Synthese (2010) 177:111–138
DOI 10.1007/s11229-009-9642-5

The No-Miracles Argument, reliabilism, and a methodological version of the generality problem

Mark Newman

Received: 29 July 2008 / Accepted: 10 July 2009 / Published online: 25 July 2009

Abstract The No-Miracles Argument (NMA) is often used to support scientific realism. We can formulate this argument as an inference to the best explanation (IBE), but doing so leads to the worry that it is viciously circular. Realists have responded to this accusation of circularity by appealing to reliabilism, an externalist epistemology. In this paper I argue that this retreat fails. Reliabilism suffers from a potentially devastating difficulty known as the Generality Problem and attempts to solve this problem require adopting both epistemic and metaphysical assumptions regarding local scientific theories. Although the externalist can happily adopt the former, if he adopts the latter then the Generality Problem arises again, but now at the level of scientific methodology. Answering this new version of the Generality Problem is impossible for the scientific realist without making the important further assumption that there exists the possibility of a *unique* rule of IBE. Doing this however would make the NMA viciously premise circular.

Keywords Scientific realism · Reliabilism · No Miracles Argument · Inference to the best explanation · Generality problem

1 The No Miracles Argument in reliabilist form

We should start by defining 'Scientific Realism'. I will follow Psillos' (2006) articulation, which indicates that there are three claims constitutive of scientific realism:[1]

[1] One should perhaps add two more theses: Methodological and Axiological. The former states that scientific method is the best *means* for deriving knowledge about the world, the latter states that truth should be the *goal* of scientific theorizing. These are slightly more contentious, so I leave them out of the definition here.

M. Newman (✉)
University of Minnesota, Duluth, MN, USA
e-mail: mnewman1@d.umn.edu

Semantic thesis: Scientific theories are truth-conditioned descriptions of their intended domain. They are capable of being true or false.
Metaphysical thesis: The world has a definite and mind-independent structure.
Epistemic thesis: Mature and predictively successful scientific theories are well-confirmed and approximately true. So entities posited by them, or ones similar, inhabit the world.

The No-Miracles Argument (NMA) has been formulated by many philosophers as an abductive defense of the *Epistemic Thesis*.[2] The argument goes as follows: Science has historically been very successful, and the only adequate explanation for this success is that our scientific theories are approximately true.[3]

But how *precisely* does the NMA support the *Epistemic Thesis*? Well, by taking the NMA to be an inference to the best explanation (IBE), and by arguing that IBE is a reliable rule of inference, the realist can claim to be using a reliable abductive rule to conclude that most successful theories are true. However, when interpreting the NMA as an IBE it has been suggested that the realist assumes the very thing he wishes to prove—he *uses* IBE to conclude scientific inference is reliable, but this rule is itself the very rule science uses. This is clearly circular.[4]

Some have responded that this circularity is not vicious because it is perfectly legitimate on externalist epistemology to use a *rule* of inference in the process of arguing for the reliability of that very same rule—it is just unacceptable to assume a *premise* that one wishes to prove.[5]

To understand this claim we need to look at how rule circularity supposedly avoids viciousness. The details of this strategy are found in Black (1958), Braithwaite (1953), and Psillos (1999). In an attempt to provide an inductive justification for induction, it has been argued that premise circular arguments are viciously circular because they appeal to reasons (premises) for accepting a conclusion where one of those reasons just is the conclusion itself. This is where one presupposes what one wishes to prove. On the other hand, rule circular arguments don't include as a premise the conclusion to which they lead. A rule circular argument concludes, C, that some rule of inference R, is justified, and that rule itself is used to move from premises, $P_1, \ldots, P_n$ to conclusion, C. This form of circularity is claimed by Braithwaite (1953), Van Cleve (1984), Papineau (1993), and Psillos (1999), to avoid vicious circularity. For example, if we consider the Straight Rule of Induction, on this account it may be justified in a rule circular manner by the following reasoning: Let R_1 represent the rule: 'If most instances of A's examined in a wide variety of conditions have been B, then conclude

2 See for example Popper (1963); Smart (1963), Putnam (1975, 1978), Boyd (1984), Leplin (1997), Bird (1998), and Psillos (1999, 2006).

3 Realists often add caveats such as that this 'approximate truth' need only refer to those specific components of our theories actually responsible for the novel success of the respective scientific theories, or that the sense of truth being used is our everyday correspondence notion.

