).[4] In the early 1990s, Patricia Kitcher undertook a sustained attempt to develop this kind of proposal. We could call the view Kitcher developed a 'cognitive processing' view. On this view,

> acts or processes of synthesis could not be performed by agents. They are unconscious activities within agents that enable them to have cognitive capacities required for agency. In Daniel Dennett's useful terminology, they are 'subpersonal' processes, not acts performed by persons. (Kitcher 1990: 122)

But Kitcher's original proposal has since fallen out of favour. As many commentators have discussed,[5] there are serious problems with this kind of view, and while recognisably similar proposals have occasionally surfaced in the literature,[6] it has largely been abandoned. Perhaps the most problematic element of Kitcher's proposal is that it views the mind in efficient-causal terms as part of the phenomenal realm of material nature, a view that many have argued is incompatible with core tenets of Kant's transcendental philosophy.[7]

Now, suppose that we want to avoid the strict naturalism of Kitcher's earlier view, but, with Sellars, we are motivated to articulate a model of spontaneity that does not draw on agential notions – one that would be available to readers who wish to deny the cognitive agency view. What might such a model look like? My project in this paper is to articulate such a model and to show that it is an attractive alternative to existing proposals. A proper understanding of the doctrine of spontaneity, I will argue, shows us that Kant ultimately views the mind and its various faculties by analogy with living beings, as having a unity that we cannot understand except in

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

teleological terms, that is, as deriving from a purpose or goal that structures its activity and organisation.[8] And as will come into view as we work through this model, end-directedness and accountability are conceptually independent: we can regard a system as end-directed without importing any agential notions. The teleological model thus opens space for an account that denies that the understanding is an efficient-causal system *without* assigning it any kind of freedom or drawing on the agential notions characteristic of the cognitive agency view.[9]

I proceed as follows. In Section 2, I describe the cognitive agency view and show that it rests on two substantive and controversial assumptions that might give us pause. My aim here is not to refute the view or even to argue against it, but to motivate us to look for an alternative model of spontaneity that would give readers who wish to question these assumptions something substantive to say about it. Section 3 presents the cognitive processing approach. While this approach does not rely on the same potentially problematic assumptions, it subscribes to a mechanistic conception of the understanding that I go on to challenge in the following sections. The key to seeing why the mechanistic view cannot be Kant's *and* to developing an alternative teleological model that resists assimilation into either of the existing alternatives is to investigate more closely the broadly rationalist metaphysical framework in which Kant elucidates the general concept of spontaneity. In Section 4, I argue that we can legitimately appeal to this framework, Kant's criticism of rationalist metaphysics notwithstanding. In Section 5, I show how this framework entails that spontaneous powers as such must be understood in teleological terms and therefore cannot be understood in terms of the efficient-causal explanatory framework appropriate to natural science. In Section 6, I then describe an implementation of the general teleological model of spontaneity for the spontaneity of the understanding in particular, and I show why the upshot is distinct from both the cognitive agency and cognitive processing approaches.

2. The cognitive agency approach

In this section, my aim is to characterise and critique the dominant interpretative trend in the literature on spontaneity.[10] As I alluded to above, a scholarly consensus has formed around an interpretation that I call the *cognitive agency* view. Now, for all its popularity, we will see that this view has substantive and controversial implications for a series of adjacent interpretative questions, and if we want an interpretation of spontaneity that does not commit us on these further issues, we should be motivated to explore alternative proposals.

According to the cognitive agency reading, Kant uses the term 'spontaneity', in the context of his account of the understanding, to signal a kind of freedom that the subject has in producing her cognitive states, in virtue of which she can be held accountable for those states.[11] And since Kant famously holds that the characteristic activity of the understanding is judgement (A69/B94), for proponents of the cognitive agency view, it is precisely in the activity of judging that we manifest the responsible freedom that constitutes spontaneity.

As we are about to see, a central underpinning of the cognitive agency view is a specific account of Kant's conception of judgement. To bring out what is distinctive about this account, it will be helpful to distinguish between what I call *purely semantic*

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

and *epistemic* accounts of judgement. Judging, on Kant's view, is an act that combines concepts in certain distinctive ways, and judge*ments*, the outputs of this activity, are complex representations comprised of concepts. We can distinguish two conceptions of the nature of this act and its corresponding output. On the *purely semantic* conception, judging simply puts contents together in certain distinctive ways, and the judgements it outputs are truth-evaluable propositions. On the *epistemic* conception, by contrast, judging is an activity that, in addition to putting contents together in distinctive ways to produce propositional contents, also involves a crucial *evaluative* component. The act of judging, on this view, involves an act of evaluation that culminates in an attitude that the judging subject takes to the proposition, and the outputs of this act are not merely truth-evaluable propositions but *attitudes* of the subject to the truth of such propositions.[12]

Given the attempt to connect cognitive spontaneity to the notion of cognitive agency, together with the assumption that the understanding manifests cognitive spontaneity through acts of judgement, the cognitive agency reader needs a conception of judgement on which acts of judgement are the kinds of acts for which the judging subject can be held accountable. The epistemic conception of judgement is thus extremely germane to the cognitive agency approach. If judgements are not merely propositional contents but acts of *endorsement* or *assent* – if to judge is not merely to generate a truth evaluable proposition but to *advance a claim* about the way the world is – it is easy to see why judgement could count as a manifestation of cognitive agency. It is thus no surprise that the overwhelming majority of cognitive agency readers assume the epistemic conception of judgement. To judge is not just to represent a certain content but to hold oneself accountable for its correctness, and that is why judging counts as an exercise of cognitive agency.

We move to a second important commitment of the cognitive agency approach if we now ask how this kind of accountability is possible. Advocates of the cognitive agency view insist that we could not be held accountable for acts of judging if we were not at least implicitly conscious of the *norms* governing judgement.[13] These norms are typically identified as either the categories themselves (see, for example, Boyle (2015: 53)) or the 'laws of understanding' considered more broadly (see, for example, Thomas Land (2021: 3146)), and our awareness of them is typically located in pure apperception. With this model in place, the cognitive agency reader is now equipped to read the entire Transcendental Analytic in normative epistemic terms.

The cognitive agency approach has a number of virtues. It draws together several important Kantian doctrines in a satisfying way, giving us a coherent and systematic explanation of the connections between spontaneity, apperception, judgement, and the categories. It also puts Kant into conversation with important work in contemporary epistemology on the ethics of belief and related topics, and it promises to forge a link between Kant's discussion of spontaneity in the theoretical domain and his discussion of spontaneity in the practical domain, where the notion of freedom as norm-responsiveness is a familiar one. It is perhaps not surprising, then, that it has become the dominant view in contemporary Kant scholarship.

Nevertheless, we have seen that this entire interpretative program rests on a very specific conception of both the nature of judgement and the laws of the understanding. Underlying the attempt to connect spontaneity, responsibility, and judgement, we have seen, are two specific interpretative claims: first, that Kant holds

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

the epistemic conception of judgement; second, that Kant conceives of the rules or laws of the understanding (which govern judgement or combination) as norms, which 'show up' (consciously or not) in judging. But these commitments are far from trivial. Each of these two claims is substantive and controversial, and I will argue that contemporary Kant scholarship gives us reasons to be hesitant about accepting them. If that is so, we should be motivated to articulate alternatives to the cognitive agency reading that do not rest on the same problematic textual foundations.

Let us start, then, with the claim that Kant endorses the epistemic conception of judgement. The first thing to note is that one does not find a single obviously agential notion in any of Kant's many characterisations of the act of judgement; instead, the primary notion is that of a distinctive *unity* of representations. It is true that notions such as *combination*, *apperception*, and *self-activity* often feature in these characterisations, and of course, it would be *possible* to gloss these notions in normative epistemic terms *if* one already assumed the cognitive agency reading. But none of these passages would convince an unbiased reader to adopt that framework in the first place.

It is also worth noting that the epistemic conception of judgement makes it awkward for the cognitive agency reader to accommodate the distinction between cognition (*Erkenntniß*) and knowledge (*Wissen*). One of the most important findings of recent Kant scholarship is the recognition that Kant systematically distinguishes these two notions.[14] On a prominent way of distinguishing them, *Wissen* is a propositional attitude that involves a specific act of assent ('holding-for-true' [*Fürwahrhalten*] of the kind analysed in the Canon), whereas cognition is a kind of referential thought that does not itself involve any act of assent.[15] As Clinton Tolley puts this point, in cognition 'nothing about the object (or anything else) needs to be claimed (held-true) at all' (cf. Tolley 2019: 3216–7). But note that on the plausible assumption that cognitions are a kind of *judgement*, the epistemic reading of judgement straightforwardly precludes this way of drawing the distinction between cognition and knowledge. After all, if all judgement involves an act of endorsement, and if cognitions are themselves a kind of judgement, it simply cannot be the case that nothing needs to be held true in acts of cognition. The cognitive agency reader thus owes us an alternative account of the nature of the distinction between cognition and knowledge. A challenge for any such account will be to explain why none of Kant's discussion of the epistemically significant attitudes of *Fürwahrhalten* in the Canon shows up in his treatment of cognition in the Transcendental Analytic.

