Reason, Freedom and Kant: An Exchange[1]

ROBERT HANNA AND A.W. MOORE
University of Colorado and St Hugh's College, Oxford

1. Robert Hanna

> Experience reveals only the law of appearances and consequently the mechanism of nature, the direct opposite of freedom. (*CPrR* 5: 29)[2]

According to Kant, being purely rational or purely reasonable and being autonomously free are one and the same thing. But how can this be so? How can my innate capacity for pure reason ever motivate me to do anything, whether the right thing or the wrong thing? What I will suggest is that the fundamental connection between reason and freedom, both for Kant and in reality, is precisely *our human biological life and spontaneity of the will,* a conjunctive intrinsic structural property of our animal bodies, which essentially constitutes human personhood and rational agency. I say 'suggest' because, obviously, no proper argument for such a conclusion could ever be worked out in a short essay. I would nevertheless like to motivate my suggestion by way of a commentary on the second part of Adrian Moore's extremely rich and interesting recent book, *Noble in Reason, Infinite in Faculty*[3] (henceforth, *NIR*).

According to what Moore calls the Radical *Conception,* nobody ever freely does the wrong thing. To do the wrong thing is to be unfree and to do the irrational thing. Or in other words, to act freely is to act rightly and to act rationally (*NIR:* 94–97). And according to what Moore calls the Radical *Picture,* not only is the Radical Conception true, but also we can incur blame for things we have not done freely, hence irrationally (*NIR:* 115–119). Strictly speaking, Kant does not defend either the unqualified Radical Conception or the unqualified Radical Picture. But there are pressures within his Critical philosophy to do so, and he comes about

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

as close to defending them as any philosopher ever has. In the second part of Moore's book, which consists in a discussion of Kant's notion of freedom and some philosophical variations on it, Moore wants to explore both the Conception and the Picture in relation to Kant's moral philosophy, with an eye to understanding Kant better and also if possible finding out some necessary truths about rational human nature.

Moore's central proposal is what he calls the Basic Idea, which is that there is a conation or nisus in all of us, *more fundamental than any other,* towards rationality (*NIR:* 128), and that this innate drive towards rationality is also experienced as what Kant calls the feeling of 'respect' for persons as ends-in-themselves and for the moral law (*NIR:* 134, 136). So being free in any way is expressing our drive to rationality, and being autonomously free is authentically and fully expressing that drive along with other free persons under moral laws, hence authentically and fully becoming ourselves as individual moral animals in an ideal community of moral animals. Kantian ethics is thus a version of what Moore calls 'conative objectivism' (*NIR:* 7), the theory that ethical thinking depends on the contents of certain psychological drive-states that we all innately share. Otherwise put, pure reason has motivational force and action-guiding content because it is essentially connected with its own unique kind of desire – the unique kind of desire that makes us *us.*

I am deeply sympathetic to Moore's proposal. In the end, what distinguishes our views, and where we will perhaps disagree, is only this: I think that the innate drive towards rationality is the same as the conjunction of our human biological life and spontaneity of the will, and that the human will is necessarily embodied. Combine that with a non-reductive view of biological concepts and facts, and call the resultant view *embodied libertarian rationalism.* In my opinion this view is the only way to get a satisfactory Kantian metaphysics of freedom of the will. Such a metaphysics would enable us to get beyond both (i) the classic Kantian dilemma, as found in the third Antinomy, between Newtonian or LaPlacean determinism on the one hand, and libertarian indeterminism on the other, and also (ii) the more recent but even more robust post-Kantian dilemma – to use Wilfred Sellars's evocative formulation[4] – between the Scientific Image and the Manifest Image of human beings in the world: that is, between Newtonian/

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

LaPlacean determinism or stochastic indeterminism on the one hand (let us call this disjunctive view, *Natural Mechanism*), and our irreducible phenomenology of free human agency on the other (let us call this *Phenomenal Libertarianism*).

In any case, in order to unpack Moore's account, I will need to say something briefly about how Moore construes Kant's conception of rationality, about the Radical Picture, and finally about the Basic Idea.

Rationality. Moore sees, as I do, a deep affinity between Kant and Wittgenstein. Rationality for Kant and for Moore is, literally, *making sense,* or constructing meanings, and this is possible only by way of an innate capacity for generating, deploying, and possessing *concepts* (*NIR:* 78–87). Some concepts are inherently action-guiding or normative, and to possess one of these concepts is to *live* by it. One central example of this is our innate capacity for doing mathematics. The other central example is our innate capacity for doing ethics.[5] One Wittgensteinian dimension of this conception of rationality as making sense or constructing meanings, is that making sense is in turn possible only in a social context and against a backdrop of shared practices. This necessary linkage of rationality and sociability is of course not at all foreign to Kant, who speaks of the public use of reason (*CPR* A738–769/ B766–797) and the necessary communicability of judgments, and the necessity of certain types of shared feelings in aesthetics and morality alike (*CPJ* 5: 203–244).[6] A robust conception of rationality naturally leads to philosophical rationalism. The thesis of rationalism, according to Moore, is the thesis that the human ability to reason, or make sense, comprehends both theoretical reason (best exemplified by mathematics) and practical reason (best exemplified by ethical thinking). Unlike classical rationalism however, which requires both God and also some sort of platonic objects or mind-independent non-spatiotemporal essences in order to explain this ability to reason, Kantian rationalism holds that we need only posit the existence of persons, or rational human animals, over and above the existence of the many different sorts of material things that populate the empirical, observable, or macroscopic natural world.