4 Arguments along these lines were first made by Fine (1986), and (Musgrave 1985, 1988). Further concerns regarding the rule circularity if IBE have also been raised by Douven (2005) and Lipton (2004).

5 This has been argued in various forms by Boyd (1984), Peacocke (1986), Papineau (1987), Bird (1998), Psillos (1999, 2006), Nola and Sankey (2007). See Busch (2008) for a response to Psillos' formulation in particular.

(probably) the next A to be encountered will be a B.' Now we argue: R_1 has usually been successful in the past, therefore, probably, R_1 will be successful in the next instance. Here we have a case of using R_1 to derive the conclusion that R_1 itself is reliable.

We can do the same thing for IBE: Let us represent IBE by R_2: 'If H explains a set of surprising data D, better than any other hypothesis, then infer that H is probably true.' Now we argue the following. (D): Science is remarkably successful. (H): it would have to be a miracle for science to be so successful if it were false (i.e. no competing hypothesis better explains the success of science). Therefore, probably, our scientific theories are true.

Here we have a case of using R_2 to derive the conclusion of this argument. Notice that R_2 is not a premise of the argument, just as R_1 was not a premise of the previous argument. The R's are being used to derive a conclusion. In the first argument the conclusion explicitly states the reliability of R_1 itself. In the second argument the conclusion is that our scientific theories are probably true. Notice that this seems to differ from the claim that R_2 is itself a reliable rule of inference. However, since R_2 is itself used by scientists in generating the surprising data D, we can infer that R_2 is a reliable rule of inference since it reliably generates successful scientific theories.

We don't yet have quite the relevant formulation though, since there is a formulation given by Richard Boyd (and Psillos) which recommends we should accept the NMA as a rule circular argument on the grounds that it is being read as a *two step* argument, rather than a single step inference as in the above form. On this more complex form of the argument the first inferential step looks to scientific uses of IBE and concludes that our best, most successful theories are derived by reliable ampliative rules of inference (IBE's). The second step uses IBE to infer from this prior fact to the conclusion that IBE is reliable at generating true theories. IBE is not included as a premise in the argument, but is used to move to the conclusion, which asserts its own reliability. Here is a statement of the argument[6] (Boyd/Psillos NMA):

1. The instrumental success of science is remarkable.
2. The best explanation of this success is that the methods of science (IBE) are reliable methods of inquiry.
3. The methods of inquiry derive from and rely upon background theories that we accept based on their success.
4. The best explanation of the reliability of our methods (IBE) is therefore that these background theories are approximately true.
5. Therefore, the best explanation of the instrumental success of science is the approximate truth of our successful theories.

But despite this reformulation of the argument in rule circular form, we still haven't really got to the heart of why we should think rule circularity is not vicious. After all, doesn't one require independent justification for the truth of an inference rule before one can use it in an argument, no matter what the conclusion? Psillos responds in typical externalist fashion:

[6] This form of the NMA can be found in Boyd (1981; 1996, p. 222), Papineau (1993), Bird (1998), Psillos (1999, pp. 78–81), Nola and Sankey (2007, pp. 349–350).

> When an instance of a rule is offered as the link between a set of (true) premises and a conclusion, what matters for the correctness of the conclusion is whether or not the rule *is* reliable that is, whether or not the contingent assumptions which are required to be in place in order for the rule to be reliable *are* in fact in place. If the rule of inference *is* reliable (this being an objective property of the rule) then, given true premises, the conclusion will also be true (or, better, likely to be true—if the rule is ampliative). Any assumptions that need to be made *about* the reliability of the rule of inference, be they implicit or explicit, do not matter for the correctness of the conclusion. Hence, their defence is not necessary for the *correctness* of the conclusion. (1999, p. 83)

An objector might respond by arguing that surely our awareness of the epistemic status of the rule *is* important—the rule must be reasonably believed to be reliable for us to justify taking the conclusion to be correct. In fact, Psillos isolates this issue as,

> The point on which the allegedly vicious nature of rule-circularity turns. For whether or not the proof of reliability is required for justification will most likely depend on the epistemological perspective which one adopts. As is well known, *externalist* accounts sever the alleged link between being justified in using a reliable rule of inference and knowing, or having reasons to believe, that this rule is reliable. (1999, p. 84)

And indeed this may be the case. Many philosophers of science, like Psillos, are willing to adopt an externalist epistemology when considering our claims to scientific knowledge. It seems too demanding, they suggest, to require that we investigate and have transparent reasons to justify the use of every rule of inference before we conclude what already seems obvious, that our methodological rules are reliable at generating success. When you have a system of reliable output generators, such as our background theories being used to generate reliably correct predictions, why further insist that we must possess exhaustive justification for each and every rule, such as IBE, before concluding that they are reliable producers of approximately true outputs?

