

Dynamics of Reason and the Kantian Project

ABSTRACT

I show why Michael Friedman's idea that we should view new constitutive frameworks introduced in paradigm-change as members of a convergent series, introduces an uncomfortable tension in his views. It cannot be justified on realist grounds, as this would compromise his Kantian perspective, but his own appeal to a Kantian regulative ideal of reason cannot do the job either. I then explain how to make better sense of the rationality of paradigm-change on what I take to be Friedman's own terms.

1. Introduction

The aim of our modified version of Kantianism [...] has nothing to do with certainty or epistemic security at all. It aims, rather, at precisely [...] universal rationality, as our reason grows increasingly self-conscious and thereby takes responsibility for itself. (Friedman 2001, 68)

Such are the last sentences of Michael Friedman's *Kant Lectures* published as Part I of the fittingly titled monograph *Dynamics of reason*. They make admirably clear what is at stake for Friedman: he sees it as philosophy's responsibility to speak up for reason, and to uphold Enlightenment values in the face of a "widespread relativistic tide" (Friedman 2001, 57). In mounting his sophisticated defence of rationality he tries to offer a way out of the incommensurability conundrum bequeathed to philosophy of science by the work of Thomas Kuhn, but he also explicitly denies that his specific way of defending the rationality of paradigm-changes would commit him to scientific realism (ibid., 117). This denial might be seen as creating an uneasy tension in Friedman's position, though (as argued, e.g., by Slowik 200X), since he stresses the convergence of succeeding paradigms as an essential part of the paradigm-transcending rationality. But belief in such convergence is something that is typically underwritten by realist commitments: why would one expect it, if not because one believes that our scientific methods allow us to learn something about the true structure of the world? And it is a striking feature of Friedman's presentation that he actually doesn't seem to

offer much of an argument for this convergence: he only introduces the historical fact that it is possible to notice a kind of convergence from the Aristotelian to the Newtonian to the Einsteinian space-time theories, which in itself obviously cannot suffice to believe that this will continue to hold true. So, is he implicitly endorsing scientific realism in stipulating the continuing convergence? And if he is not, what sense can we make of his stress on the necessary convergence? A careful consideration of these questions might help us to reach a better understanding of the precise nature of Friedman's defence of universal rationality. It will turn out that we should take his summary in the opening quote more seriously than most people might be tempted to do.

2. The relative a priori and the dynamics of reason

Friedman is probably best known for his forceful reintroduction of the notion of constitutive, a priori principles in contemporary philosophy of science. Taking his lead from Hans Reichenbach's early work he argues that Kant's views on the a priori prerequisites for natural science still hold promise if we distinguish between on the one hand the *constitutive* role of a priori principles, and on the other hand their *supposed unrevisability*. He stresses that a detailed analysis of the structure of Newtonian mechanics and special and general relativity shows that we should discern three different layers within these theories. On a first level we have purely mathematical structure (e.g. Euclidean geometry and calculus for Newton's theory); on a second level we have a special class of principles, which allow the coordination of mathematical structures with empirical phenomena (e.g. Newton's three laws of motion, which allow one to identify the inertial frames of reference in which one can then first unambiguously relate mathematically characterized paths of motion to their empirical counterparts); and on a third level we have the theory's properly speaking empirical laws (e.g. the law of universal gravitation). The most important point about this layered structure is that the different levels have an epistemologically asymmetrical function: the first two levels first allow us to ascribe truth values to the empirical claims of the theories, and thus are *constitutive* with respect to the third level. This asymmetry also implies that it makes no sense to think that the constitutive principles themselves could be tested empirically, since all meaningful empirical tests must already presuppose them. But this does not imply that for that reason they would be unrevisable, as the historical progression from the Newtonian to the Einsteinian space-time frameworks forcefully makes clear.