But how can a person's pure reason be ethical? Ethical rationalism would seem to imply, implausibly, that ethics is a kind of formal science; that moral principles, because strictly universal, are

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

utterly insensitive to context and cultural difference; and that, as formal and pure, reason has no action-motivating or action-guiding force. One way of responding to these worries is provided by the Radical Picture.

The Radical Picture. Recall that in Moore's language, the Radical Conception is that being free, being rational, and doing the morally right thing are one and the same, from which it follows that no one ever freely does evil, and that all evil is irrational. The Radical Picture then adds the further idea that we must be held morally responsible for at least some things that we have not done freely. Thus the Radical Picture provides a response to the basic worries about ethical rationalism by essentially *identifying* reason, right action, and freedom of the will. What could be less formal, less disengaged, and more action-oriented than freedom?

This brings us to Kant's conception of freedom of the will. Moore focuses his account on two central aspects of Kant's theory: (1) his distinction between *Willkür* and *Wille,* and (2) his metaphysics of freedom.

Now what, more precisely, *is* the human will, according to Kant? The answer is that *Willkür,* or the power of choice, is the power of intentional causation, that is, effective desire; by contrast *Wille,* or the will, is the power of self-legislation, or giving ourselves either instrumental or non-instrumental reasons for the determination of choice (*MM* 6: 213–214). To act on the basis of *Willkür* is always to move our animal bodies on the basis of our desires. This can of course occur in a Humean way by means of instrumental reasoning according to the hypothetical imperative. Since instrumental reasoning is itself a form of self-legislation, it involves what we might call the 'impure' *Wille.* To act on the basis of the pure *Wille,* however, is to constrain and differently determine our *Willkür* by recognizing the categorical imperative, which, as recognized, provides a universal overriding non-instrumental reason for action. So to act on the basis of pure *Wille* is to do the right thing as determined by our own pure practical reason, no matter what the external and psychological antecedents, and no matter what the consequences. This two-levelled conception of the human will, in turn, allows us to understand the Radical Picture. The nub of this understanding, as Moore expresses it, is this:

> Both irrational acts and rational acts qualify as exercises of freedom, but whereas the former qualify simply through the agent's choice to act

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

> in one way rather than another, the latter qualify in another way too, namely through the agent's compliance with his own or her own most fully autonomous judgment about how that choice is to be made. (*NIR:* 119)

Here is the crucial point. Even when we are acting unfreely, immorally, and irrationally, it remains true that our capacity for acting freely, non-instrumentally, and rationally in the sense of pure *Wille* is undiminished, despite the fact that we have not adequately realized that capacity in that context. *Only a being with an undiminished capacity for pure practical reason can act unfreely, immorally, and irrationally.* Hence we remain morally responsible even for things that we have done unfreely and irrationally in the sense of pure *Wille,* provided that we have also done them freely, instrumentally, and rationally in the sense of *Willkür.* This is because the capacity for pure *Wille* counterfactually guarantees that even if, given the same set of external and psychological antecedents, together with the fact that it had been in our selfish or even benevolent interest to do something morally wrong, then we still could have gone ahead and done the right thing instead of the wrong thing we actually did.

But this explanation of the Radical Picture will not ultimately work without a metaphysics of freedom of the will, according to which the capacity for pure *Wille* has real causal efficacy. What would Kant say about this? Moore construes Kant as an incompatibilistic compatibilist. Kant's view of freedom is incompatibilistic because he thinks that Newtonian/LaPlacean determinism and libertarian indeterminism are mutually metaphysically inconsistent. But Kant is also a compatibilist who thinks that if we are transcendental idealists and thereby adopt distinct phenomenal and noumenal standpoints on the will, then despite the fact that we take ourselves, from the phenomenal standpoint, to be acting at best comparatively freely or unfreely, instrumentally, and impurely rationally in the sense of *Willkür,* nevertheless from the noumenal standpoint we can also take ourselves to be acting under the regulative idea of autonomy or pure rational freedom, that is, take ourselves to be acting absolutely freely, non-instrumentally, and purely rationally in the sense of pure *Wille.*

As Moore correctly notes, Kant's incompatibilistic compatibilism, even when construed according to the two standpoint theory of the phenomenon-noumenon distinction, as opposed to

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

the two world theory, is a metaphysical mystery, bordering on complete unintelligibility (*NIR:* 104–113). Timeless, indeterministic, natural-law violating libertarian agency in a spatiotemporal, deterministic, nomologically-governed physical world is a non-starter, even as a regulative idea. And that is because the existence of a deterministic physical cause both explanatorily and metaphysically excludes the timeless cause, and timeless causal overdetermination seems absurd. How then can we make sense of the Radical Picture in terms of freedom? In order to do this, Moore runs a variation on Kant and proposes the Basic Idea.