The second claim underpinning the cognitive agency view – that the laws of the understanding are (or function as) norms that govern the act of judgement – is equally controversial. There is direct textual evidence that speaks against attributing to Kant a normative conception of the laws of understanding, and there is also a growing body of literature that argues for an alternative, 'constitutivist' conception of these laws.[16] Now, I make no claim to settle these issues here. But if we are hesitant about accepting a normativist conception of the laws of understanding, we should be equally hesitant about accepting an account of spontaneity that entails normativism.

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

3. The cognitive processing model

Suppose, then, that we were motivated to articulate an alternative model of spontaneity – one that would be available to readers who wish to deny the two controversial claims presupposed by the cognitive agency view. What resources would we find for such a view in the existing literature?

Kitcher's (1990) seminal work on Kant's transcendental psychology is perhaps the most influential statement of a kind of view that eschews agential notions in its account of spontaneity. On this sort of view, which I call the *cognitive processing* approach, *all* that is meant by spontaneity in the context of Kant's theoretical philosophy is a kind of 'structural creativity' in the mental operations of the understanding. This sort of view has largely fallen out of favour in the literature,[17] but we have reasons to explore it since it does not require either the epistemic conception of judgement or the normativist conception of the laws of understanding on which the cognitive agency view depends.

To begin, here is Robert Hanna articulating this view:

> A cognitive faculty is spontaneous in that whenever it is externally stimulated by raw unstructured sensory data as inputs, it then automatically organizes or "synthesizes" those data in an unprecedented way relative to those inputs, thereby yielding novel structured cognitions as outputs (B1–2, A50/B74, B132, B152). So cognitive spontaneity is a *structural creativity* of the mind with respect to its representations. (Hanna 2020)

In this kind of view, spontaneity can be understood simply in terms of the fact that the understanding can 'hold together' or 'structure' representations in a way that is not *given* in sensibility, which is, therefore, novel or creative relative to what is already cognitively available. According to the cognitive processing reader, the kind of spontaneity that the understanding displays in organising contents is not to be identified with any kind of freedom or control that the subject has over the generation of those contents. The process is *automatic*; it is simply a sub-personal psychological process that puts representations together in accordance with rules, which, as causal laws, are merely descriptive of these very processes. Thus, the cognitive processing proposal tends to invoke a *descriptivist* account of the laws of understanding; as Kitcher puts the point, 'rules govern syntheses only as the law of gravity governs the movements of the planets: Theorists can appeal to these rules to describe what is happening' (Kitcher 1990: 83). What we are seeing, then, is that the cognitive processing model, unlike the cognitive agency reading, is compatible with denying both the epistemic conception of judgement and the normative conception of the laws of understanding. Indeed, the view comports more naturally with a *purely semantic* conception of judgement and a *descriptivist* account of the laws of understanding. Readers who share the reservations put forward in Section 2 should thus be motivated to re-visit the cognitive processing model.

However, I believe that the cognitive processing model as it currently stands suffers from a fundamental shortcoming. Advocates of the reading, in resisting the cognitive agency approach, have tended to treat the understanding's activity as

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

automatic, involuntary, and *mechanical*. Consider, for example, Susan Neiman's forceful expression of this mechanistic conception of the understanding:

> The understanding provides the conditions under which anything can be perceived as an object, yet it does so in a way Kant describes as reactive, dependent on the sensible world as the occasioning cause of its activity (VII; 71). Its operations are *routine, automatic, and mechanical*. (Neiman 1994: 49, my emphasis)

Now, admittedly, not all commentators express this commitment quite as explicitly as Neiman does here, but I take it to be a standard commitment of cognitive processing views such as Neiman's that there is some sense in which the operations of the understanding are simply part of the causal order of nature. To the extent that this is so, one would expect that the functioning of this faculty is explicable in the same causal mechanistic terms by which we explain natural phenomena in general, and indeed, Kitcher explicitly claims that '[t]he I that thinks will be phenomenal and causally determined' (Kitcher 1990: 139–40).

It is crucial that we distinguish *this* claim – that the understanding's operation is mechanical and causally determined – from the further claim that authors such as Neiman take themselves to be clarifying here, namely, that its operations are automatic and involuntary. The former claim, I am about to argue, cannot be Kant's view, for it is fundamentally at odds with his general account of how spontaneous powers function. As we will see, a proper acknowledgment of the spontaneity of the understanding is straightforwardly incompatible with a mechanistic account of the understanding's functioning – thus far I am in agreement with the cognitive agency view.[18] But it does *not* follow from this point, I will argue, that the understanding's actions cannot be automatic and involuntary in just the way the cognitive processing reader describes. Therefore, in rejecting the mechanistic conception of the mind that cognitive processing readers have tended to assume, we *do not* have to assign the understanding of the kind of freedom characteristic of agency, which means that there is room for an alternative view that resists assimilation into either the cognitive agency approach or existing cognitive processing approaches.

4. Kant's rationalist inheritance

For the remainder of this paper, my aim is to spell out an alternative to the cognitive agency view that also avoids the mechanistic implications of cognitive processing views. This is the teleological model that I have been alluding to, and my development of the view will take place in three sections. The view that I will attribute to Kant draws on the metaphysics of causal powers that he inherits from the Wolffian tradition;[19] in this section, I briefly introduce that framework before addressing concerns that it is incompatible with key tenets of Kant's Critical philosophy and therefore cannot inform an interpretation of the doctrine of spontaneity. With these concerns set aside, I move to the next section for a deeper exploration of this framework, explaining the place of the notion of spontaneity in it. It is here that I draw the connection between spontaneity and teleology. In the final section, I clarify

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

the proposal and emphasise the respects in which it breaks with the cognitive agency view.

The teleological reading that I will develop draws heavily on the metaphysical causal framework within which Kant explains the general concept of spontaneity. All substances, in this view, are causally active insofar as their powers serve to bring about the realisation of certain 'accidents', or properties. The historical tradition in which Kant writes applies this framework to the mind itself, conceiving of the mind (or soul) as itself a substance that is characterised by its possession of certain *powers* or *faculties*. These powers are assigned an important explanatory role, for it is in virtue of having these powers that the soul enters into and sustains its characteristic states – representing, feeling, desiring, willing, perceiving, and so on. Kant inherits this framework from his rationalist predecessors: Not only does Kant continue to use its characteristic terminology (that of faculty, power, act, and activity) throughout the critical period, but also he never once qualifies his use of these terms or distances himself from the traditional understandings of these concepts as they are developed in the Wolffian tradition. Within this framework, *spontaneity* and *receptivity* are well-defined technical notions, and I will argue that a proper appreciation of how they feature in this context illuminates the doctrine of cognitive spontaneity as it is developed in the first *Critique*.

Now, many commentators, understandably eager to distance Kant from the dogmatic doctrines of his predecessors, have ignored this framework in their attempts to give an account of the doctrine of spontaneity on the assumption that any rationalist metaphysics at all would be inconsistent with Kant's Critical project.[20] One worry is that treating the mind as a substance is incompatible with specific claims Kant makes in the Dialectic (particularly in the Paralogism chapter). And beneath it lurks a still deeper concern, which is that such a metaphysical framework violates Kant's thesis of noumenal ignorance: his claim that we cannot have theoretical cognition of things in themselves or of anything that is 'beyond the boundaries of experience' (A296/B535, see also, for example, A30/B45, A44/B62, A256/B312). We see this limitation of human cognition displayed in Kant's restriction of the application of the categories to appearances. Yet I aim to argue that the mind and its constituent powers are not part of phenomenal nature and that these powers cannot, therefore, be accounted for in terms of the efficient causal framework of natural science. My insistence on conceiving of the mind as a substance with causal powers, therefore, appears to be a violation of Kant's restriction on the use of the categories to what is given in sensibility. Before I proceed to make my positive argument, I will address each of these concerns in turn.