This does imply, however, that we are once again confronted with the spectre of radical incommensurability, as it is no longer clear how we can rationally decide to switch constitutive principles: it seems that that these principles in the first place make possible rational agreement, because they first constitute what Friedman calls “the empirical space of reasons” (ibid., 85). But this problem, which arises exactly because we can no longer hold on to Kant’s belief in the absolute unrevisability of constitutive principles, can be overcome by paying attention to another distinctively Kantian insight, Friedman claims. Kant’s keen analysis of the crucial role of constitutive principles was only possible because of his specific understanding of the task of philosophical analysis vis-à-vis scientific practice. He saw that the most fruitful position for philosophy was that of a “transcendental” meta-discipline which critically investigates the presuppositions of science. Putting metaphysics on the sure path of science implied giving up the false idea that philosophical analysis could proceed independently from science to give the latter its true foundations, but also resisting the idea that it had to be conceived as an extension of scientific insights to a further, supposedly metaphysical domain. And it is exactly philosophy considered along these lines that historically provided the resources to make paradigm-change intelligible and made it possible to switch constitutive frameworks through a process of consensus-creating reasoned argumentation. Friedman argues for the latter point along two lines, with the first taking clear predominance: he first presents a subtle historical narrative, and he then gestures towards a philosophical argument. The historical story explains in convincing detail how Einstein’s revolutionary moves were only possible because of his engagement with the philosophical debates on space-time that crucially included Helmholtz and Poincaré. And these philosophical debates in turn were a direct outcome of a critical investigation of Kant’s transcendental analysis of the conditions of possibility of (applied) mathematics. Side by side to the successful application of the Newtonian framework to ever more empirical phenomena (the Kuhnian process of normal science) was developing a lively dialogue on the “quid juris” of this framework, a dialogue which was initiated by Kant’s own analyses but further informed by developments internal to these mathematical and empirical sciences (just as Kant’s views were informed by the state of these sciences at his time). And it are these philosophical reflections which as “philosophical meta-paradigms or meta-frameworks” play “an indispensable role, by serving as source of suggestions and guidance – for orientations, as it were – in motivating and sustaining the transition from one paradigm or conceptual framework to another” (ibid., 46). Einstein could only have made his revolutionary moves because he operated against this level of philosophical meta-reflection; and exactly because his moves had this background, they

had to be taken serious and considered as live options in the science of his days. In the place of an empirical testing of the constitutive principles comes a sustained philosophical inquiry of their right to claim special status, which in the right circumstances (Kuhnian crisis) can help open up the conceptual space for the crafting of a new “empirical space of reasons”. Kuhn’s views on history of science only raise the spectre of irrationality because they unduly neglect the parallel evolutions in history of philosophy.

In arguing for this point of view, Friedman introduces the notion of “communicative rationality”, which he takes over from Jürgen Habermas (ibid., 53ff). This notion is intended as an articulation of the Enlightenment faith in human rationality, by offering the ideal of “non-coercively uniting, consensus creating power of argumentative speech” (Habermas, quoted in Friedman 2001, 54). And this is exactly the kind of rationality that not only characterizes normal science (due to a shared empirical space of reasons), but also paradigm-change (due to the presence of philosophical meta-paradigms sustaining well-articulated “argumentative speech”). Conclusion: we can still uphold our best mathematical sciences as exemplars of human rationality.

It is after having added historical evidence that we can indeed describe paradigm-changes as communicatively rational, that Friedman adds that, “from a philosophical point of view”, “we can thus view the evolution of succeeding paradigms or frameworks as a convergent series, as it were, in which we successively refine our constitutive principles in the direction of ever greater generality and adequacy” (ibid., 63). This claim is backed up by historical evidence (the possibility of retrospectively approximating features of the old framework within the new one, and the possibility of “prospectively” seeing the new framework as arising out of the old through a continuous development – a continuity only made possible by the philosophical meta-framework, of course), but as already indicated in the introduction, its scope is actually much wider than can be established by purely historical evidence and the latter must thus be supplemented by philosophical argumentation – as Friedman himself indicates by introducing it as following “from a philosophical point of view”. This is further underscored by his switch to explicitly normative language in repeatedly stating that we *should* always be able to view paradigm-change in this way. These strong statements are backed up by an appeal to Kant’s conception of the regulative use of reason, which offers an ideally completed state of scientific progress as a *focus imaginarius* guiding all empirical research (cf. also § 3.2 below), but which Friedman now extends to also make possible “trans-historical universality” within the domain of shifting constitutive principles. This idea is inspired by Cassirer’s way of rethinking Kant’s transcendental idealism in the light of revolutionary changes, but Friedman

stresses that, more than Cassirer, he still wants to leave room for relativized constitutive principles within this ideal progression (ibid., 66, fn. 80) – providing temporary points of rest as it were.

