Fictionalism and Illusion: Comments on Kraus

Katharina Kraus’ *Kant on Self-Knowledge and Self-Formation* provides us with a comprehensive account of self-knowledge, in all of its modalities, in Kant. Kraus provides us with the textual and conceptual resources to navigate Kant’s rather extensive vocabulary pertaining to self-knowledge—inner sense, inner perception, inner experience. This already marks out her book as an ambitious contribution to the existing literature, and her own original and provocative interpretations of these notions will provide an indispensable point of departure for future scholarship on the topic. This is not all Kraus accomplishes with this impressive volume, however. Along the way, she offers an innovative hylomorphic account of the faculties of the Kantian mind. She outlines a novel “expressivist” account of the “I think.” And she also elaborates a positive use of the idea of the soul that looks beyond its limited scientific application in a narrow psychological endeavour and assigns it a central role in the process of self-formation which constitutes a theme Kant actively explored in a variety of texts over the course of his career.

Clearly, there are a number of themes with respect to which one might fruitfully engage Kraus’ text. The one I will opt to pursue here is one that is broached in her fifth chapter, namely, the nature of reason’s idea of the soul in its application to inner appearances. This is a (surprisingly) broad theme, and there are a number of questions that might be distinguished within it, such as what does the idea of the soul represent (if anything)? is it true, or ‘truth-apt’? and *in virtue of what* is it able to do the work it does in guiding our investigation of inner appearances? While I think that Kraus and I disagree on what Kant’s answers are to most (if not all) of these questions, it should be said in advance that this disagreement takes place against a backdrop of considerable consensus between us. This includes (but is not limited to) the importance, even indispensability, of the idea of the soul for psychology, and more generally our commitment to taking seriously Kant’s account of reason’s positive use in the Appendix to the Dialectic, rather than regarding it as a regrettable outgrowth of certain outdated teleological assumptions on his part. Indeed, I believe that this background of consensus extends further than Kraus herself suggests in her discussion of my published views on the subject in the fifth chapter. In order to get a sense of this we’ll need to begin with Kraus’ initial distinction between “noumenalist” and “fictionalist” interpretations of the idea of the soul, to which she opposes her own novel “context of intelligibility” interpretation.

Kraus introduces the distinction between noumenalist and fictionalist views of the idea of the soul in the context of highlighting two key features of the idea of the soul for Kant. On the one hand the idea of the soul is “indispensably necessary” for the acquisition of inner experience, and on the other hand it has a “regulative use” with respect to inner appearances in that it projects a unity that guides the use of the understanding but falls short of determining a mental unity (cf. Kraus p. 172). According to Kraus, the noumenalist and fictionalist interpretation of the idea of the soul have in common that they take the regulative use of the idea to result in the generation of “descriptive statements about a given reality” (p. 187) which are either taken to be true or as truth-apt (so potentially false). The noumenalist view takes these statements to pertain to an underlying noumenal reality, to describe or involve the description of the soul as the sort of simple substance that was under discussion in the Paralogisms chapter. Given this apparent inconsistency with the negative part of Kant’s project, Kraus notes that this view is not well-represented in the literature. It has recently gained defenders (such as Wuerth and Proops) even if this is at the cost of diluting the significance of, for instance, the concept of substance at issue (and, to my mind, consequently squandering any wished-for continuity with a recognizably metaphysical project). The fictionalist view, by contrast, takes these descriptions as primarily of heuristic usefulness since it allows that they could turn out to be false and the idea of the soul on which they are based to be “an *untrue fiction* (or *mere illusion*)” (K p. 188). Kraus will depart from both of these views, since she takes the regulative use of the idea of the soul to issue in merely prescriptive rather than descriptive statements. But she also thinks that both fall afoul of the two features noted above: the noumenalist view in offering an presuming an overly (or illicitly) metaphysical view of the “indispensable necessity” requirement, and the fictionalist view in offering too weak a reading of the regulative feature to be effective.