The first concern, then, is that Kant appears to explicitly rule out that we can legitimately characterise the mind as a substance in the first Paralogism, the conclusion of which is that 'I, as thinking being (soul), am a substance' (A349). Now, we know that Kant thinks there is something wrong with the inference that leads to this conclusion; the question is whether Kant thinks we should (1) reject the conclusion entirely, (2) reject this particular way of reaching the conclusion, or (3) reject one particular interpretation of the conclusion. My own view is that he recommends (3): Kant's opposition is not to the claim that the soul is a substance but to one specific interpretation of that claim. There is clear textual evidence that Kant, in fact, *endorses* the conclusion once properly formulated, for he explicitly says that 'everyone must

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

necessarily regard Himself as a substance, but regard his thinking only as accidents of his existence and determinations of his state' (A349; cf. 'Progress in Metaphysics', 20: 270–1; A400). This much, he seems to think, is entirely unobjectionable. What Kant objects to is the attempt to argue for a specific claim about the *nature* of the soul as substance – that the soul is simple and indestructible.[21] Motivating such a claim would require an 'empirically usable concept of substance' (A349), which Kant denies is available for this purpose. Nevertheless, we can make use of a concept that is devoid of any empirical (and specifically temporal) content, and in doing so, we can legitimately think of the mind as a substance – indeed, we have seen Kant claim that we must do so. The rationalist's mistake lies in attempting to employ a concept of substance to make specific synthetic a priori claims about the nature of the mind; in doing so, the rationalist illicitly relies on sensible temporal content that restricts the concept of substance to empirical use. But Kant's diagnosis of this mistake does not entail a wholesale dismantling of the metaphysical framework of causal powers within which the rationalist tradition situates the mind.

The reading of Kant's claim in the first Paralogism depends on a 'thin' understanding of the concept of substance, according to which a substance is just an existing thing that is a subject of inherence to certain accidents. Nothing in so conceiving of the mind requires one to endorse any claims about the temporal properties or the underlying nature of substance, and indeed, we might think that it is exactly such a conflation that is at the basis of Kant's criticism. Yet even this thin characterisation of the mind as a substance apparently violates Kant's restriction on the legitimate use of the categories to the domain of experience, for the mind itself, the subject of thought, is not given in experience. It appears then, that conceiving of the mind as a substance with causal powers, as my account proposes, violates Kant's self-imposed restriction on the scope of our knowledge and cognition.[22] In answering this concern, I will suggest that there is a use of the categories 'beyond' experience that is unproblematic and that is consistent with the 'thin' conception of substance that I have appealed to in my answer to the first concern. Such a use of the categories, I will suggest, does not give us theoretical cognition of objects but nevertheless plays an important role in Kant's Critical philosophy, and serves as a reminder that our knowledge is not exhausted by what we can theoretically cognise.

We can start by reminding ourselves of the restriction that Kant places on our use of the categories. This is a restriction we see Kant reiterate again and again in the first *Critique*, and it is worth citing several representative passages, both to bring out the claim itself and the frequency with which it is made:

> The categories consequently have no other use for the cognition of things except insofar as these are taken as objects of possible experience ... [The] further extension of concepts beyond our sensible intuition does not get us anywhere. For they are then merely empty concepts of objects, through which we cannot even judge whether the latter are possible or not – mere forms of thought without objective reality ... Our sensible and empirical intuition alone can provide them with sense and significance. (B148–9)

> Thus the schemata of the concepts of the understanding are the true and sole conditions for providing them with a relation to objects, thus with

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

> **significance**, and hence the categories are in the end of none but a possible empirical use … (A146/B185)
>
> Without schemata, therefore, the categories are only functions of the understanding for concepts but do not represent any object. This significance comes to them from sensibility … (A147/B187)
>
> [W]e cannot even define a single … [category] without immediately descending to conditions of sensibility, thus to the form of the appearances, to which, as their sole objects, they must consequently be limited, since, if one removes this condition, all significance, i.e., relation to the object disappears. (A240–41/B300)

What Kant appears to insist on over and over again, not only in these passages but also consistently and throughout the *Critique* (see also B145; A139/B178; B288–9), is that without relation to what is given in intuition, the categories are without 'significance' (*Bedeutung*), or 'relation to an object'. It certainly appears as if Kant is claiming in these passages that there can be no meaningful application of the categories beyond experience. How, then, can I insist that Kant thinks of the mind – something that is not given in experience – as a substance with causal powers? Both concepts, <substance> and <cause>, are categories, and Kant would appear to be violating his own restriction on the use of these categories if he indeed conceived of minds in those terms. It would appear, then, that my view has Kant making use of the categories that violate his own self-imposed restrictions on their significance.

Note, however, that this objection to my view rests on an interpretation of Kant's restriction thesis that bears the interpretative cost of making him grossly inconsistent. For it is well-documented that Kant does, in fact, employ the categories, and the category of cause specifically, beyond what is given in intuition.[23] Few commentators, for example, would deny that Kant makes such an unrestricted use of the categories in his practical philosophy, where he conceives of transcendental freedom as a distinctive non-sensible kind of causality. The proposition that we are transcendentally free might not amount to theoretical cognition, but Kant surely regards it as meaningful. Yet if the categories are meaningless beyond the boundaries of experience, he is not entitled to this attitude. Faced with this realisation, we could, with Jacobi, reject Kant's system as inconsistent, but a second, more charitable approach, would be to pursue an interpretation of the restriction of the categories that is compatible with them admitting of a super-sensible employment that is both meaningful and epistemically important. As Ameriks puts it,

> This strategy seems forced on us by minimal charity, for without anything like it, Kant would be talking wild nonsense in his many general discussions of things in themselves, in his many positive characterizations of them from a practical perspective, and in his many slightly more specific remarks about them as having 'distinctive and inner predicates'. (Ameriks 2003: 29–30)

Is there, then, a way of acknowledging the categories' lack of significance outside of what is given in experience that is nevertheless consistent with the claim that such a

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

use of the categories can be meaningful and important, and even perhaps indispensable for Kant's critical enterprise?

A number of commentators have suggested that there is. On the kind of view they defend, the categories have a purely intellectual content that is independent of the pure sensible content that is provided for them by the transcendental time-determinations of their schemata.[24] These commentators all admit that, in these contexts, Kant's use of the categorial concept (because it is devoid of a priori sensible temporal content) does not constitute theoretical *cognition* of the objects so characterised, but they insist that it is nevertheless a meaningful application of the concept that does important work within Kant's Critical philosophy. Eric Watkins, for example, distinguishes between two senses of 'significance' (*Bedeutung*): one is quite demanding and requires reference to an object. But there is another, less demanding sense of *Bedeutung* that does not require reference to objects, and therefore does not require any temporally determinate schematisation of the concepts in question. If the category is taken in this sense, as a representation of a property of an object, that is, as a representation of the way objects as such must be, regardless of whether they are given sensibly, then we have a grasp of a 'thinner' pure intellectual content of the concepts of the understanding. Here is Watkins:

> Kant's Critical project requires that the categories have meaning [*Bedeutung*] prior to their schematization, because we must use the unschematized categories to think things in themselves and we are surely thinking of things in themselves as having certain properties, even if we cannot refer to determinate or specific things in themselves by means of them (i.e., even if they do not have a specific *referent*). In other words, if the categories did not have meaning, we could not think of things in themselves and Kant's ethics (as one important instance in which thoughts of things in themselves are required) would be fundamentally inconsistent with his metaphysics. (Watkins 2002: 204)

Their use in the relevant contexts, admittedly, does not give us theoretical *cognition* of the objects in question. Cognition does require that the object be 'given' or shown to be 'really possible' – it requires the more demanding kind of significance characteristic of schematised concepts – but the use of the categories outside of theoretical cognition is not for that reason illegitimate or meaningless.

We might nevertheless still wonder of what use it might be to employ the categories in this way: even if it is not 'wild nonsense', as Ameriks puts it, to employ categories beyond experience, doing so does not yield theoretical cognition, and therefore we might worry that it is epistemically insignificant, that it can do no important work for Kant within his critical philosophy. However, as several commentators have noted (see especially Hogan 2009 and Chignell 2007 and 2010), Kant distinguishes a rich variety of epistemic states beyond theoretical cognition, such as belief (*Glaube*) and 'knowledge' (*Wissen*),[25] and plausibly thinks, at least in the context of our knowledge of ourselves as transcendentally free, that the application of the category of cause beyond what is sensibly given can yield *Wissen* even if it does not satisfy the strict requirements on theoretical *Erkenntniß*. We have reason to conclude, then, that Kant's restriction of the significance of the categories is consistent with not

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

only the meaningfulness but the epistemic value of applying them beyond the boundaries of sensible experience. It is against the backdrop of such a reading that I would like to claim that, although we do not theoretically cognise the subject of our thoughts, we nevertheless have knowledge of the mind as an existing thing (a substance), that is causally active, that is, that has certain faculties or powers that ground the inherence of accidents such as intuitions, perceptions, thoughts, desires, and so on in it.[26]

With these obstacles cleared, I now want to move on to show that we can develop an attractive interpretation of cognitive spontaneity according to which we should conceive of the operations of the understanding in the teleological, but non-agential terms that Kant develops to articulate his theory of organisms.