But why exactly do we need this regulative use of reason in our defence of rationality? At this point Friedman is disappointingly vague. It would seem that the appeal to philosophical meta-paradigms already does the job of securing communicative rationality, in offering the possibility of coming to reasoned intersubjective agreement. Why does he want to add an appeal to an ideal of reason as giving extra, “trans-historical” direction to this process of argumentative deliberation? I think that a glimpse of the largely hidden motivation can be seen in the way Friedman introduces a summary of his views with the question: “how ... can the proposal of a radically new conceptual framework be ... both rational *and responsible*?” (Ibid., 66; my emphases.) But by thus indicating that reasoned intersubjective agreement might not be enough for fully universal rationality, this extra appeal to “responsibility” could lead us back to the suspicion that this hidden motivation is actually one of a kind with scientific realism – as realists typically portray it as the *responsibility* of scientists to uncover the world’s true structure, which true structure would then ground the convergence of scientific theories. As already mentioned, it is clear that Friedman himself does not intend his argument in this way, as he explicitly warns us that his views on the necessary convergence of successive frameworks must not be understood as “ever better approximations to a radically external world existing independently of the scientific enterprise itself”, since the “original Kantian conception of objectivity ... was explicitly intended to undermine such a naively realistic interpretation of scientific knowledge” (ibid., 67). If we thus want to come to a better understanding of Friedman’s possible agenda in stressing the need of “responsible” paradigm-change – and especially its relation to the series of converging frameworks – we must start by taking a closer look at this original Kantian conception of objectivity and its possible relation to present-day discussions on scientific realism. This will prove to be especially interesting because we will see that Kant’s own appeal to the regulative use of reason is closely connected to some of the issues that are at stake in these discussions.

3. Transcendental idealism and scientific realism

In the following I will propose a reading of Kant’s transcendental idealism which has to meet two constraints: it should be formulated at high enough a level of abstraction, such that it also can be used in helping to characterize Friedman’s position (this implies remaining as

uncommitted as possible on the vexed issue of the status of a priori forms of sensibility – Friedman definitely owes us further clarifications about the position such forms occupy in his modified Kantianism); and it should be a reading which fits Friedman’s claim that being Kantian implies not being a scientific realist.

3.1 *The transcendental stance*

One way to understand Kant’s transcendental idealism is seeing it as the outcome of a decision to take at face value the fact that we, human beings, *have to think* in confronting reality – and to let this fact provide its own norms in our philosophical views on this confrontation.¹ Let me try to unpack this rather cryptic statement a bit. All classical (i.e., transcendently realist) metaphysical positions have started from the supposition that reality has a determinate structure and that we must thus judge our knowledge by trying to assess how it fares in bringing this structure to light. But ascribing a determinate structure to reality is of course already the result of an act of thinking, whereas *this* structure is supposed to be completely independent of all human activity. Rather than deciding with the rationalists that reality in itself must already have an intellectual structure (a conclusion that is traditionally backed up by an appeal to God), or with the empiricists that we can never apprehend this independent structure (exactly because we are always forced to think), transcendental idealism decides to turn philosophical reflection away from this postulated structure and direct it on precisely the human acts of thinking which first allow determinate structures *to have sense for us*. Reflecting on what it means that structures can have sense for us implies that we have to ask what it actually means to be *thinking about* something, which then brings us to the central Kantian question: how does our thinking relate itself to its objects? The answer consists in noticing that this relation is (and can only be) brought about by *presupposing* certain (“objectivating”) conditions which then allow us to recognize and re-identify objects under (universal) concepts. Thinking posits its own norms (the categories) by which any human can transcend her particular subjective perspective and relate this to an objective world shared by all human beings. And it is because of this norm-bound activity that we are first able to judge on the truth and falsity of empirical claims referring to this objective world – which is why Kant calls his transcendental analysis of these norms a “logic of truth”

¹ Kant-interpretation is of course a lively philosophical field in its own right, and I have neither the space nor the intention to defend my particular understanding in any detail here (but for the fact that it should meet the two constraints just mentioned). Let me just point to Allison 2004 and Longuenesse 1998 as two important works which have greatly influenced my reading of Kant.