I will return to Kraus’ criticism of the fictionalist view later (I will not return to a consideration of the noumenalist view, since I share her doubts about its plausibility). Before that, I think it’s important to get clear on what fictionalism is, at least in one of its variants, since I am among those charged with holding the view. I must say that I do not see my own interpretation reflected in Kraus’ description, though this is not perhaps for the reasons one would expect. Put simply, my own view is not that the idea of the soul *might* be an untrue fiction or *could turn out to be* an illusion, but that in fact it *is* a fiction, and (at least some) descriptive statements generated through it are *in fact* false. (I think ‘fictionalism’ might be an apt term for my view of the idea of the soul, when understood according to our ordinary understanding of ‘fiction,’ but it is rather more committed than the version canvassed by Kraus.) To avoid misunderstanding, but also to effect the contrast with Kraus’ own context-of-intelligibility view, it will be important to make a few distinctions here. First, and most importantly, we might distinguish between two sorts of descriptive statements that the idea of the soul can be said to generate: (i) those that pertain to or purport to determine the soul (‘the soul is a persisting substance,’ or ‘the soul is or has a fundamental power’) and (ii) those that pertain to inner appearances (claims relating to the causal relation between a comparatively fundamental power and a representation, without explicit reference to the soul). In light of this, we can conceive of a variety of sorts of fictionalism which concern one or the other, or both, of these sets of descriptive statements. So, one might be a fictionalist insofar as one takes descriptive statements about the soul itself as merely truth-apt or actually false; one might be a fictionalist insofar as one (also) takes statements of the second sort to be merely truth-apt or actually false. The most natural combination of these claims would be holding versions of the first *and* the second (perhaps ‘strong fictionalism’ is appropriate here), and my sense is that Kraus thinks that fictionalism one the first issue *implies* fictionalism on the second. But one can also conceive of hybrid forms, where one holds one but not the other. As it happens, my own view is (probably) best characterized as just such a hybrid as I take it that such descriptive claims *about the soul* are *in fact* false, but that resulting claims about inner appearances are, or can be, true (which I take it is as much as we can say about *any* putative cognition projected by reason before the understanding establishes its truth).

Kraus might welcome this view as ultimately ‘stranger than fictionalism’ in both relying on the *actual* falsehood of the idea of the soul at stake in the investigation of inner appearances and nonetheless allowing for the truth of the resulting cognitions. She might welcome it, that is, because it makes matters all the worse for my version. Kraus criticizes fictionalism inasmuch as “the possibility of falsehood makes any explanation of the principles’ [i.e., the ideas of reason] bindingness ineffective” (K, p. 191). Kraus considers that an adherent of fictionalism might cite the example of idealizations, or the utility of approximations, in science which involve “inaccurate or even false assumptions” in support of the claim that a possibly-false claim can yield true cognitions. Yet, she rejects the parity of cases here since “the existence of a noumenal thing that may not exist at all” cannot be close enough to a given thing to constitute such an approximation (K p. 192). Since my view maintains the outright *falsity* of this idea, and not just the *possibility* of its turning out to be false, it would appear to be worse off than other fictionalist alternatives.

And yet, I would contend that my version of fictionalism is better off than its cousins, and in a number of respects. First, I need to briefly explain my position whose essential difference from Kraus’ view (and from some other forms of fictionalism) consists in taking seriously Kant’s doctrine of transcendental illusion. As I read Kant, the idea of the soul is generated through reason’s effort to think the subject of every thought (the ‘I’ of the ‘I think’) as an absolute unity, not conditioned by the manifold of representation. The content of this idea, as the representation of the *unconditioned* subject of thinking, is something that cannot be encountered in experience and so is rightly designated as transcendent. Yet, through a mechanism that Kant does not describe in detail (but whose results he presupposes), that which is represented through the idea *appears* to us as given as an object, and specifically as the persisting object of inner sense. This is to say that the idea gives out the illusory appearance as if it is given to us immediately in inner sense, as a result of which we can determine it through the categories (as Kant had himself believed possible in his own pre-Critical rational psychology). Accordingly, while the object of the idea of the soul cannot be given in a possible empirical synthesis, it nonetheless appears as the persisting, and so phenomenal subject of inner sense. In any case, Kant’s exposure of this illusion constitutes the root of his critique of rational psychology, which traditionally had relied on just such a presupposition of immediate empirical access to the soul. However, the exposure of the illusion does not serve to dispel it entirely, as Kant notes that unlike the errors that it occasions, the underlying illusory seeming does not go away; rather, it is *natural* and *unavoidable* such that it “irremediably attaches to human reason [...] even after we have exposed the mirage” (A298/B354).