5. Kant's metaphysics of causality

I turn now to provide a sketch of the core elements of the causal metaphysical framework and explain how I see spontaneity as fitting into it. At the centre of this rationalist inheritance is the notion of causal power (*Kraft*).[27] On Kant's view, substances like individual bodies and minds, have 'accidents'. This means that they exist in particular ways, or they are in particular states. The idea of causal power is meant to explain how changes in those states of substances come about. When a causal power of a substance is exercised, Kant says that the substance acts (*handeln*): it actualises an accident, or it effectuates its existing in a particular state. Kant also says that this act or activity (*Tätigkeit*) *grounds* the inherence of the accident in the substance – which just means that it explains or is part of the reason for the substance's existing in the way that it does. Here is how he puts it in the *Metaphysik Mrongovius*:

> A substance, insofar is it contains the ground of that which belongs to the being of one thing, acts <agirt; G: handelt>; ... substance, insofar as its accidents inhere, is in action, and it acts insofar as it is the ground of the actuality of the accidents ... (Met-Mron, 29: 772-3[822-3])

Now, it is important to note that Kant distinguishes two fundamental kinds of activity. The first he calls, somewhat paradoxically, passivity or suffering (*Leiden*). The idea here is that some powers function simply to 'receive' certain properties from other powers – consequently, Kant calls such powers or faculties receptive (see, for example, 29: 823). The activity of a receptive power, therefore, constitutes a kind of suffering – the substance changes as a result of some external power affecting it. But, there is also a second kind of activity, which Kant calls *spontaneity* or self-activity (*Selbsttätigkeit*).[28] Now, it is this very concept, of course, that Kant uses to explain the nature of the power or faculty of understanding, and which he claims sets it apart from the *sensible* faculty of cognition, which he claims is *receptive*.

Both passivity and self-activity, as species of the genus, activity, can be understood in terms of the general conception of activity as that which grounds the inherence of an accident in a substance. What distinguishes the two kinds of activity is how the substance grounds the inherence: either a substance can ground the inherence of an accident through external or outer power, in which case it is passive (Met-Mron, 29:

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

823), or the substance can ground the inherence of an accident through its own internal or inner power, in which case it is spontaneous or self-active (Met-L1/Pölitz, 28: 268, 285; Met-Volckmann, 28: 448).

When a substance acts *spontaneously*, then, it does not receive an accident from an outer power, but it 'brings forth' its own accidents, which is to say, it changes its state from what Kant calls an 'inner principle'. As Kant says, '[T]he act <actus> of spontaneity cannot proceed from an outer principle' (Met-L1/Pölitz, 28: 285; cf. Volckmann, 28: 448; L2/Pölitz, 544). What that means is that the change in the state of the substance does not depend on any *external* power; instead, the change in the substance is determined by its nature.

Now that we have an initial handle on the general distinction between receptivity and spontaneity, I want to argue that, for Kant, each of these kinds of activity is subject to a distinct form of explanation: while beings with exclusively passive powers can be explained within the efficient-causal framework of natural science, spontaneous beings cannot. Such beings require an alternative explanatory framework that makes essential reference to final causes, ends, or purposes.

There is a specific kind of substance that Kant thinks can be characterised *exclusively* in terms of passivity – namely, what he calls 'phenomenal substance' or 'matter'. Kant says that

> [a]ll matter is lifeless, has no faculty for determining itself, . . . *For every matter remains in motion or at rest until it is altered by something else. Matter thus has mere receptivity or passivity*. (Met-Mron, 29: 913, my emphasis)

The matter is passive because all of the changes in its states of motion and rest must be explained through external powers: nobody can begin to move or stop moving by itself, but always only in relation to another body that exercises a repulsive or attractive force on it. All change in the state of matter is thus attributable to an external power, which is why Kant characterises matter in this passage as 'mere receptivity'.

Now, Kant explicitly identifies the activity of passive beings, whose determinations, or changes of state, always depend on external powers, as 'mechanical'. And he contrasts this 'mechanical' activity with a second kind of 'animal or practical' activity, which determines a power from an inner principle:

> But there are still two kinds of activities: *one of these is mechanical*, and is produced by an external power; the other is animal or practical. Here the power is determined from an inner principle. (Met-L1/Pölitz, 28: 254; cf. Met-Volckmann, 28: 449)

Here, Kant contrasts the kind of activity distinctive of material nature with the kind of activity distinctive of animal nature, and he identifies only the former as *mechanical*, and therefore explicable in terms of the efficient-causal order of material nature. This efficient-causal order requires that every event be explained by a prior event, and therefore that the cause of any change be found in a further external power. Thus, as long as we explain events by citing external powers, we stay within the domain of natural scientific explanation, which explains why matter, as a purely passive

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

substance, is amenable to such explanation, for, as we have seen, every change of a material body is traceable to the exercise of an external power.

However, as soon as we attempt to explain an event by citing an inner rather than external principle, our explanations can no longer be accommodated within a mechanistic framework. Accordingly, Kant says explicitly that we can never appeal to the second kind of activity, spontaneity, to explain any event in nature:

> If I want to explain an event in the world, and I derive this from the general laws of nature, then that is a natural event. In the world as a series of appearances, we *cannot and must not* explain any event from spontaneity <*ex spontaneitate*>. (Met-Mron, 29: 862, my emphasis)

Since mechanistic explanation is by its nature restricted to passive causal powers, it follows directly that spontaneous activity is outside of the explanatory remit of the science of material nature.

Rather than surrendering spontaneous activity as rationally unintelligible, in the second part of the *Critique of the Power of Judgement*, Kant develops a distinct mode of causal explanation that is suitable for irreducibly spontaneous beings – teleology. This kind of explanation can make sense of a being that acts in accordance with an inner principle, for it is a mode of explanation that cites ends or purposes. When a substance acts spontaneously, there is no external cause that we can appeal to in order to explain the change of its state. Instead, we must appeal to something inside the substance that directs and produces the change, and teleology is a mode of explanation precisely designed to explain how the changes of a substance flow from its inner nature. The nature of a substance is the inner principle that we must appeal to in order to explain changes in its state; these principles ground the inherence of accidents in the substance (cf. Met-Mron, 29: 935). If the inner principle is by itself sufficient to ground an accident, that is, if we do not need to appeal to any external power to explain the change in its state, then the power that realises the accident is spontaneous. Now, in that case, the act that actualises the accident in the substance is an endeavour, or a striving. There is, thus, a distinction between a *principle*, which is part of the nature of a thing (and is a ground of possibility, not yet *actualised*), and the striving, which is the act or 'actualization' of the power. Hence, Kant says that '[b]etween faculty and power lies the concept of endeavor <conatus; G: Bestrebung>' (Met-L2/Pölitz, 28: 565), having noted that a faculty is the possibility of a power (cf. Met-Mron, 29: 823). Teleological explanation explains how the accidents of a substance result from these inner strivings and thus makes intelligible how a substance's states could be grounded by its inner principles.

If I am right that this form of explanation is appropriate to spontaneous activity, then, we should expect that Kant's explanation of the mind as a spontaneous power would have a teleological dimension. And indeed, Kant often characterises the mind and its various faculties in teleological terms that directly draw from the account of teleological explanation developed in the third *Critique*. There he claims, for example, that each faculty that forms part of the mind has, as part of its nature and so as a constitutive 'principle', an *interest* or a *goal* (5: 120; cf. also A642/B670).[29] We can think of this interest in broadly Aristotelian or Leibnizian terms as a striving to realise itself as the power that it is:

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

> To every faculty of the mind one can attribute an interest, that is a principle that contains the condition under which alone its exercise is promoted. (5: 120)

It should, therefore, not come as a surprise that Kant characterises the understanding, too, in explicitly teleological terms. He says that 'the understanding, ... is aimed at an end which is necessary for it' (5: 187, my emphasis), and that we cognise 'the lawful unity in a combination ... as in accordance with a necessary aim (a need) of the understanding' (5: 184, my emphasis). In both of these passages, Kant makes it clear that this is an aim or end that the understanding has essentially and not merely accidentally or voluntarily, and this is exactly what we should expect given that the understanding is spontaneous and given the connection we have seen between teleology and spontaneity. In this context, it is also worth emphasising that Kant characterises the mind and not only the organism as alive: 'the mind [*Gemüt*] for itself is entirely life (the principle of life itself)' (5: 278); thus, it requires no stretch of the imagination to suppose that Kant would apply the same general form of explanation in his theory of mind as we have seen characterises his account of animal life.