The Basic Idea. On Moore's Kantian approach to reason and freedom, to be free is to be rational, and to be rational is to make sense. But what, apart from an ability for noumenal causation or transcendental freedom, could adequately align and relate pure reason and freedom? The first part of Moore's proposed answer is that rational freedom is making *new* sense, or rational creativity (*NIR:* 65–66, 71–78, 121–122). This is the same as creating radically new concepts, and then living by them. Moore connects this idea again to Wittgenstein, but this time to the early Wittgenstein of the *Tractatus.* To create and live by a radically new concept is 'to exercise one's will in such a way that the world "becomes an altogether different world. It must, so to speak, wax and wane as a whole. – The world of the happy man is a different one from that of the unhappy man"' (*NIR:* 125).

In this way, unfreedom and irrationality are ways of wilfully refusing to make new sense, or ways of wilfully refusing to be rationally creative. And because they are wilful, we are personally responsible for this refusal.

The Basic Idea then adds this thesis: we posses an innate nisus or drive, *more fundamental than any other,* towards rationality (*NIR:* 128). Freedom and rationality are thus the full expression and realization of this most fundamental creative drive, whereas unfreedom and irrationality are the self-suppression and wilful non-realization of this creative drive. So our most fundamental drive is to realize ourselves as autonomous creative rational animals in Kant's sense. As rational animals, we are all fundamentally trying to become authentic persons in an ideal community of other persons, and to create meaning in our lives by progressively conforming ourselves to the categorical imperative. And to refuse to try to be as rational as possible in this sense is to be inauthentic, and to refuse to be true

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

to ourselves. In this way, Moore's Basic Idea beautifully interweaves threads of existentialism and Wittgenstein's philosophy with the vital cord of the Critical philosophy.

But here is a worry about the Basic Idea. If I have understood him correctly, Moore himself is a conceptual or explanatory incompatibilist, because he holds what he calls the Incommensurability Thesis, which is that '*exercise of the concept of* physical determinism precludes *exercise of the concept of* freedom' (*NIR:* 114, Moore's emphasis). But conceptual or explanatory incompatibilism is logically consistent with metaphysical or ontological compatibilism (*NIR:* 120), just as conceptual or explanatory non-reductionism in the philosophy of mind is logically consistent with metaphysical or ontological reductionism.

So as it stands it seems to me that the Basic Idea is logically consistent with Natural Mechanism. We could be at once naturally mechanized *and also* such that we possess an innate drive, more fundamental than any other, towards rationality. But if so, then we are at best only *phenomenal* libertarian rationalists. And then it is all really a tragic illusion because we do not literally act freely and literally move our own limbs, either by means of *Willkür* and impure practical reason, or by means of pure *Wille* and pure practical reason. In fact, we are nothing but naturally mechanized puppets epiphenomenally dreaming that we are persons. But if that is true, as Jerry Fodor observes in a closely related context, then practically everything we believe about anything is false, *and it's the end of the world.*[7]

So what I would propose instead, is an interpretation of Kant's theory of freedom of the will and of Moore's Basic Idea which takes libertarian rationalism and conative objectivism to entail the denial of both incompatibilism and compatibilism, that is, to be *neither* incompatibilist *nor* compatibilist.

Consider first compatibilism. Compatibilism says that freedom of the will and natural mechanism can co-exist. On my interpretation of Kant's theory of freedom and Moore's Basic Idea, compatibilism is false. This is because according to this interpretation, all causation bottoms out in event-causation, and there are no events that are at once free and naturally mechanized. And since all individual substances and agents are complex events, there are also no individual substances or agents that are at once free and naturally mechanized. All the conscious animals and in particular the

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

rational animals and their actions are both alive and spontaneous, and not naturally mechanized.

Consider now incompatibilism. Incompatibilism says that freedom of the will and natural mechanism cannot co-exist. On my interpretation of Kant's theory and Moore's Basic Idea, incompatibilism is also false. This is because according to this interpretation there can be a natural world parts of which are naturally mechanized, and parts of which are not naturally mechanized. Living organisms, for example, are not naturally mechanized. As Kant puts it, there could never be a biological Newton who could explain the generation of even a single blade of grass (*CPJ* 5: 400). Most relevantly, conscious animals and in particular rational animals are not naturally mechanized. They are alive and spontaneous, 'because the mind for itself is entirely life (the principle of life itself)' (*CPJ* 5: 278). And they have got freedom of the will. So the thesis that there is a strong continuity between biological life and the spontaneity of the will, when combined with an emergentist and non-reductive approach to biological facts, entails the denial of incompatibilism.

Here is another way of putting the same crucial point. It does not follow from the fact that something is free, that it *violates* the laws of natural mechanism. We can do only those things that are *permitted* by the laws; but at the same time the laws themselves together with the settled facts *do not necessitate* our intentional actions, even if what merely happens to us (as opposed to what we will or do) still *contingently conforms* to the laws. In a precisely similar way, in a moral context, as Kant points out, we can morally do only those acts that are permitted by the moral law (universalizability); but at the same time the law itself does not necessitate our intentional actions (*ought* does not entail *is*), even if what merely happens to us (as opposed to what we will or do) still contingently conforms to the law. It is also true that for Kant we can actually will or do things that only contingently conform to the moral law, if we have done them for reasons other than the moral law itself. But that leaves the distinction between something's being *permitted by the law,* something's being *necessitated by the law,* and something's *contingently conforming to the law,* perfectly intact.