There have been long and detailed debates in the epistemology literature concerning this point. And in fact it is interesting that the debate between internalists and externalists should have such an impact on our more localized debate over scientific realism. It appears that those who would defend scientific realism by appeal to the NMA must adopt an externalist epistemology in order to retrieve it from vicious circularity. Those who reject scientific realism may not have to adopt internalism if there are other arguments that can defeat realism, but at the least if the debate is over the NMA they must adopt internalism.[7]

[7] Although see (Hudson, 2004, pp. 193–211) for an approach that attempts to synthesize internalism and reliabilism in a scientifically informed manner.

2 Reliabilism and the generality problem

The form of externalism appealed to on the reading of the NMA above is one known as Reliabilism. It is our task to establish not whether reliabilism itself is a thesis worthy of adopting for our epistemology, but rather that *if we do* adopt reliabilism, on what grounds is it so adopted, and are those grounds in conflict with the scientific realist reading of the NMA? Specifically, I will argue that a defense of Scientific Realism that depends on an externalist reading of the NMA (one that treats it as an IBE), must presuppose an answer to the Generality Problem. However, an answer to the Generality Problem on externalist grounds requires commitment to realism about some particular science of the mind, and this itself requires presupposing either IBE or scientific realism. The reliabilist won't have a problem with assuming IBE is reliable even though he may be ignorant of that fact, but he does run into trouble with one other assumption he must make, namely that IBE is a *type* of inference which is even in principle *uniquely specifiable*. I will argue that making this *metaphysical* (rather than epistemic) assumption entails his reading of the NMA is not just circular, but *viciously* circular.

The first task then is to specify a fair characterization of reliabilism. This is a position that suggests we have reasonable justification in adopting a belief if that belief is formed as the result of a reliable connection with the truth. Like Scientific Realism, Reliabilism comes in a variety of formulations. Some appeal to the indicated connection between inputs and outputs, which make beliefs reliably formed, as being due to the laws of nature (Armstrong 1973). Some argue the connection is counterfactual (Nozick 1981). Others take the connection to be a causal process or a method that wouldn't generate true output beliefs if they did not actually hold. For example, I would not come to form the belief that there is a cat in front of me unless I was caused to believe this fact by natural laws (nomologically); unless it were true; or unless there were a process or mechanism causing me to believe it. Importantly, these connections, whatever they may be, between my beliefs and the world, are external to my collective background beliefs in that not only must they be true, but also it is not necessary that I be *aware* that the relevant connections hold in order that my beliefs be justified. The most popular form of reliabilist approach, (and the kind adopted by realists in their reading of the NMA), is a *process* reliabilist account (due to Goldman amongst others).[8] On this account, a belief is justified if and only if it is produced by a reliable cognitive process—one that has a high enough number of true beliefs as output in proportion to true inputs.[9]

Because this externalist account of justification does not actually require the subject be aware that he is using a reliable cognitive process to generate his beliefs, the

[8] Perhaps not all realists adopt this epistemology, but it is at least to be found in Psillos (1999), Bird (1998), Nola and Sankey (2007), Papineau (1993).

[9] This kind of reliabilist account which takes the relevant extent of reliable processes to be limited to the *internal* cognitive states of our subject, may superficially avoid the difficulties of establishing reliable *scientific* methods. However, limiting reliability constraints only to those in our bodies will still generate the 'embarrassment of riches' as Alston calls the Generality Problem below. Notice that this account need not commit to actual processes, but can perfectly well be given a counterfactual or modal interpretation. Nothing of what follows hangs on such a choice.