The task in the final part of the paper will be to describe the way in which Kant applies this teleological mode of explanation to the specific case of the faculty of understanding. Before we proceed to that task, however, I want to address a concern that one might raise about my appeal to teleology in this context.[30] Specifically, one might worry that Kant's approach to teleology in the third *Critique* does not treat it as an independent and 'alternative' mode of explanation, one that could in principle be pursued in the absence of mechanistic explanation, but as a form of explanation that is, at best, a heuristic device that can supplement our efforts to understand the world in mechanistic terms. Such explanations could never *replace* efficient-causal explanations of organisms, and in fact, Kant seems to think they will always be 'defective' from a scientific perspective.[31] Therefore, my suggestion that the fundamental explanation of the mind (and spontaneous beings in general) can and must be teleological might appear to assign a role to teleological explanation that Kant's doctrines in the third *Critique* preclude it from having.

My response to this line of objection is to suggest that the limitations that Kant places on teleological explanation in the third *Critique* take place in the specific context of a discussion of a proper explanation of biological organisms, that is, *material* organised beings. I do not need to deny any of the prevailing wisdom about Kant's views on the relationship between teleological and mechanical explanation in the domain of biology, the study of such beings. I do not need to claim that teleological explanations are *alternatives* to mechanical explanations of biological organisms, for example, and nor do I need to deny that mechanistic explanations must always, in principle, be available for biological organisms.[32] The important point to note is that these claims are made within a discussion of material organised beings, and the mind, while it is an *organised* being, is *not* a material being.[33] For this reason, the mind is not even in principle susceptible to mechanistic explanation, as the following two passages make clear:

> That thinking is thus a mechanism is absurd, this would mean to make thinking in one's own consciousness into an object of the outer senses. Matter

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

> can indeed be a necessary requirement for the support of our thoughts, but thinking itself is not mechanical. (Met-Volckmann, 28: 449)
>
> But we cognize matter, as object of outer intuition, in its extension and in its figure. Thinking takes no extension, no figure. - Therefore the soul must also be something other than matter. If this investigation has no positive use, it still has a negative one, which consists in this, that we do not fall into the mistakes of the materialists and explain the actions of the soul physico-mechanically . . . (Met-Mron, 29: 905)

Whereas the mind 'must . . . be something other than matter', and cannot, therefore, be explained 'physico-mechanically', biological organisms are material beings, and to the extent that anything is part of material nature, it must admit of explanation in mechanical terms. Teleology must have a place in the explanation of organised beings since they manifest spontaneity, but it must also be reconciled with mechanistic explanation insofar as such beings are part of material nature.[34] Therefore, in the context of our investigation of biological nature, we will never be able to dissociate teleology from mechanistic explanation and pursue the former in the absence of the latter. But since in the context of an account of the mind, a mechanistic explanation would be a materialist 'mistake', the choice is between teleology and no explanation at all, and the evidence, I have suggested, is that Kant chooses the first option. However, his willingness to pursue teleology absent mechanistic explanation in relation to the mind is entirely compatible with his insistence that in our study of material nature, teleology must never float free of mechanism.

The task now will be to describe the way in which Kant applies this teleological mode of explanation to the specific case of the faculty of understanding.

6. A teleological model

What we have seen up to now is a principled reason why the cognitive processing reader must be wrong in claiming that the understanding, as a spontaneous power, can be situated within the efficient-causal framework of material nature. In this section, I would like to sketch the outlines of an alternative teleological model of the understanding, which goes beyond existing cognitive processing proposals while managing to resist assimilation into the cognitive agency approach.

On the teleological model that I am proposing, the understanding, as a spontaneous power, must function in accordance with an inner principle, which is also a goal or end that reflects the nature of the faculty that the understanding is. To cash out the proposal, we therefore need to (1) identify the inner principle of the understanding, (2) explain how it could function as an end or goal of the understanding's activity, and (3) explain why we do not need to appeal to any agential notions in characterising the way in which this principle governs the understanding's activity.

Now, if we approach the first *Critique* with a background understanding of Kant's general metaphysics of spontaneous powers, what becomes striking is the frequency with which he refers to the *unity of apperception* as a 'principle' (emphasis added in the following) of the understanding. Consider:

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

> Now the unity of the manifold in a subject is synthetic; pure apperception therefore yields a *principle* of the synthetic unity of the manifold in all possible intuition. (A117; cf. A122)

> Combination does not lie in the objects ... but is rather only an operation of the understanding, which is itself nothing further than the faculty of combining a priori and bringing the manifold of representation to the unity of apperception, which *principle* is the supreme one in the whole of human cognition. (B135)

> The first pure cognition of the understanding, therefore, on which the whole of the rest of its use is grounded ... is the *principle* of the original synthetic unity of apperception. (B137)

In each of these passages, Kant claims that the unity of apperception is a principle that makes possible the synthetic unity or combination of the manifold of representation. In doing so, he claims that it 'grounds' our entire use of the understanding. The understanding, of course, is the faculty that we use to think of the objects of intuition (A50/B75), so Kant's claim appears to be that the unity of apperception grounds our thinking of objects. Now, we have seen in our discussion of the metaphysics of causal powers that acts, the expression of the inner principle of a substance, ground the inherence of accidents, and in doing so constitute the exercise of a distinctive power that the substance has. My claim here is that the unity of apperception is such an inner principle. Expressed through the act of apperception, this principle grounds the inherence in a subject of a distinctive kind of representation: thoughts.

Kant, of course, uses the term 'principle' in different ways in different contexts within his critical philosophy, but what I have tried to do in the previous section is to outline a way in which this term is used within his general account of causal powers. What we have seen, specifically, is that spontaneous powers act in accordance with an inner principle. Since the understanding is a spontaneous power of cognition, we should therefore anticipate that it too has an inner principle that is the ground of its activity. My suggestion, borne out, I think, by the passages just quoted, is that the unity of apperception is a good candidate for being such a principle. As I have shown above, spontaneous powers in general are those whose changes of state are grounded on an inner principle, where an inner principle is a goal for realising the power as the kind of power it is. And this is how I propose to understand the unity of apperception: as the goal or end of the understanding's activity, through the pursuit of which the understanding realises itself as the faculty that it is.[35]

What is the faculty of understanding? Most fundamentally, the understanding is the faculty for thinking of objects of sensible intuition (A51/B75). That is, understanding is the faculty in whose nature it is to make intelligible the representational contents received through the senses: making these representations intelligible (thinking objects rather than just intuiting them) just is understanding them. We therefore cannot make sense of the activity of the understanding without supposing that it has the goal of making representations intelligible. Given that this is the nature of understanding, we can see why the unity of apperception should function as an inner principle of the understanding's activity. For this, unity is just the

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

form of intelligibility itself – the very form that the activity of the understanding aims to bring about in the sensible manifold. Kant claims in several places that the unity of apperception is identical to the *form* of thinking (A346/B404; cf. A685/B713, B411, B427). It is the form that representations must have in order to be thought – the unity without relation to which a representation 'in me ... could not be thought at all' (B132). In order to realise itself as the faculty of thinking, then, the understanding must strive to bring '[the manifold of] representations to the unity of apperception' (B135). In doing so, the understanding operates in accordance with an end of intelligibility that is part of its nature.

A complete interpretation of Kant's theory of apperception is beyond the scope of this paper, but what I do want to emphasise is that this way of thinking of the nature of the understanding as a spontaneous faculty and of its fundamental act, apperception, does not and need not appeal to any agential notions. On the view of apperception that I have in mind, apperception is *not* consciousness or awareness of any activity, and nor is it an awareness (even 'implicit' awareness) of any rules. In my view, when the understanding confronts a given sensible manifold, it *automatically* responds in a distinctive way, and it outputs conceptual representation as an immediate result of this activity.[36] To make the sensible manifold intelligible is to impose a logical structure on the representations given in sensibility that allows us to think it, and apperception is the act through which the manifold acquires such a structure. Even though this activity is goal-directed (its aim is to bring given representations to the unity of apperception), neither the activity, nor the subject whose activity it is, is aware of itself or of this goal, or, for that matter, of the rules that would enable it to successfully reach that goal. Its functioning is sub-personal, automatic, and routine in very much the way that the cognitive processing view conceives of the workings of the understanding. Consequently, there is also no sense in which we could hold the subject accountable for producing or not producing any kind of representational output. Supposing that nothing interferes with its functioning, the understanding will simply unify representations in the way that it does because of its nature as the faculty of thinking.