This point is intimately connected to Kant's idea, developed in the First Introduction to the *Critique of the Power of Judgment,* that there is an explanatory and ontological gap between what in

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

the first *Critique* he had called the 'transcendental affinity' of nature (= its transcendentally nomological character), and its 'empirical affinity' (= its empirically nomological character) (*CPJ* 20: 208–211; see also *CPR* A122–128, B163–165). And this in turn is intimately connected to the problem of 'empirical laws.'[8] More specifically, Kant is committed to the thesis that even allowing for the existence of universal, transcendental laws of nature, and also for the existence of general mechanistic laws of nature, it does not automatically follow that there are specific empirical laws of nature 'all the way down.' Indeed, nature might still be lawless and chaotic in its particular empirical details. If we take this problem seriously, then it is arguable that for Kant in the third *Critique* the assumption that nature is pervasively deterministically nomological is merely a regulative but not constitutive principle of the understanding, which could then fail to apply to *all* of the material objects studied in natural science. In that case, then neither the universal transcendental laws nor the general mechanistic causal laws of nature would determine the specific behaviours and natures of *all* material objects. And in particular they would not determine the specific behaviours and natures of non-animal organisms, non-rational animals, or rational animals.

Now assuming that this suggestion is correct, what can close the nomological gap? The answer is that transcendentally free rational animal choices produce natural causal singularities, and *one-time laws,* and thereby *freely complete nature.* Transcendentally free agents thus create new unique empirical causal-dynamic laws of nature that fall under, and are *permitted by,* but are not *compelled or necessitated by,* the general laws of natural mechanism. This in turn is the same as what Moore calls creating novel concepts or new sense. If we frame this point in terms of properties rather than concepts, then what I am saying is that for Kant in the third *Critique,* in order to explain the behaviours and natures of living organisms, including of course the behaviours and natures of rational human animals, we are theoretically obliged to posit the existence of causally efficacious emergent properties that naturally arise from self-organizing complex dynamical systems.[9] Given Kant's anti-Humean view that empirical causal-dynamic laws are intrinsic to the events they nomologically govern,[10] it then follows that these laws themselves are also emergent and 'one-off.' Nature is not mechanistic either all the way down or all the way through:

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

it is only partially naturally mechanized, but also partially alive and partially spontaneous. As transcendentally free rational animals with embodied wills, we enrich and ramify the causal-dynamic nomological structure of material nature by being the authors of its most specific empirical laws. In this way not only do we make a causal difference, we also freely *make* nature, in part and on an appropriately human scale. As finite and radically evil, we are most certainly not gods. But we *are* small-time creators. And how much more power over nature could we really want?

But what then is nature? On Kant's view, nature contains nothing but material or spatiotemporal events and substances, yet some of them are not naturally mechanical but are in fact biologically alive and thereby instantiate some emergent non-mechanical intrinsic structural properties, and in particular the property of being conscious and rational. To put a twist on Josiah Royce's famous definition of idealism ('the world and the heavens, and the stars are all *real,* but not so *damned* real'[11]), the natural world is everywhere *physical,* but not so *damned* physical. On this view of reason and freedom, then, biological life and mind are one and the same, and they are dynamically emergent intrinsic structural properties of a neutral non-mechanical, non-mental 'gunk' or fluid aether (*OP* 21: 206–233) that consists of a system of dynamic events and forces, and consciousness is continuous with animal life in suitably complex, suitably structured animals. Some of those animals are rational human animals or persons. Thus the natural world contains, in addition to natural mechanisms and biological/mental facts, a further set of dynamically emergent intrinsic structural properties, which together with the natural mechanisms and biological facts, jointly constitute human persons and their living, embodied, spontaneous wills.

In this way, we can make Kant's embodied libertarian rationalism depend on the idea that our innate drive towards rationality is the same as the conjunction of our human biological life and spontaneity of the will, which in turn is necessarily embodied, given that the mind is identical to life. Another way of putting this is to say that if biological life and mind are the same, then since human rationality includes conscious mind, it follows that rationality is necessarily embodied, and that the embodiment of rationality is identical to our capacity for free choice. *The human will, for better or worse, is rationality incarnate.* Yet another way

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

of putting it is to say that the human will, whether as *Willkür* or as pure *Wille,* is necessarily spatiotemporally located and materially real, neurobiologically real and alive, irreducible to natural mechanisms, causally efficacious, unprecedented or temporally underdetermined, inherently creative, inherently perverse, self-guiding, theoretically reasonable, practically reasonable, and morally sublime.

2. A.W. Moore

I am extremely grateful to Robert Hanna for the great care with which he has read my book, and for the great generosity with which he has engaged with it.[12] Although I believe that there are several misunderstandings, some of which are pretty serious and one of which I shall try to correct in this reply, I am also aware of how much of the blame lies, not in his reading of the text, but in the text itself.[13]

Correcting that misunderstanding is one of two principal aims that I have. The other connects with the thesis which Hanna develops in the latter part of his essay, in contradistinction to some of my own ideas, and which he calls 'embodied libertarian rationalism'. Embodied libertarian rationalism is a thesis with two components: first, that the biological life of a human being and the spontaneity of that human being's will together constitute a structural property of his or her animal body, what we might call the human being's *vitality;*[14] and second, that manifestations of this vitality occur in the slack left over by mechanistic laws of nature, which, although they determine some of what happens in nature, do not determine everything that happens there. Hanna sees this thesis as both exegetically important, in as much as it has a grounding in Kant's texts, and philosophically defensible in its own right. He presents it as part of the best answer to that fundamental Kantian question, 'How can pure reason be practical?' The second of my aims is to say something about where I think embodied libertarian rationalism stands in relation both to my own ideas and to Kant's.