This account of illusion is important because it is precisely because of its harmful cognitive effects that Kant feels impelled to outline a positive use for the ideas of reason. For my own view, it is likewise important because my claim is that the idea of the soul is able to carry out this positive function *precisely in virtue of its illusory appearance*. There is no doubt that Kant regards the illusory appearance of the soul, as the persistent object of inner sense, as false and deceptive. The Deduction turns on the unavailability of an abiding subject of inner sense, and as containing the thought of absolute unity the soul *could not* actually be given in experience (though ascertaining this requires exposing the representation’s origin in reason’s activity rather than reflection on inner experience). There is also no doubt, at least as far as I’m concerned, that for Kant the ideas of reason are only able to contribute to our cognitive enterprise insofar as we take them in their illusory capacity. Kant makes this point generally concerning illusion in the following well-known passage concerning the *focus imaginarius*, also cited by Kraus, and indeed in connection with one of the features (of ‘indispensable necessity’) Kraus had identified as important above:

Now of course it is from this that there arises the deception, as if these lines of direction were shot out from an object lying outside the field of possible empirical cognition (just as objects are seen behind the surface of a mirror); yet this illusion (which can be prevented from deceiving) is nevertheless indispensably necessary if besides the objects before our eyes we want to see those that lie far in the background, i.e., when in our case, the understanding wants to go beyond every given experience (A644-5/B672-3; emphasis mine)

The point I want to emphasize here is that the ideas make possible the discovery of further cognition, i.e., permit of a positive use, precisely in virtue of “illusion,” precisely because of their illusory appearance as objects given in their own right. As Kant later claims, the object of experience is as it were “derived from the imagined object of this idea as its ground or cause” (A670/B698—my emphasis). In the case of the soul, then, inner appearances only permit of being brought into systematic connection, and further appearances cognized that are not immediately given, through the illusory appearance that the soul is given as an object of inner sense.

Returning to Kraus’ criticism of the fictionalist view, the indispensability of illusion when it comes to the systematization and discovery of inner appearances suggests that Kant himself thinks that its possible to gain true cognition from (what are ultimately) false principles. Another passage that says much the same thing, and this with respect to an explicitly psychological theme, follows soon after. So, after Kant had claimed that the idea of an absolutely fundamental power only “pretends to objective reality” (A650/B678), he goes on to emphasize the indispensability of this *pretension* for any attempt to introduce systematic unity into the empirically-discovered faculties and powers of the mind. I don’t know whether this is a fact that makes (what I take to be) Kant’s view more or less appealing *philosophically* (and to be honest, my interest in outlining it consisted in its utility in accounting for the errors of Kant’s predecessors). But a version of Kant’s own example here might support its intuitive plausibility. We might imagine seeing a mirror at the end of a sidewalk set up at a 45 degree angle so that when we are walking along the sidewalk we can see our friend approaching from around the corner. The illusory appearance here is that our friend is ahead of us and approaching us directly. If this is all artfully concealed, we might fall prey to this illusion, but given sufficient experience (through which the understanding is instructed how to use the illusory reflection to judge regarding what it represents) we could infer what is really the case from this (false) appearance, namely, that our friend is actually around the corner, and even accurately gauge his distance, pace, etc. on its basis. Indeed, this example might also be instructive in mounting a reply to Kraus’ criticism that fictionalism cannot draw on the scientific utility of approximations to buttress its case. In my version, the “approximation” of the illusory seeming of the idea of the soul is not to an underlying *noumenal* reality. It is, rather, to something (putatively) given in inner sense, as a *persisting* subject of thought (and which as a matter of fact Kant and many others actually *took* to be so given), just as the illusion through the mirror is something that could very well be taken as given in experience (and of course this only enhances its deceptive potential).

As for Kraus’ last objection to the fictionalist view—that it fails to account for the bindingness of reason’s principles with respect to the investigation of inner appearances—my reply is, as far as I can see, much the same as hers. The prescriptive force of the idea of the soul lies in the fact that without it we cannot undertake any coherent investigation of the domain of the purely inner. Without an idea of the soul, there is no way in which we can make a principled distinction between empirically inner and outer appearances such that we can investigate the former wholly apart from the latter. In this way, we might conceive of the prescriptiveness of the idea of the soul along the lines of a commitment we undertake as a ‘cost of doing business’ in such an enterprise, where we need to secure the “integrity of the mental” against, for instance, the incursion of “windy hypotheses” through materialistic reductions of psychical phenomena to physical ones. This looks to me much like the normativity assigned to the idea of the soul in Kraus’ “context of intelligibility” view, that is, where the idea “must be presupposed to set a stage [...] within which a certain *kind of cognition* is first intelligible” (K p. 172).