(A63/B87). And it is the decision to take the functioning of these norms seriously which then brings the transcendental idealist to the view that it literally doesn't make sense to ask what structure the world would have in itself, independently of our human way of knowing it: this is simply not the right kind of question.

In stressing the fundamental decision lying behind transcendental idealism I follow Henry Allison in construing it as a philosophical *stance*. But it is important not to lose sight of the fact that this stance has important "metaphysical" implications. Most important among these is the absolute primacy of form over matter, which follows upon the highlighting of the necessity of a priori conditions in our cognitive approach towards the world. There simply are no facts of the matter independently of our cognitive activity which first allows us to ascribe structure to the world. It is important that this does not imply that the cognitive norms governing this structure-constituting activity would be arbitrary for that reason, as is shown by the transcendental deduction in which Kant proves their right to normative status by grounding them in the transcendental unity of apperception. It is of course impossible to enter here into the intricacies surrounding this proof, but it might be useful to recall that its kernel consists in the fact that the categories are shown to be indispensable to the constitution of a *unified experience* – they reflect acts of providing necessary unity to the merely given and as such indeterminate sensations.

There is one more aspect about Kant's transcendental idealism that will turn out to be relevant, and this concerns the thorny issue of "the thing in itself". On Allison's one-world account (which is explicitly endorsed in Friedman 1996, 441, n. 18) it is stressed (1) that talk about "things in themselves" does not refer to a separate class of putative entities which would be radically different from the mere appearances which are subject to our cognitive activity; and (2) that such talk is an unavoidable and necessary consequence of a transcendental perspective. Since in occupying such transcendental perspective a philosopher reflexively isolates our necessary cognitive contribution in confronting objects in the world, this implies a different (distinctively philosophical) way of considering these same objects than in our natural empirical considerations. And because of this reflexive act of isolation it must also be possible to bracket these constitutive conditions and refer to these objects as they are when considered independently from all constitutive conditions. It must be stressed immediately that such a consideration must remain void of cognitive content – we cannot genuinely think about objects when we consider them in this way, but for the *indeterminate* thought of an object in general. But it is not insignificant for that reason, because in doing so we highlight the fact that not their existence, but only their existence *as* objects depends on us: our knowledge of

the world is grounded, but not in determinate things independent from our approach to them. And this must stand as a warning against the illusions with which metaphysical realism presents us (which includes the trap of subjective idealism). We can add that the mere possibility of this dual consideration also teaches us something else of further significance: that our cognitive approach to things, although not arbitrary, does have an essential element of contingency.

3.2 *The regulative ideal and scientific realism*

By now I can be rather brief, as the most important point has already been established in pointing out the metaphysical implication of the transcendental perspective. Since the metaphysical assertion “that the world has a definite and mind-independent natural-kind structure” is supposedly “a basic philosophical presupposition of scientific realism” (Psillos 1999, xix), Kant can apparently not be counted among the scientific realists (although it must be said that the use of “mind-independent” can be ambiguous, as Kant certainly is not a subjective idealist who believes that empirical reality is a product of our minds – a more careful formulation might be to speak about standpoint-independent structure). This is also all that Friedman seems to have in mind when he offers his curt dismissal of realism. We can say a bit more about the relation of transcendental idealism to the contemporary discussions on scientific realism, though; and this will prove relevant in assessing Friedman’s convergence claims.