To see more clearly what I have in mind, we can compare how the understanding functions with how even the most basic organisms function. Organisms, too, have an inner principle that is the ground of their behaviour. This inner principle accounts for the fact that the organism's behaviour is goal-directed: it behaves so as to preserve, sustain, and promote its life. Now, depending on the kind of organism it is, there will be a distinctive set of rules that would allow it to flourish, develop, preserve, and sustain itself as the kind of being it is, provided nothing interferes with its functioning. It can do so without ever being conscious of any of these rules, even implicitly; moreover, it can do so without ever representing its goal to itself or setting that goal for itself. To function in accordance with those rules is just part of what it is to be that organism. Now, my claim is that the understanding is, in this respect, like an organism. It, too, has a goal that is constitutive of its nature – the goal of making the sensible intelligible. It too must follow certain rules in order to achieve this goal. But just as in the case of the organism, it is not implicitly or explicitly conscious either of the rules themselves or of the goal for which it functions.

We can, therefore, give an account of how the understanding functions that is fully teleological but does not invite any agential notions. We do not hold organisms

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

accountable for their behaviour regardless of our appreciation that their behaviour is teleologically oriented, because we recognise that insofar as they simply behave in accordance with their nature, their behaviour is in no sense up to them. Similarly, we do not hold subjects accountable for the particular cognitive states that the understanding generates because we recognise that producing these states is in no sense up to them.

So, while I agree with the cognitive agency view on one count, namely that the transcendental subject and the functions of the understanding cannot be explained within the efficient-causal framework of material nature, I depart from this view in two fundamental ways.

First, the cognitive agency account holds that subjects are at least implicitly aware of the rules governing their cognitive activities. However, I argue that the purposive nature of the understanding does not imply that the subject is aware of its activity, or of the goal of that activity, or, indeed, of any rules that would facilitate reaching that goal. It simply generates representations without recognising its function as goal-directed. In this sense, it resembles an organism. A bee, for instance, instinctively follows behaviours without awareness of the laws governing them – its activity is as automatic and unreflective as that of a watch. While a biologist (at least Kant's biologist) may interpret the bee's behaviour teleologically, the bee itself has no conception of its actions as goal-oriented. Likewise, from the perspective of transcendental philosophy, understanding functions teleologically, yet neither the subject nor understanding itself is aware of its purpose or the rules that direct it.

Second, and connectedly, the goal of rendering the sensible intelligible is neither one that the understanding freely adopts nor one that it could renounce. It is entirely consistent with the teleological model that the understanding be necessitated to act in the way that it does; what matters is that the source of this necessity is an internal principle resident in the understanding itself. Moreover, this has implications for the status of the laws that Kant says govern the understanding: consistently with the teleological model, we can think of those laws as purely descriptive, rather than attributing to them the normative status that, we saw above, is central to the cognitive agency reading.[37] If one wishes to insist on calling them norms, that is fine too; however, we must recognise that a subject's acting in accordance with them would not be the basis for holding her *accountable* for her cognitive activity in any meaningful sense.

7. Conclusion

Throughout this paper, I have argued that there is an alternative to existing interpretations of spontaneity. We neither need to accept that Kant's doctrine of spontaneity requires that the cognitive subject be *held accountable* for the operations of the understanding in any interesting sense nor must we accept an account that regards the understanding as mere mechanism. We do not have to choose between the cognitive agency model, which imports a problematic conception of the nature of judgement and the laws of understanding, and the cognitive processing approach, which brings with it a mechanistic conception of the understanding that seriously distorts Kant's view. Departing from both of these models, I have shown that there are resources within Kant's philosophy to articulate an alternative teleological account of

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

spontaneity, which fits Kant's specific account of the spontaneity of understanding into his broader metaphysics of spontaneous powers in general. The result is, I hope, a philosophically intriguing view of the intellect that will be the basis for further productive discussions of transcendental psychology and the place of transcendental self-consciousness in Kant's account of the faculty of understanding.

Acknowledgements. I am grateful to the audiences at Aarhus Kant Week (2019), the UKKS Annual Conference in Bristol (2019), the NAKS Pacific Study Group at Loyola Marymount University (2020), the ENAKS Virtual Dream Commentator series (2021), and the Kant and Cognitive Science Workshop hosted by the Digitales Kant Zentrum NRW and Universität Siegen in Imperia, Italy (2023). For their constructive criticism and thoughtful engagement, I am especially thankful to Hannah Ginsborg and Béatrice Longuenesse, who commented on an early draft of this work through the NAKS Virtual Dream Commentator series. I am also indebted to the members of various workgroups and participants in informal discussions who offered valuable feedback on earlier drafts, including Lucy Allais, August Buholzer, Tim Jankowiak, Patricia Kitcher, David Landy, Michael Oberst, Jim O'Shea, Karl Schafer, Andrew Stephenson, Brian Tracz, Clinton Tolley, and Eric Watkins. I am grateful as well to two anonymous reviewers for their helpful and incisive comments. Finally, I am deeply grateful to Max Edwards, whose advice and feedback guided this project at every stage, and who read more drafts of this paper than anyone reasonably should.

Notes

1 Kant distinguishes between two senses of the term 'understanding': a broad and a narrow sense (A130–31/B169). In its narrow sense, the term refers to the specific cognitive faculty that Kant analyses in the Transcendental Analytic. The broad sense encompasses all the higher faculties of cognition, including the understanding in the narrow sense of the term, the power of judgement, and reason. While Kant takes the understanding in both senses to be spontaneous, in this paper, I will be exclusively interested in the claim that the understanding in the narrow sense is spontaneous.

I cite the first *Critique* using the customary A/B pagination. All other texts are cited using the Akademie Ausgabe (Kant 1900) pagination. All translations are from the Cambridge Edition of Kant's works (1992) unless otherwise indicated.

2 I use the term 'cognitive spontaneity' to mark Kant's use of the notion of spontaneity in his theoretical philosophy, as distinct from his use of the notion in his practical philosophy. Hereafter, I will use the term 'spontaneity' to refer specifically to cognitive spontaneity, and, more specifically still, to the spontaneity of the understanding (in the narrow sense of the term).

3 Aside from the authors mentioned, this kind of view is also defended by Allison (1990, 1996), Ellis (2017), Kohl (2015, 2023), Land (2021), Longuenesse (2017), McLear (2020), Pollok (2017), Pippin (1987).

4 For further discussion of Sellars' work on relative spontaneity, see Brink (forthcominga).

5 See especially Allison (1996: 59); Aquila (1989: 31), Hurley (1994: 160–2), Pippin (1987).

6 Accounts that view the processes of the understanding primarily in sub-personal, non-agential terms are present in Neiman (1994: 48ff.), Dickerson (2003), and Friedman (1996).

7 Pippin especially puts this objection at the centre of his critique (Pippin 1987: 451).

8 The claim that the understanding functions teleologically is not unprecedented. Béatrice Longuenesse (1998) develops an account on which we cannot make sense of the understanding's activities of synthesis without positing judgement as its goal. Karl Schafer (2023) has recently argued that generalising Longuenesse's account of the teleological unity of the understanding enables us to see the faculties of the mind as forming a teleological system, but Schafer departs from Longuenesse in drawing the unifying principle of this system from the aims of the highest cognitive faculty, namely, reason, rather than the lower intellectual faculties. Courtney Fugate (2014) also develops an account of the teleology of the intellect as a whole and of understanding as a component thereof; but to the best of my knowledge, no one has attempted to explain what Kant's doctrine of cognitive spontaneity amounts to in teleological *but non-agential* terms, as I do in this paper.

9 To head off a potential misunderstanding of the view that I will advocate, my claim is only that end-directedness and imputability are conceptually *independent*, not that they are *mutually exclusive*. But the

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

fact that these two notions can be separated means that it is insufficient to motivate the cognitive agency view by showing that the understanding is not part of the efficient-causal framework of material nature.

10 The literature I am referring to here specifically concerns the nature of the understanding as a cognitive faculty. There is an adjacent debate around the ethics of belief in Kant, which focuses more narrowly on Kant's discussion in the Canon of the various epistemic attitudes (of 'holding-for-true' or *Fürwahrhalten*) that a cognitive subject can take up. There is disagreement about whether or not Kant's view is best characterised as involuntarist or voluntarist when it comes to these specific attitudes; cf. esp. Chignell (2007) and Cohen (2013) for representative views. As we will see below, it is not clear that this discussion has any direct bearing on characterising the sense in which the understanding and its activities are spontaneous, and so I will not engage directly with this rich literature in what follows.

11 What I say in what follows applies regardless of how we spell out the details of what accountability amounts to in this context.