reliabilist is apparently able to circumvent several important problems in epistemology, the most notable of which is how we can have justified beliefs at all. The skeptic challenges us to justify how inductive knowledge is even possible given that we will have to appeal to an inductive argument to justify induction. The externalist avoids this circularity problem by adopting the position that our subject not actually be required, even in principle, to *defend* his claim that any belief he holds is justified. The only requirement is that the subject *in fact be* justified because his belief forming process stands in a reliable connection to the facts that constitute its content. So, for example, on the reliabilist account, to be justified in believing that there is a cat in front of me I need not be capable of giving good reasons for why I believe it. All that need be the case is that I am using a reliable cognitive process to generate the belief, given that the input to that process is indicating there is a cat in front of me.[10] A couple of well known accounts include Goldman's:

(G) If S's believing *p* at *t* results from a reliable cognitive belief-forming process…then S's belief in *p* at *t* is justified. (1986, p. 347)

And Alston's is very similar but specifies what *type* of relevant cognitive belief-forming process is in question:

(A) The relevant type for any process token is the natural psychological kind corresponding to the function that is actually operative in the formation of the belief.[11]

Of course accounts of knowledge which reject the need for a subject to give an account of how his beliefs are justified may appear counterintuitive to begin with, but they do provide a very convenient means of answering not only the problem of induction, but also concerns one may have about internalism, where it seems overly demanding to require a subject be able, even in principle, to provide a full account of the reasons for his beliefs. After all, it would appear uncharitable to accuse someone of failing to be justified in believing a scientific law, such as the Ideal Gas Law, merely on the grounds that they either fail to justify where they obtained the belief, or they can't provide a step-by-step account of why their source is reliable. The reliabilist therefore, has at hand a theory apparently amenable to solving significant problems in epistemology.

There is however a serious problem faced by reliabilist accounts, to which they have as yet failed to provide an adequate response.[12] The problem they face is known as 'The Generality Problem' and it can be characterized in the following way: each

[10] Some reliabilists also require we not have reasons *against* using the process in question. This additional requirement might leave the scientific realist subject to a pessimistic induction from the history of science—surely if some successful science has been wrong in the past then we have good reasons to believe IBE is not reliable after all. We see however that modern scientific realists work very hard to specify just which parts of our best theories are truly responsible for their success, and they tie this to IBE while casting aside the erroneous entities, laws, processes, etc. Entity realism, structural realism, and work by Kitcher (1993) and Psillos (1999) all point to this strategy, which nicely supports reliabilism.

[11] Quoted in Conee and Feldman (1998, p. 377)

[12] Note that Alston's definition here is actually *in response* to the following problem, but I will suggest it fails to solve the problem by following concerns raised by Conee and Feldman (1998).

token psychological process that leads to a belief will be an instance of potentially many different *types* of process, and it is only these types that can be assessed for reliability, since singular token instances cannot be more or less reliable. The trouble is that there is no *unique* psychological process type that describes any given process since there are indefinitely many psychological process types for any given token instance. Even worse, these different types almost invariably differ in their reliability at producing true beliefs. This entails that without an account of precisely which types of belief-generating processes are relevant to producing any given belief, there is no way to evaluate the process for reliability. Thus, the reliabilist requires a relevance rule for attributing reliability, but because each token process is multiply describable under a multitude of types, there is no possibility of providing such a rule.

Examples of this problem typically appeal to the manner in which we might describe a simple case of belief formation from a visual experience. I come to believe, for instance, that there is a cat in front of me in virtue of a specific token process. Is this process to be described as a 'visual process', a 'cognitive process', a 'cognitive process occurring on some particular Sunday evening in London', the 'visual process occurring at 7:36 pm on a Sunday in mid-January', etc.? My token belief forming process fits all of these types, but which one is the appropriately reliable type that we are to evaluate?

In fact, following Feldman (1985) we can divide this problem further. One element he calls the 'Single Case Problem' arises when we give such a narrow account of the relevant psychological process that there is only one instance of the relevant type, and that is the very token we are talking about to begin with. In this case a true belief will be the result of a fully reliable process (that process, having only occurred once, must be fully reliable if its reliability is a ratio of input to output truth values). If the single case results in a false belief, then the process must then be fully unreliable. Yet, this doesn't capture what might actually be the case, since even reliable processes can sometimes turn up false beliefs.[13]

The other element of the Generality Problem is the 'No-Distinction Problem' and this arises when our description of the relevant type is so broad that we can develop beliefs of problematically diverse reliability even though they fall under the same type. This situation arises because the beliefs are formed by tokens that have very different degrees of reliability despite being of the same type. The problem when we combine these two strands of the Generality Problem is then to find a type of psychological process that is not too narrow and not too broad.