To begin, then, with the misunderstanding. This concerns what I call the Incommensurability Thesis. Hanna cites the definition of the Incommensurability Thesis that I give in my book: 'exercise of

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

the concept of physical [determination] precludes exercise of the concept of freedom' (*NIR:* 114, emphasis removed).[15] The idea is that these two concepts are incommensurable, not incompatible. In other words, it is not that there is some conceptual rule that prevents their co-application; it is rather that the conceptual rules that govern one of them do not govern the other at all. Suppose that someone asserts, of some given action, that it was physically determined. He or she is not thereby committed to denying that it exhibited freedom as well. Rather, what he or she thereby does is to 'bracket', or to put to one side, the question of whether it exhibited freedom, so that the question of whether it exhibited freedom does not so much as arise, at least while what is at issue is whether the action really was physically determined. An analogy that I use in my book to illustrate this idea is the contrast between the two following claims that someone might make in the course of a game:

(1) The next move in this game cannot be a pawn move because if White moves any of his pawns, then he will place himself in check.
(2) The next move in this game cannot be a pawn move because it is a game of draughts.

The 'cannot' in (1) is like the 'cannot' of incompatibility; the 'cannot' in (2) is like that of incommensurability. There is, of course, much more to be said about this idea of incommensurability; and the distinction between incommensurability, on the one hand, and various different species of compatibility and incompatibility, on the other hand, is by no means always sharp. But I hope that these comments give some indication of what I have in mind.

A brief caveat before I go any further. I am presenting the Incommensurability Thesis as 'my' thesis. And I do indeed believe that, suitably construed, this thesis is correct. But I claim no originality for it, nor do I make any attempt to defend it in my book. It is a thesis that I mention almost parenthetically. It does not play the significant rôle in my thinking that I think Hanna thinks it plays. The bulk of what I say in the second part of my book, the part with which Hanna is concerned, is impervious to the Incommensurability Thesis, and would, I hope, survive its rejection. Be that as it may, I do endorse this thesis, and I do think that the question of how it relates to theses that Hanna and Kant endorse remains of great interest.

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

Now Hanna presents the Incommensurability Thesis as though it were a variation on the theme of Davidson's anomalous monism.[16] He explicitly draws a comparison with what he calls 'conceptual non-reductionism' in the philosophy of mind, which he says is logically consistent with what he calls 'ontological reductionism'. I am not entirely sure what he means by these terms, but I take this to be an allusion to the Davidsonian idea that, although mental concepts are quite independent of physical concepts, still they may apply to the very same things: mental events may *be* physical events. This makes the Incommensurability Thesis consistent with a free action's *being* physically determined. Or to put it in Hanna's own terms, it makes the Incommensurability Thesis consistent with a free agent's being 'naturally mechanized'. What this in turn means, Hanna complains, is that the freedom in question is not real freedom. It is at best only 'phenomenal' freedom, a feature of how our own agency strikes us – which, if our own agency is in fact naturally mechanized, is, in Hanna's evocative phrase, 'a tragic illusion'. As Hanna sees it, the problem with the Incommensurability Thesis is that it is a version of classical compatibilism: it leaves us with a freedom which, precisely because it is compatible with natural mechanism, is not the real article. It is in this spirit that Hanna advocates his rival view, embodied libertarian rationalism, which he claims is *neither* compatibilist *nor* incompatibilist. And he further claims that this rival view has a grounding in Kant.

I want to turn the tables completely here. Just as Hanna contends that my view is a version of classical compatibilism, whereas his is neither compatibilist nor incompatibilist, I want to contend that my view is the one that is neither compatibilist nor incompatibilist, whereas his is a version of classical *in*compatibilism. And where Hanna wants to claim that Kant's view is likewise neither compatibilist nor incompatibilist, I want to claim that, on the contrary, Kant's view is in some sense both. That, it seems to me, is precisely what makes Kant's view ultimately unsatisfactory.

As regards my insistence that the Incommensurability Thesis is neither compatibilist nor incompatibilist, *that* – in a way – is its whole point. The chess/draughts analogy was supposed to illustrate this. If what you are playing is draughts, then there is no question of the next move's being a pawn move. If what you are playing is the language game of freedom, then there is no question of your saying that an action is physically determined. *Pace* Hanna,

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

the Incommensurability Thesis is *not* consistent with a free action's being physically determined. On the contrary, it casts 'This free action is physically determined' as a piece of nonsense.