12 Longuenesse argues that these attitudes can have different degrees of epistemic force depending on 'the relation between the particular proposition . . . and other propositions that belong to the system of knowledge available to me to back up my statement, whether or not I am, currently, consciously assessing the truth of those propositions' (Longuenesse 2017: 27). We might also think, with Boyle (2015: 24) and Allison (1990: 37), that the evaluative component is not a distinct act that takes place in isolation from the rule-governed synthesis that generates the content that is under consideration.

13 This notion of 'implicit consciousness' shows up in the literature on cognitive agency quite often. Land (2021), Boyle (2015, 2024), and Longuenesse (2017) all appeal to such a notion in discussing the sense in which we can be responsive to the norms that govern our acts of thought and judgement. They all insist that we need not be *explicitly* conscious of the grounds that we have for judging as we do, but that we must nevertheless recognize that these acts are constrained by norms in some implicit sense. Boyle's recent work (2024) makes a serious effort to articulate this conception of implicit consciousness in terms of the Sartrean notion of 'nonpositional consciousness', which he elucidates by appeal to the idea of a mode of presentation. For critical discussion, see Brink (forthcomingb).

14 For particularly influential treatment, see Chignell (2007) Smit (2009), and Watkins and Willascheck (2017).

15 Chignell (2007); Leech (2012: 261); Tolley (2019); Watkins and Willaschek (2017).

16 For constitutivist readings, see especially Merritt (2018), Nunez (2018), Tolley (2006). For primary texts that clearly invite a constitutivist reading, see especially, CPJ, 5: 186–7, Logic, 9: 11.

17 Kitcher herself moves away from the kind of overtly functionalist, cognitive processing framework that she develops in her first monograph in her later work. As we will see, Neiman (1994) appears to accept this kind of view, and while Hanna (2020) clearly articulates the position, he does not ultimately defend it. Consequently, there have been no recent attempts to work out this view in systematic detail.

18 For cognitive agency readers who criticize precisely this aspect of Kitcher's position, see Allison (1996) and Pippin (1987).

19 Several commentators take this framework seriously in their exploration of various topics in Kant's Critical philosophy. See, especially, Ameriks (2000) and Watkins (2005). More recently, Boyle (2020), Smit (2009), Wuerth (2014), Dyck (2014), Tolley (2021), McLear (2020).

20 For an explicit repudiation of such attempts, see Kitcher (2017).

21 For readings along the same lines, see Ameriks (2000: 67), Wuerth (2014: 166–7), Dyck (2014: 200ff.).

22 I am very grateful to an anonymous referee for encouraging me to think harder about how my account can be made compatible with Kant's restriction of the categories, and the paragraphs to follow constitute my attempt to do justice to this important concern.

23 See, for example, Ameriks (2000, 2003); Chignell (2010); Hogan (2009); Watkins (2002), Van Cleve (1999).

24 See, especially, Chance (2018), Edwards (2023), Watkins (2002).

25 Following Chignell (2007), I translate 'Glaube' as 'belief'; however, 'faith' may arguably be a more appropriate rendering, as it better reflects the concept's narrower scope compared to contemporary understandings of belief. As Chignell observes, neither translation is entirely satisfactory: for modern readers, 'faith' is often too closely associated with religious contexts, whereas 'belief' fails to convey what is distinctive about Kant's use of 'Glaube'.

26 At the heart of the issues that I raise here is a deeper question about Kant's critical methodology, and the status and justification of the starting assumptions of the study that he undertakes in the *Critique of*

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

Pure Reason, but that is beyond the scope of my concern in this paper, which is simply to show that there is room, despite Kant's restrictions on theoretical cognition, to think that the faculty psychology and theory of causality that he inherits from his rationalist predecessors are consistent with his critical project.

27 Eric Watkins' landmark study (2005) brings out the crucial role of the metaphysics of causal powers in Kant's theoretical philosophy.

28 See, for example, Met-L1/Pölitz, 28: 268, 285; Met-Volckmann, 28: 448.

29 I am using the terms 'goal', 'aim', 'end', and 'interest' as at least extensionally equivalent for purposes of my discussion here. I am not committed to denying that we could or should not make fine-grained distinctions between these concepts, or that we could not track such distinctions in Kant's use of them. What I am committed to is that these are all teleological concepts and that they all show up in Kant's discussion of how the mind functions, and in connection to the faculty of understanding in particular. I should note, however, that there is a distinction between an 'inner principle' and a goal. An inner principle, in this context, is just a part of the nature of a thing that grounds changes in its state (which we distinguished above from the 'striving' through which these changes take place). Some inner principles ground the inherence of accidents teleologically, and they are, therefore, constitutive *goals* or *ends* – that is, goals or ends that are part of the nature of the being. But not all principles function in this manner – notably, those that require the cooperation of external powers do not. I would like to thank an anonymous reviewer for encouraging me to be more explicit about the way that I am using terminology here and throughout this section of the paper.

30 I am very grateful to an anonymous referee for impressing this objection on me.

31 There is a tendency in the literature on the third *Critique* to downplay the explanatory significance of teleology. On this sort of view, typified by readers such as McLaughlin (1990, 2014) and Geiger (2021), Kant thinks of teleological explanation as a useful heuristic device but denies that any part of reality is *in principle* mechanically inexplicable, contingent human cognitive limitations notwithstanding. This kind of view finds support in statements like the following from the first introduction to the third *Critique*:

> We can and should be concerned to investigate nature, so far as lies within our capacity, in experience, in its causal connection in accordance with merely mechanical laws: for in these lie the true physical grounds of explanation, the interconnection of which constitutes scientific cognition of nature through reason. (CPJ, First Introduction 20: 236)

32 Such explanations will, of course, as Kant insists, also be compatible with any teleological understanding that we have of these organisms (as the resolution to the Antinomy requires), and that too is not a problem for my account here.

33 For excellent discussions of Kant's immaterialism, see Ameriks (2000) and Watkins and Willaschek (2017).

34 For discussion of the complex issues regarding the exact nature of this reconciliation, see, for example, Breitenbach (2008), Geiger (2009), Goy (2015), McLaughlin (1990), and Watkins (2009).

35 I should emphasise that I do not take this reading of the doctrine of apperception simply to fall out of the fact that Kant refers to the unity of apperception as a 'principle'. An anonymous reviewer suggests that Kant at least sometimes uses that term in a non-teleological register, and if that is so, then, the status of apperception as a 'principle' does not immediately entail the teleological reading I propose. I am happy to acknowledge that Kant's use of the term 'principle' is not univocal. Nevertheless, my background reading of spontaneous powers in general, with the linkage it establishes between teleology and spontaneity, entails that the understanding, as a spontaneous power, must have an inner principle that teleologically governs its changes of state. In my reading, therefore, when Kant presents us with the inner principle of the understanding *qua* spontaneous power, he must be using the term 'principle' in a teleological register, regardless of the further uses to which he puts the term in other contexts.

36 In doing so, the subject is not committed to any way the world is. I would go so far as to say that the transcendental subject is never committed to any claims about what the world is like, what is true or what is false about it.

37 The claim that we are accountable for our acts of cognition, as I understand it, is meant to be grounded in the claim that the laws and rules of the understanding function as norms for us. Thus, on this view, these cognitive acts must be in our control (in some or other sense), and that must mean, at least,

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

that it is possible that our acts of cognition do *not* necessarily conform to the rules of the understanding. It seems to me that a rule can function as a norm for engaging in a certain kind of activity only if it is possible to perform the activity *without* conforming to the norm – at least if norm-responsiveness is meant to ground accountability for the kinds of act in question. So, if there is a rule that, say, beliefs be formed on the basis of evidence, then for it to be appropriate to hold me accountable for the beliefs I hold, it must be possible for me to form beliefs that violate that rule. Were it not possible to violate the rule, such a rule would describe what it is to act in a relevant way, rather than prescribing a certain way of acting, and thus could not function as a norm, at least in the sense that I am understanding that notion for the purposes of this paper. Now, what I am suggesting is that such non-conformity is not possible in the case of the rules and laws of the understanding (I am in broad agreement with Tolley (2006) here). It simply is not up to us whether, how, and when we apply the categories to what is given in sensibility. My target in this context is, admittedly, a very specific interpretation of the Analytic, according to which norm-responsiveness grounds a kind of accountability that we have for our acts of cognition. I do not mean to suggest that my account precludes *all* kinds of normativity; Ginsborg's (2012, 2014) conception of primitive normativity, for example, might well be relevant to our characterization of the content of our cognitive and mental states, but I do not think it obviously has implications for holding a subject accountable for those states; so, I leave open here the question of whether my view is compatible with proposals such as Ginsborg's.