Scientific realists typically believe in the convergence of scientific theories to an accurate account of the true structure of the world because they hold on to some variant of the no-miracle argument (cf. Slowik 200X, who claims that Friedman’s claims about convergence can only be intelligible when underwritten by this argument), but appeal to this argument is of course ruled out by the transcendental idealist stance: the idea that we could explain the success of science by its approximately mirroring the true structure of the world, where the latter is independent of all (scientific) theorizing, is just a vestige of transcendental realism. But what is more important, for Kant there is also no reason why we *should* explain the success of science along these lines. First, we have a constitutive framework which allows us to come to objective knowledge. Second, we have the regulative ideal of reason which directs us towards a progressive theoretical determination and unification of all aspects of reality – we make progress in science because that is what we actively *have to* search for. This is of course not an “explanation” that would satisfy the realist, which is only to be expected as

Kant denies all sense to such explanation by (what for him must be void) postulation. But seen from the transcendental perspective it does show something relevant about what it is about science that brings people to think that we need to, or even could give such explanation. It is this regulative ideal of reason, actually a projected rather than given unity (a *focus imaginarius*), which through hypostatization fosters the illusion that reality would actually exist as completely and systematically determined; and Kant's claim on the unavoidability of the transcendental illusions of reason can thus also explain the pull that realism continues to exert.

It is important to point out that Kant explicitly warns against interpreting this regulative ideal as merely a heuristic or methodological device (e.g. A653/B681 – a passage that serves as motto to Friedman 2001): without it we could not even coherently use the understanding, and thus would not be able to ascribe empirical truth to any statement (A651/B679). How this is to be understood in detail is at least as much a matter of contention as any other claim in the Kantian corpus, but I think we can at least distil the following. The understanding, which is the faculty that forms concepts reflecting objects does so with an eye to finding unity in experience (it is driven to do so by its own immanent norms, cf. the transcendental deduction), but in doing so it cannot look beyond what is actually given to it in actual perception (which is of course a form of the problem of induction). The faculty of reason must thus assist the understanding through its ability to integrate different concepts; and the only way that this can actually help us to go farther than what perception can directly teach is by effecting this integration under the regulative ideal of complete systematicity in nature (allowing only “projectible” predicates as concepts, by necessarily interpreting observed order as belonging to larger, ideal structures of order). There are two things of immediate importance to notice: the regulative ideal of reason is dependent on the constitutive principles of understanding for its operation; but it nevertheless has an a priori normative force of its own (as is the case for the constitutive principles, systematicity could never be empirically falsified) which is again grounded in the possibility of unitary experience (rather than in meta-physical beliefs about the “true” structure of the world), i.e. in the possibility to use our understanding in forming truth valued statements about an objective reality.

4. Reason's responsibility

In sketching this background to Friedman's denial of scientific realism, we have actually gathered the elements which will now allow us to diagnose what seems to be a fatal ambiguity

in his characterization of inter-paradigmatic rationality through the idea of a convergent series of constitutive frameworks, guided by a regulative ideal of reason. It is not that such an idea would commit him to realism, as his reference to the Kantian pedigree of his position is indeed enough to ward off such suspicion. It is rather the idea itself of having a series of *constitutive* principles being guided by a *regulative* ideal which does not make sense when seen from within this Kantian perspective.

We have seen that in Kant's system, the regulative ideal of reason is indeed genuinely normative – and thus a candidate to ground communicative rationality – but we have seen that it is also dependent on a constitutive framework *within* which the ideal can further guide the understanding. In Cassirer's modified version of Kantianism, which Friedman cites as the important source of inspiration for his own views, this dependency is inverted. The regulative ideal of systematicity takes absolute precedence, and a constitutive framework can only be defined as the set of principles which turn out to be preserved in the progress from theory to theory towards the projected complete system, and which as such are the conditions of every possible theory. But what about Friedman's proposal? At any stage of development in a mature physical theory we do have a set of constitutive principles which determine the space of reasons and allow us to form truth valued statements about an objective reality. Within any such framework it makes sense to project the ideal of nature as a fully unified system, but what kind of unified world could be “trans-historically” projected to span all these frameworks? Empirical truth makes sense only within a constitutive framework (it is this possibility which grounds the latter's special status), so speaking about the truth of the changing frameworks is precisely what cannot make sense (also according to Friedman (*ibid.*, 118)) – but as our exposition in § 3.2 should have made clear, this also takes away the ground under the transcendental justification of the normative status of the Kantian regulative ideal. The problem is of course that both Kant's and Cassirer's regulative ideal plays a role in (further) articulating what is ultimately *one* empirical space of reasons, and that it is not clear what sense it makes to speak about “convergence” of synthetic structures in the absence of such unitary empirical world – whether this would be interpreted realistically or transcendently. I already noted in § 2 that Friedman gestures towards an argument rather than really providing one when introducing his normative remarks on the convergence of frameworks. It now seems that he actually gestures in the wrong direction.