As regards my reservations concerning Hanna's claim that his own view is neither incompatibilist nor compatibilist, let us consider how Hanna defends this claim. He defines incompatibilism as the view that freedom and natural mechanism cannot co-exist; he defines compatibilism as the view that freedom and natural mechanism *can* co-exist; and he distances himself from each. But there is an equivocation here on 'co-exist'. What he means by 'co-exist', when he distances himself from incompatibilism, is 'exist in the same world'. What he means by 'co-exist', when he distances himself from compatibilism, is 'exist in the same thing (event, substance, agent)'. This makes his claim to be neither an incompatibilist nor a compatibilist something of a sham. And if what is at stake is what is usually at stake in philosophical discussions of these issues – roughly, whether it is possible for *everything* in nature to be naturally mechanized *and* for nature to contain freedom – then Hanna's view is straightforwardly incompatibilist. He thinks that this is not possible.

On Hanna's view, which he also takes to be Kant's view, if human beings ever act freely, then this must be because natural mechanism does not determine everything that happens in nature. It must be because natural mechanism leaves gaps, within which freedom operates. And the way in which freedom operates within these gaps is by filling them with what Hanna calls 'causal singularities', that is to say, if I understand him correctly, events that are governed by laws, but by laws of a maximally specific kind: 'one-time' laws that govern those events and those events alone.

In attributing this view to Kant, Hanna draws an analogy with the way in which the moral law, although it is a constraint of sorts on what human beings do, leaves gaps of permissibility within which freedom can operate. I have several misgivings about this analogy. First, Hanna says that the moral law no more necessitates all that we do than mechanistic laws of nature necessitate all that we do, adding in parenthesis '*ought* does not entail *is*'. But the fact that *ought* does not entail *is*, which is basically a fact about the moral impermissibility of some of what we do, seems to me to be completely beside the point here, and indeed out of tune with the analogy. (The fact that *ought* does not entail *is* has no counterpart

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

in the case of mechanistic laws of nature.) If the analogy is to be a reasonable one, then the question of necessitation in the moral case should be with respect to morally permissible worlds, just as the question of necessitation in the case of natural mechanism is with respect to worlds that do not violate any mechanistic laws of nature. But as far as *that* question goes, *ought* does entail *is:* whatever ought to happen, in a morally permissible world, does happen. This is related to Hanna's claim that some of what happens to us 'contingently' conforms to mechanistic laws of nature. In what sense of 'contingently'? With respect to worlds that do not violate any mechanistic laws of nature, *nothing* that conforms to those laws does so contingently (for conforming to those laws is a precondition of happening at all). With respect to a broader range of worlds, say logically possible worlds, *everything* that conforms to those laws does so contingently (for the laws themselves are contingent). Similarly in the moral case.

True, in the moral case, there does seem to be some distinction between actions that conform to the moral law as a matter of necessity and actions that do so merely contingently – the very distinction to which Hanna subsequently draws our attention. But that is an entirely different matter, which has no analogue, as far as I can see, in the case of natural mechanism. That is a matter of its being possible to characterize actions without reference to what motivates them. The point is this. Given such a characterization, we may be able to see that the action in question conforms to the moral law. But it is then a further question whether the agent is acting morally or not: that depends on whether or not the moral law is what is motivating him. If the moral law *is* what is motivating him, then, relative to his motivation (and prescinding from complications concerning any 'special disfavour of fortune' or 'the niggardly provision of a stepmotherly nature' [*GMM* 4: 394]), it is no mere contingency that his action conforms to the moral law. If the moral law is *not* what is motivating him, then, relative to his motivation, it *is* a mere contingency (*GMM* 4: 397–400). But to repeat: I see no analogue of this in the case of natural mechanism.

There is still of course the idea that the moral law leaves gaps of permissibility within which freedom can operate (which may indeed be all that Hanna means by saying that *ought* does not entail *is* – although, if that *is* all he means, then he is guilty of expressing himself in a misleading way). It is worth noting,

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

however, that this idea, like the idea that we can freely do what is impermissible, allows for exercises of freedom that are beyond the control of pure reason, which means that it is like the idea that we can freely do what is impermissible in another respect too: although it is certainly to be found in Kant (*GMM* 4: 439 and *CPrR* 5: 66), it is arguably 'un-Kantian'.

Be that as it may, there is still the question of whether Kant believes that natural mechanism leaves analogous gaps; gaps which are filled by 'causal singularities', serving as the loci of human freedom. Hanna, it seems to me, gives little in the way of evidence for the claim that he does. He appeals to the passage from *Critique of the Power of Judgment*, in which Kant says that 'it would be absurd for humans ... to hope that there may yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws that no intention has ordered' (*CPJ* 5: 400). But that passage can be interpreted as making a quite different point, about the possibility of teleological principles supervening on a completely naturally mechanized subvenient base.