References

Allison, Henry E. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

—— (1996) *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. New York: Cambridge University Press.

Ameriks, Karl (2000) *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*. Oxford: Oxford University Press.

—— (2003) *Interpreting Kant's Critiques*. New York: Oxford University Press.

Aquila, Richard (1989) *Matter in Mind: A Study of Kant's Transcendental Deduction*. Bloomington, IN: Indiana University Press.

Boyle, Matthew (2015) 'Die Spontaneitaet des Verstsandes bei Kant und einigen Neokantianern'. *Deutsche Zeitschrift für Philosophie*, 63, 35–55.

—— (2020) 'Kant on logic and the laws of the understanding'. In Sofia Miguens (ed.), *The Logical Alien* (Cambridge, MA: Harvard University Press), 117–44.

—— (2024) *Transparency and Reflection: A Study of Self-Knowledge and the Nature of Mind*. New York, NY: Oxford University Press.

Breitenbach, Angela (2008) 'Two views on nature: a solution to Kant's antinomy of mechanism and teleology'. *British Journal for the History of Philosophy*, 16, 351–69.

Brink, Claudi (forthcominga) 'Kant's distinction between absolute and relative spontaneity reconsidered'. *British Journal for the History of Philosophy*. Published online: 08 May 2025. https://doi.org/10.1080/09608788.2025.2478979.

—— (forthcomingb) 'Nonpositional consciousness and the form of thought: Kant and Boyle on self-consciousness'. *Canadian Journal of Philosophy*.

Chance, Brian A. (2018) 'Pure understanding, the categories, and Kant's critique of Wolff'. In Kate A. Moran (ed.), *Kant on Freedom and Spontaneity* (Cambridge and New York: Cambridge University Press), 30–47.

Chignell, Andrew (2007) 'Belief in Kant'. *The Philosophical Review*, 116, 323–60.

—— (2010) 'Real repugnance and belief about things-in-themselves: a problem and Kant's three solutions (including one about symbols)'. In Benjamin J. Bruxvoort Lipscomb and James Krueger (eds.), *Kant's Moral Metaphysics: God, Freedom, and Immortality* (Berlin: De Gruyter), 177–209.

Cohen, Alix (2013) 'Kant on doxastic voluntarism and its implications for epistemic responsibility'. *Kant Yearbook*, 5, 33–50.

Dickerson, Adam B. (2003) *Kant on Representation and Objectivity*. Cambridge: Cambridge University Press.

Dyck, Corey W. (2014) *Kant and Rational Psychology*. New York, NY: Oxford University Press.

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

Edwards, Maximilian (2023) 'Synthetic attributes and the schematized categories'. *Kantian Review*, 28, 21–40.
Ellis, Addison (2017) 'The case for absolute spontaneity in Kant's *Critique of Pure Reason*'. *Con-Textos Kantianos*, 6, 138–64.
Friedman, Michael (1996) 'Exorcising the philosophical tradition: comments on John McDowell's *Mind and World*'. *The Philosophical Review*, 105, 427–67.
Fugate, Courney (2014) *The Teleology of Reason: A Study of the Structure of Kant's Critical Philosophy*. Berlin: De Gruyter.
Geiger, Ido (2009) 'Is teleological judgement (still) necessary? Kant's arguments in the Analytic and in the Dialectic of teleological judgement'. *British Journal for the History of Philosophy*, 17, 533–66.
—— (2021) *Kant and the Claims of the Empirical World: A Transcendental Reading of the Critique of the Power of Judgment*. New York: Cambridge University Press.
Ginsborg, Hannah (2012) 'Meaning, understanding and normativity'. *Aristotelian Society Supplementary Volume*, 86, 127–46.
—— (2014) *The Normativity of Nature: Essays on Kant's Critique of Judgment*. Oxford: Oxford University Press.
Goy, Ina (2015) 'The antinomy of teleological judgment'. *Studi Kantiani*, 28, 65–88.
Hanna, Robert. (2020) 'Kant's theory of judgement'. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford CA: Stanford University.
Hogan, Desmond (2009) 'How to know unknowable things in themselves'. *Noûs*, 43, 49–63.
Hurley, Susan L. (1994) 'Kant on spontaneity and the myth of the giving'. *Proceedings of the Aristotelian Society*, 94(1), 137–64.
Kant, Immanuel (1900–) *Gesammelte Schriften*. 29 vols. Ed. German Academy of Sciences (and predecessors). Berlin: De Gruyter (and predecessors).
—— (1992–) *The Cambridge Edition of the Works of Immanuel Kant*. Ed. P. Guyer and A. Wood. Cambridge: Cambridge University Press.
Kitcher, Patricia (1990) *Kant's Transcendental Psychology*. New York: Oxford University Press.
—— (2017) 'Kant on the faculty of apperception'. *British Journal for the History of Philosophy*, 25, 589–616.
Kohl, Markus (2015) 'Kant on freedom of empirical thought'. *Journal of the History of Philosophy*, 53, 301–26.
—— (2023) *Kant on Freedom and Rational Agency*. Oxford: Oxford University Press.
Land, Thomas (2021) 'Epistemic agency and the self-knowledge of reason: on the contemporary relevance of Kant's method of faculty analysis'. *Synthese*, 198, 3137–54.
Leech, Jessica (2012) 'Kant's modalities of judgement'. *European Journal of Philosophy*, 20, 260–84.
Longuenesse, Béatrice (1998) *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Trans. C. T. Wolfe. Princeton, NJ: Princeton University Press.
—— (2017) *I, Me, Mine: Back to Kant and Back Again*. Oxford: Oxford University Press.
McDowell, John (1994) *Mind and World*. Cambridge, MA: Harvard University Press.
McLaughlin, Peter (1990) *Kant's Critique of Teleology in Biological Explanation: Antinomy and Teleology*. Lewiston, ME: E. Mellen Press.
—— (2014) 'Mechanical explanation in the "Critique of the Teleological Power of Judgment"'. In Eric Watkins and Ina Goy (eds.), *Kant's Theory of Biology* (Boston: De Gruyter), 149–66.
McLear, Colin (2020) 'On the transcendental freedom of the intellect'. *Ergo. An Open Access Journal of Philosophy*, 7, 35–104.
Merritt, Melissa (2018) *Kant on Reflection and Virtue*. Cambridge: Cambridge University Press.
Neiman, Susan (1994) *The Unity of Reason*. Oxford: Oxford University Press.
Nunez, Tyke (2018) 'Logical mistakes, logical aliens, and the laws of Kant's pure general logic'. *Mind*, 128, 1149–80.
Pippin, Robert B. (1987) 'Kant on the spontaneity of mind'. *Canadian Journal of Philosophy*, 17, 449–75.
Pollok, Konstantin (2017) *Kant's Theory of Normativity: Exploring the Space of Reason*. New York: Cambridge University Press.
Schafer, Karl (2023) *Kant's Reason: The Unity of Reason and the Limits of Comprehension in Kant*. Oxford: Oxford University Press.
Sellars, Wilfrid (1970) "' . . . this I or He or It (the thing) which thinks . . . "'. *Proceedings and Addresses of the American Philosophical Association*, 44, 5–31.

https://doi.org/10.1017/S1369415425100782 Published online by Cambridge University Press

Smit, Houston (2009) 'Kant on apriority and the spontaneity of cognition'. In S. Newlands and L. M. Jorgensen (eds.), *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams* (Oxford: Oxford University Press), 188–251.
Tolley, Clinton (2006) 'Kant on the nature of logical laws'. *Philosophical Topics*, 34, 371–407.
—— (2019) 'Kant on the place of cognition in the progression of our representations'. *Synthese*, 197, 3215–44.
—— (2021) 'The metaphysics of powers in Kant and Hegel'. In Julia Jorati (ed.), *Powers: A History* (New York: Oxford University Press), 243–70.
Van Cleve, James (1999) *Problems from Kant*. New York: Oxford University Press.
Watkins, Eric (2002) 'Kant's transcendental idealism and the categories'. *History of Philosophy Quarterly*, 9, 191–215.
—— (2005) *Kant and the Metaphysics of Causality*. Cambridge: Cambridge University Press.
—— (2009) 'The antinomy of teleological judgment'. *Kant Yearbook*, 1, 197–222.
Watkins, Eric, and Marcus Willaschek (2017) 'Kant's account of cognition'. *Journal of the History of Philosophy*, 55, 83–112.
Wuerth, Julian (2014) *Kant on Mind, Action, and Ethics*. Oxford: Oxford University Press.

Cite this article: Brink, C. Cognitive Spontaneity and the Organisation of the Understanding. *Kantian Review*. https://doi.org/10.1017/S1369415425100782