Conee and Feldman (1998) have analyzed various attempted solutions to this most critical problem for the reliabilist. My concern here is not to evaluate their critique, but rather to suggest that because the reliabilist must appeal to some form of 'local' realism[14] about psychology or cognitive science or neuroscience (as Alston as well

[13] Note that Goldman (1986) and **?** suggest that we can solve the Single Case Problem by appealing to a dispositional account of types, rather than to a frequency account. It is not important for the argument of this paper whether their approach succeeds.

[14] I am using the terminology 'local' realism to designate a realist attitude towards *some* particular scientific theory based on scientific evidence, which is to be contrasted with 'global' realism which indicates the philosophical thesis of the scientific realist supported by the NMA: realism towards *any* particular scientific theory.

as Conee and Feldman recognize) he cannot use reliabilism to support his preferred reading of the NMA.

3 Reliabilist defenses of NMA viciously presuppose scientific realism or IBE

Thus far we have seen that the scientific realist who appeals to an IBE reading of the NMA must adopt an externalist epistemology. The preferred version adopted by most realists is known as reliabilism, but this account appears to suffer from the Generality Problem. In this section I will briefly look at proposed solutions to the Generality Problem and argue that they either presuppose scientific realism or IBE. In either case the realist is caught in a vicious circle because the particular assumptions being made are no longer merely epistemic, but are metaphysical—and this makes the IBE reading of the NMA premise circular rather than rule circular.

3.1 Proposed solutions to the generality problem

It has been supposed that a solution to the Generality Problem requires the reliabilist to provide an account of reliable belief forming processes that narrow down the relevant types of cognitive processes to such a degree that it avoids the No Distinction Problem, yet does not fall into the Single Case Problem. Depending on how one characterizes the cognitive process, the account will need to appeal to some way of carving-up the mind's processes into specific and distinct mechanisms of belief formation that are of a single type. So, when I believe that I see a cat in front of me, no matter how I specify the mechanism that generates that belief, it must be a singular description.

We have already seen one attempt at providing the appropriate relevance rule for reliabilism. Alston suggests that the way to select the relevant process type for any given token of belief formation is to look for the *natural psychological kind* corresponding to the function actually operative in the generation of the belief. This suggests that there is only one psychologically real type of belief forming process in any given token instance of belief formation. What is this real psychological type? It is that which *functionally operates* for the given input and output. This means that for any given input (sense perception for example), there is only one real function that operates to generate the specific belief output. For example, when looking at my cat, the Generality Problem raises the difficulty of multiple types for my particular token sensory experience. I might be looking at the cat as a whole, or maybe experiencing just its face, or its legs, or only its fur. The input is describable under many different types of process. What Alston suggests is that there is only one actually operative type in any given instance of my forming the belief that there is a cat in front of me. There is only one real type of input operated on to generate the output (the belief that there is a cat in front of me).

Alston's suggestion is a potential solution to the Generality Problem because it specifies a reliable process for each token of belief formation. The process is sufficiently narrow as to avoid the No Distinction Problem (there is only one type being instantiated), and is sufficiently broad as to avoid the single case problem (we are still dealing with a natural kind type, not a single instance). However, Conee and Feldman point out that even for one specific input there are multiple functions that could generate the specific output required. This reintroduces the No-Distinction Problem.[15] One simple way of illustrating this point is by noting that just as my input may be sensory information about a cat, the processing of that data may still be about a cat but include other sensory cues, or background inputs, such as my beliefs about cats. The original output and input pair will remain the same, but there may be multiple other processes functionally operative. An analogy might help. Take as an input and output pair the ordered set $\{2, 8\}$. The operative function could be (x^3), or $(x + x + x + x)$, or $8[x - (x - 1)]$, or,…There are an indefinite number of possible functions for this input/output pair, so what use is it to suggest that we are able to specify the correct function under our natural psychological kinds?