I think that Kant accepts determinism, the thesis that everything that happens in nature is completely determined by its antecedent conditions in combination with mechanistic laws of nature. Furthermore, I think that he wants to combine this with both libertarianism, the thesis that some of what we do we do freely, and incompatibilism, the thesis that determinism and libertarianism, thus defined, are in some sense incompatible with each other.[17]

This shows what I mean when I claim that Kant is in some sense both a compatibilist and an incompatibilist. The way in which Kant thinks he can have his cake and eat it is by assimilating the incompatibility between determination and freedom that he endorses to the incompatibility between rest and motion. There is a sense, a perfectly straightforward sense, in which rest and motion are incompatible with each other. We can all agree that a physical object which is at rest cannot at the same time be in motion. Nevertheless, a physical object, a luggage rack say, can be both at rest relative to a train and, at the same time, in motion relative to an embankment. The same sort of relativism, Kant thinks, applies in this case. He believes that an event can be both completely determined by natural mechanism, when considered from one point of view, and free, when considered from another.[18]

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

The second of these points of view involves reference to an atemporal reality beyond the world of nature, in which free agency is ultimately to be located and with respect to which the world of nature is mere appearance. This is why I cannot ultimately accept Hanna's idea that, for Kant, freedom operates in gaps that mechanistic laws leave within the world of nature, still less that it does so by filling these gaps with 'causal singularities' – by creating 'one-time' laws – where this in turn is to be understood in such a way that freedom is essentially embodied. I think that Kant's writings abound with material that tells against this interpretation. One example is the section from *Critique of Pure Reason* entitled 'Resolution of the Cosmological Idea of the Totality in the Derivation of the Occurrences in the World from their Causes' (*CPR*, A532–558/B560–586), which seems to me more or less decisive.

I shall not say much more about this now, even though there is much more (obviously) to be said. This is not least because I doubt whether there is much more that I can say that is not both exceedingly familiar and, for anyone who reads Kant differently, unpersuasive. But I shall add just one point, and then indicate, very briefly, why I think that Kant's reconciling project fails (which is incidentally not for the reasons that Hanna suggests).[19]

The point that I want to add is this. I do take Kant to be committed to a kind of incompatibilism, and not to the Incommensurability Thesis. There are some crucial passages in which he might be interpreted in either way. But, much as I would like to, I cannot ultimately read him as holding the Incommensurability Thesis – even though I do think that, if he had held it, then his conception would not have been vulnerable to my main objection.[20]

That objection is as follows. There needs to be an answer to the question, 'Which of the things that we do exhibit freedom?' If Kant's conception is to have any chance of being taken seriously, then it must also have some chance of connecting with the imputations that we are antecedently inclined to make. Thus John cannot be said to have acted freely when he suddenly jumped at that gunfire, nor when he came down with flu last week. But now: what *are* the imputations that we are antecedently inclined to make? If there is anything in this area that we are antecedently inclined to do, then it is to revise our imputations in the light of further knowledge. We think twice about saying that a shoplifter is acting of her own free will when we discover that she is a kleptomaniac. But –

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

and this is the crucial point – what we are antecedently inclined to do, if we become persuaded of determinism *and* become persuaded of the incompatibilism on which Kant insists, is to deny that there is any freedom at all. It is of no avail for Kant to argue that his reconciling project shows that we do not need to do this. The reconciling project comes one consideration too late. It is what we are *antecedently* inclined to do that dictates what is available to be reconciled.

Notes

1 This paper is a revised version of a one-on-one discussion presented at the 'Free Will, Agent Causation, and Kant' conference at the University of Sussex in June 2005. We would like to thank the British Academy and the University of Sussex, whose support made the conference possible; Lucy Allais, who organized the conference; and the other conference participants, whose comments and questions helped guide the revision of the discussion.

2 For convenience, we refer to Kant's works infratextually in parentheses. The citations include both an abbreviation of the English title and the corresponding volume and page numbers in the standard 'Akademie' edition of Kant's works: *Kants gesammelte Schriften,* edited by the Königlich Preussischen (now Deutschen) Akademie der Wissenschaften (Berlin: G. Reimer [now de Gruyter], 1902-). We generally follow the standard English translations, but have occasionally modified them where appropriate. For references to the first *Critique,* we follow the common practice of giving page numbers from the A (1781) and B (1787) German editions only. Here is a list of the relevant abbreviations and English translations:

CPJ *Critique of the Power of Judgment*, trans. P. Guyer and E. Matthews (Cambridge: Cambridge University Press, 2000).

CPR *Critique of Pure Reason*, trans. P. Guyer and A. Wood. (Cambridge: Cambridge University Press, 1997).

CPrR *Critique of Practical Reason*, trans. M. Gregor, in *Immanuel Kant: Practical Philosophy* (Cambridge: Cambridge University Press, 1996), pp. 133–272.

GMM *Groundwork of the Metaphysics of Morals*, trans. M. Gregor, in *Immanuel Kant: Practical Philosophy*, pp. 37–108.

MM *Metaphysics of Morals*, trans. M. Gregor, in *Immanuel Kant: Practical Philosophy*, pp. 353–604.

OP *Immanuel Kant: Opus postumum*, trans. E. Förster and M. Rosen (Cambridge: Cambridge University Press, 1993).

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

3 A. W. Moore, *Noble in Reason, Infinite in Faculty: Themes and Variations in Kant's Moral and Religious Philosophy* (London: Routledge, 2003).

4 W. Sellars, 'Philosophy and the scientific image of man,' in W. Sellars, *Science, Perception and Reality* (New York: Humanities Press, 1963), pp. 1–40.

5 See O. O'Neill, *Constructions of Reason* (Cambridge: Cambridge University Press, 1989), ch. 2.

6 See P. Guyer, *Kant and the Experience of Freedom* (Cambridge: Cambridge University Press, 1993).

7 See J. Fodor, 'Making mind matter more,' in J. Fodor, *A Theory of Content and Other Essays* (Cambridge: MIT Press, 1990), pp. 137–59, at 156.