It is probably for this reason that, notwithstanding his stress on the convergence of frameworks, Friedman at one point speaks about “approximation to a final, ideal *community of inquiry*” (Friedman 2001, 64, my emphases). Such an ideal community would have

“achieved a universal, trans-historical communicative rationality on the basis of ... fully general and adequate constitutive principles” (ibid.), which is of course a description of what Kant believed he had achieved! I suggest that we thus better understand Friedman’s view on what secures the ultimate rationality of paradigm-change as a very serious call to uphold the full promises of the Kantian philosophical project itself – but as exactly that: a project. And this makes good sense of his appeal to the *necessary* role of philosophy in making possible revolutionary change, as this role is actually a modification of Kant’s original transcendental deduction into a necessarily inconclusive transcendental reflection. And as such reflection cannot but start from the (cognitive) situation we are in (as was already the case for Kant), it also makes good sense of the requirement that a new constitutive framework should *necessarily* evolve continuously from the older one. And finally, there is still a way in which we could see this as “progress”: we do actually *learn* something – not because we gain insight in the empirical structure of reality, and even less because we would gain insight in the transcendent structure of reality, but because we gain insight in how *we* do, and can, cognitively approach empirical reality.

In this sense paradigm-change is both less and more than communicatively rational. It is less, because it is not objectivity-constituting (which it for Habermas is, cf. the quote in Friedman 2001, 54; and which it also would have to be if it were linked to Kant’s or Cassirer’s regulative ideal); but it is more, because it does exemplify another deep feature of human rationality. We saw in § 3.1 how a transcendental perspective unavoidably brings with it a distinction between things as they appear and things as they are in themselves. This was of the utmost importance to the Kantian project in that it curbs the pretensions of reason, and thus actually confronts it with its own responsibility – which is to grasp its (cognitive) autonomy. And this finally brings us back to Friedman’s curious invocation of *responsibility* besides (communicative) rationality. What the possibility of paradigm-change, and the concomitant historicization of reason, shows is that the contingency that characterizes our approach to reality runs even deeper than Kant could see; but this only further highlights that we have no choice but to take full responsibility for it – which now also implies that we must be ready to question even our most cherished basic framework when situation demands it. And it is in this sense that we *should* see ourselves as approaching, and as necessarily *only* approaching, the ideal Kantian situation: because only thus can we fully uphold the ethical-cognitive ideal that is part and parcel of the transcendental idealist stance. Only thus can “reason grow increasingly self-conscious and thereby take responsibility for itself.”

REFERENCES

- Allison, Henry E. (2004), *Kant's Transcendental Idealism*. 2nd revised and enlarged edition. New Haven and London: Yale University Press.
- Friedman, Michael (1996), "Exorcising the Philosophical Tradition: Comments on John McDowell's *Mind and World*", *The Philosophical Review* 105: 427-267.
- Friedman, Michael (2001), *Dynamics of Reason*. Stanford: CSLI Publications.
- Kant, Immanuel (1781/1787), *Kritik der reinen Vernunft*. Herausgegeben von Wilhelm Weischedel (1974). Frankfurt an Main: Surhkamp Verlag.
- Longuenesse, Béatrice (1998), *Kant and the Capacity to Judge*. Princeton and Oxford: Princeton University Press.
- Psillos, Stathis (1999), *Scientific Realism. How Science Tracks Truth*. London and New York: Routledge.
- Slowik, Edward (200X), "Spacetime and Structuralism: Epistemological Realism or Relativized A Priorism?", manuscript available at PhilSciArchive.