An answer to this problem requires, as Alston acknowledges, a strong version of psychological realism, one that assumes there is only one, *unique* natural kind function for each and every token of belief formation:

> The viability of a reliabilist theory of justification or knowledge hangs on the viability of psychological realism. If there is not an objective fact of the matter as to what input-output function is utilized in a given belief formation, then reliabilists are helpless before the Problem of Generality, and they may as well pack up their bags and go home….According to my psychological realism, exactly one of those possibilities is realized in this case. And whichever one is realized, it is the reliability of that function (or of the correlated mechanism or process) that is crucial for the epistemic status of the belief. (1995, p. 365)

And here is where we first run across a potential circularity problem for reliabilist IBE readings of the NMA. If, as suggested here, a defense of reliabilism requires psychological realism, then how can the reliabilist who is a scientific realist presume to defend realism generally without begging the question?[16] That is, by assuming local realism regarding psychology isn't the realist leaning on the reliability of IBE (as used in psychology to determine cognitive processes) in order to defend IBE's operation at the higher level of scientific theories more generally? The concern is that defending reliabilism from the Generality Problem requires local scientific realism about psychology, but to secure this local realism requires the reliability of IBE—the very rule in question. The only other alternative for securing local realism about psychology would seem to be an appeal to the global philosophical thesis of scientific realism in the first place, and this would clearly be premise circular (assuming scientific realism to justify local psychological realism, and then using that to justify scientific realism).[17]

---

15 As mentioned in footnote 12, Alston thinks he has a solution to the Single Case Problem, but that doesn't affect this argument.

16 Note that although Alston opts for cognitive psychology, one's local realism might rather be about predicates and relations from other sciences, or even common sense for that matter. Any good naturalized defense of realism will opt for some scientific account, but whichever it is, there still remains the assumption of local realism. I shall in what follows use cognitive psychology as an example, but the points I make will work if one is instead a realist only about, for example, neuroscience.

17 Conee and Feldman recognize that there is a problem for reliabilists who lean heavily on a local realism of one form or another, and consider Ralph Baergen's appeal to IBE as a reliabilist response to the problems facing Alston. They suggest IBE is problematic for reliabilists only on purely pragmatic grounds. What

But the reliabilist has a plausible answer to this supposed problem: externalist epistemology does not need to provide any justification for the formation processes involved in generating beliefs. So long as a reliable process *is in fact* being used, then the belief is justified. If a token instance of belief formation really does use a unique natural kind type of reliable psychological process, then the belief is justified. Who cares if we cannot specify the precise relevance rule for that type? If it is instantiated, then we are justified.

And this is where things get interesting. To appreciate the reliabilist's move here we must highlight an obvious distinction; epistemic assumptions versus metaphysical assumptions. The reliabilist is being accused of making the epistemic assumption that psychological realism is correct—that we know there is a unique natural-kind-belief-forming-process-type for each and every token. The reliabilist responds by pointing out his position does not require justification for this epistemic assumption, it only needs the world to cooperate in the right way—for it to be true that there is a unique natural-kind-belief-forming-process-type for each and every token. This latter assumption is metaphysical, not epistemological. Only internalists require both assumptions to be justified.

3.2 Trouble for realists

So, the reliabilist apparently has an adequate response to the Generality Problem: although we may not be aware of it, so long as our beliefs are actually formed by reliable processes delineated by the natural kinds found in psychology, then we are justified in our beliefs.

What I want to do now is argue that things are not so simple for the reliabilist. In particular, it is in the *metaphysical* assumption that trouble arises for the scientific realist. What is the metaphysical assumption being made by reliabilists? Well, there are many (that there is an external world is clearly one example), but the assumption essential to the externalist response to the Generality Problem is this: it is metaphysically *possible* that there is a unique natural-kind-type-cognitive-process which is reliable and is instantiated by a token process of belief formation.

Now this metaphysical assumption seems *prima facie* harmless. It seems merely to be stating that it is possible there is a cognitive process being used in any token of belief formation. This would in fact seem to be necessary—how could we explain belief formation in the absence of some sort of mechanism? Something has to be going on after all. But the problem lies in the assumption that this process, whatever it really is, has to be of the *natural kind* variety. If one is a scientific realist adopting a naturalist approach to epistemology, as is reliabilism, then one is making the assumption that whatever science ends up being correct in describing the types of cognitive processes involved in forming beliefs, that science will have predicates describing a set of natural

Footnote 17 continued
I want to show is that the problems for the reliabilist are far more serious—they are led into vicious premise circularity.

kinds. These terms will carve up the world differently in different sciences, but all will nevertheless describe the natural kinds in the world.