Yet, all of these considerations are circling the deeper difference between my fictionalism and Kraus’ context of intelligibility view. This, I take it, comes down to an interpretative disagreement regarding the importance of the doctrine of transcendental illusion for understanding the claims of the Dialectic and its Appendix. It is perhaps reflective of the priority I assign to Kant’s claim of the naturalness and unavoidability of transcendental illusion, that I was surprised to see Kraus characterize the possibility that the soul is a “mere illusion [...] to which no real entity corresponds” (K. p. 188) as a *problem* with a putative interpretation of the Dialectic rather than, as I had presumed, an indispensable starting point. In any case, it is this commitment to the interpretative priority of the doctrine of illusion that underlies the key differences between Kraus and I. For me, it is *in virtue of* its illusory appearance that the idea of the soul *represents* the persistent subject of inner sense, and while this representation is false, it is indispensable for gaining further (true) cognition of inner appearances and for securing the integrity of the mental in which such a systematic investigation is possible in the first place. Kraus, by contrast, does not take the idea as a representation at all but as a mere *presentation* (*Darstellung*), and claims that it is possible for it to do the work we both assign to it in this presentational capacity.

But if this is so, and here’s my principal question for Kraus, what happens to the doctrine of illusion in the Dialectic? Kraus provides scant mention of the doctrine in the course of her discussion, and even then it is not rigorously distinguished from the unnatural and avoidable errors consequent on it (K pp. 166, 190), and so I’d be keen to have her elaborate on her own views about the nature and importance of the doctrine for Kant. Looking at Kraus’ text, one possible interpretation of transcendental illusion might be readily available, in connection with her presentational interpretation of the idea of the soul. As she writes,

the act of *presupposing*  the object described by an idea should be understood as *an a priori* act of *presentation*. In presenting, or exhibiting such an object to the mind, reason produces procedural rules for how to seek systematic unity with respect to given empirical cognitions. (K p. 200)

As Kraus goes on to emphasize, this presentation does not “amount to the assertion of a truth-apt claim about some given object, not even a hypothetical claim, since this would inevitably and illegitimately amount to the hypostatization of the idea” (Ibid.). While Kraus does not make this connection explicit, though it is suggested by the mention of ‘hypostatization’ here, she might very well take transcendental illusion to consist in the appearance of what is in fact a mere *presentation* as a *representation* of the idea of the soul (i.e., as an object). That Kraus takes illusion in this way would explain one difference between her reading and my own—that on her view the work accomplished by the idea of the soul is done *apart* from its illusory appearance (here, as a *representation*) rather than in virtue of it, as is the case in my view, which again only serves to enhance the temptation posed.

Even so, I harbour doubts about whether such an account (if it is Kraus’ or not) can live up to the ‘naturalness’ and ‘unavoidability’ criteria that Kant clearly sets for properly *transcendental* illusion. On the ‘naturalness’ count, I have difficulty conceiving of how a presentation conceived of merely as a “procedural rule” could give rise to being *naturally* mistaken for a representation of an object. Perhaps the naturalness of such an appearance follows more readily from conceiving of the idea as a “model, sketch, outline, or pattern” (K p. 198) though that these are taken to present a “transcendental thing” rather than a real one makes any such appearance less natural. Such an account fares worse, I think, on the ‘unavoidability’ point since there is little that is enduringly tempting in such a conflation—having been alerted to the merely presentational character of the idea of the soul, why would the metaphysician still be tempted to take it as something more, especially given the fact that the idea can perform its important regulative function apart from any such illusory appearance? What does such a transcendental illusion have in common with empirical illusions, which continue to incite us to deception even after they have been exposed as mirages?

But this is all based on conjecture on my part, and I thus look forward to Kraus’ explanation of her take on the nature and role of illusion in Kant’s Dialectic. I do think that, in the end, it is important not to marginalize this doctrine or defang it such that the errors it occasions lose all conceivable appeal for the metaphysician. For, as Kant memorably writes, the illusions fostered by these ideas are sophistries “of pure reason itself, and even the wisest of all human beings cannot get free of them; perhaps after much effort he may guard himself from error, but he can never be wholly rid of the illusion, which ceaselessly teases and mocks him” (A339/B397). The bar Kant sets for transcendental illusion is thus rather high, which is only befitting its central role in accounting for the errors of traditional metaphysics that Kant himself once fell prey to. Indeed, it seems, he never completely put the threat they posed behind him.

Corey W. Dyck