8 The problem is how to understand both the apparently a priori epistemological and also strongly modal status of these laws, in view of the fact that they are explicitly held to be empirical. See, e.g. H. Allison, 'Causality and causal laws in Kant: a critique of Michael Friedman,' in P. Parrini (ed.), *Kant and Contemporary Epistemology* (Netherlands: Kluwer, 1994), 291–307; G. Buchdahl, *Metaphysics and the Philosophy of Science* (Cambridge: MIT Press, 1969), pp. 651–65; G. Buchdahl, 'The conception of lawlikeness in Kant's philosophy of science,' in L. W. Beck (ed.), *Kant's Theory of Knowledge* (Dordrecht: D. Reidel, 1974), 128–50; P. Guyer, *Kant's System of Nature and Freedom* (Oxford: Oxford University Press, 2005), ch. 2; M. Friedman, *Kant and the Exact Sciences* (Cambridge: Harvard University Press, 1992), chs. 3–4; M. Friedman, 'Causal laws and the foundations of natural science,' in P. Guyer (ed.), *The Cambridge Companion to Kant* (Cambridge: Cambridge University Press, 1992), pp. 161–99; W. Harper, 'Kant on the a priori and material necessity,' in R. Butts, R. (ed.), *Kant's Philosophy of Physical Science* (Dordrecht: D. Reidel, 1986), pp. 239–72; R. Walker, 'Kant's conception of empirical law,' *Proceedings of the Aristotelian Society*, 63 (1990): 243–58; and E. Watkins, 'Kant's justification of the laws of mechanics,' in E. Watkins (ed.), *Kant and the Sciences* (New York: Oxford University Press, 2001), pp. 136–59.

9 See: H. Haken, *Principles of Brain Functioning: A Synergetic Approach to Brain Activity, Behavior, and Cognition* (Berlin: Springer, 1996); A. Juarrero, *Dynamics in Action* (Cambridge: MIT Press, 1999); J. S. Kelso, *Dynamic Patterns* (Cambridge: MIT Press, 1995); Port and T. Van Gelder (eds), *Mind as Motion: Explorations in the Dynamics of Cognition* (Cambridge: MIT Press, 1995); E. Thelen and L. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge: MIT Press, 1994); F. Varela, *Principles of Biological Autonomy* (New York:

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

Elsevier/North-Holland, 1979); and A. Weber and F. Varela, 'Life after Kant: natural purposes and the autopoietic foundations of biological individuality,' *Phenomenology and the Cognitive Sciences,* 1 (2002): 97–125. The notion of self-organization used by contemporary theorists of complex systems dynamics is slightly broader than Kant's, in that it includes non-living complex systems as well, e.g. the rolling hexagonal 'Bénard cells' that appear as water is heated. Kantian self-organizing systems are all holistically causally integrated or 'autopoietic,' such that the whole and the parts mutually produce each other.

10 See E. Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005), ch. 4.

11 J. Royce, *The Letters of Josiah Royce* (Chicago: University of Chicago Press, 1970), p. 217.

12 Like Hanna, I shall refer to my own book as *NIR*.

13 Among the misunderstandings that I shall not discuss, two, I think, are worth mentioning in a footnote. One of these concerns what I call the radical picture. Hanna rightly points out that, while I do not claim to find the radical picture in Kant, I do claim to find pressures in Kant's system to endorse it. Hanna develops this point in terms of Kant's *Wille/Willkür* distinction, as though that were the place where I took the pressures to be greatest; actually, that is the place where I take Kant to be doing most to keep the radical picture at bay. The second misunderstanding comes in Hanna's claim that I identify rational freedom with the creation of new concepts. I certainly put a heavy emphasis on the creation of new concepts as a paradigm of rational freedom. But I am just as keen to recognize rational freedom in the exercise of old concepts. The creation of new concepts and the exercise of old concepts have much in common, and I do not want to suggest that only when the element of autonomy that is characteristic of both reaches the intensity that is characteristic only of the former does it constitute freedom.

14 This is my term, not Hanna's.

15 I have slightly modified the definition, replacing 'determinism' by 'determination'. I take this to be an inessential difference, though the modified version is somewhat more convenient for my current purposes.

16 D. Davidson, 'Mental events', reprinted in his *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).

17 In my book I tell an old joke which I take to illustrate Kant's extraordinary ambitions here (*NIR:* 209, n. 1). I shall hereby allow myself the indulgence of repeating this joke. Two people are in bitter dispute with each other about whether some proposed course of action can be justified. They consult a sage. To the one who says that the course of

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press

action can be justified the sage says, 'You are right.' To the one who says that it cannot be justified the sage says, 'You are right.' A bystander protests, 'But they can't both be right: their views are incompatible.' Turning to the bystander, the sage says, 'And you are right too.'

18 I am here drawing on material from *NIR*, theme two, §3.

19 I shall be drawing on material from *NIR*, theme two, §5.

20 See further, with references, *NIR*, pp. 114–15.

https://doi.org/10.1017/S1369415400000832 Published online by Cambridge University Press