That we can rely upon our best sciences to carve up the world into natural kinds entails that the metaphysical thesis adopted by reliabilists requires a slight, but important, amendment. This amendment is that not only is it metaphysically *possible* that the processes we use to generate beliefs fall under natural kind descriptions, but that it is also metaphysically *actual*. That is, the scientific realist must adopt a form of reliabilism that assumes our cognitive belief forming processes are actually real natural kind types. The metaphysical assumption is now the following: it is metaphysically *actual* that there is a unique natural-kind-belief-forming-process-type that is reliable and is instantiated by a token process of belief formation.

This refinement to the metaphysical assumption is less acceptable than its predecessor. There are at least two reasons for this. For the reliabilist more generally, it is still an open question whether we need accept that the world must have a natural kind structure, regardless of how our best scientific theories currently describe it. There is no logical incoherence in the thought that the world is not carved into natural kind types. On the other hand, whether this is metaphysically possible is going to depend on how one construes 'metaphysical possibility'. Assuming this notion captures all that is possible according to the laws of nature, the reliabilist still needs to secure the claim that the laws of nature exhaust all possible descriptions of the events in the world. If there is room outside the laws of nature for events to occur, then it is metaphysically possible that the world does not have exclusively a natural kind structure. I won't pursue this line of thought further, since there is a much larger problem lurking.

The problem I refer to is simply that for the reliabilist to justify his assumption that the world actually does have a natural kind structure (regardless if science ever actually discovers it or not) he is leaning on an inferential rule. If the reliabilist is a scientific realist then that rule is IBE. After all, the scientific realist we are addressing is a naturalist and sees IBE as a reliable rule of inference in science. Naturalists typically think we should adopt the methods of science as best we can to develop new knowledge. Therefore, our scientific realist will want to adopt IBE as his rule of inference for even metaphysical assumptions. This particular instance of the rule might for example say that our best theories indicate that the world has a natural kind structure. These theories provide the best explanation for phenomena we observe around us every day, as well as in the laboratory. Therefore, we should accept that the world has a natural kind structure.

If this is the case, then the NMA is in trouble. The reason is this: by appealing to IBE to justify the metaphysical assumption that the world actually does have a natural kind structure, the realist faces something I will call the Methodological Generality Problem (MGP).[18] The problem is similar to the previous generality issue, but now applied to methodological processes, such as IBE, instead of cognitive processes. Here's the problem in a form that mimics the original formulation of the Generality Problem:

18 Note that this is not a simple accusation of rule-circularity regarding IBE. If one were to level this claim at the realist, he'd just respond as before—that we are justified in using a rule, if it is reliable.

(MGP): Since there are indefinitely many IBE process types for any given token instance, there is no *unique* IBE process type that describes any given process of inference.[19]

Even worse, these different types almost invariably differ in their reliability at producing true beliefs. This entails that without an account of precisely which types of IBE processes are relevant to generating any given belief we might have, there is no way to evaluate the process for reliability. Thus, the reliabilist requires a relevance rule for attributing reliability, but because each token of IBE is multiply describable under a multitude of types, there is no possibility of providing such a rule.

This may sound peculiar at first blush, but let me explain in a little more detail why it is we should think IBE itself suffers from the Generality Problem. Essential to grasping this argument is the point that this is a metaphysical concern, not an epistemic one. The notion of IBE is not metaphysically coherent—at least not as a *unique* process of inference. If the reliabilist cannot show IBE to be unique, then his answer to the Generality Problem falls apart. If the realist who is a reliabilist assumes that IBE is metaphysically coherent as a unique inferential rule, then he is assuming *as a premise* something essential to his use of IBE—that the rule is even metaphysically possible—and this is viciously circular.

Let's start this line of argument with an analogy to the case we've already been dealing with, that of perceiving a cat, and then I'll move on to reasons for suspecting IBE's uniqueness on evidential grounds. Recall that in our example the original Generality Problem arises when I come to believe that I see a cat—there are indefinitely many different types that may be instantiated when I form that belief. I might be looking at the whole cat, its legs, cat parts, etc. The input will be the same in terms of the information entering my mind, but the processes that could be functionally operative may differ. These processes may dramatically vary in reliability, so there is no way to justify the reliability of my belief without a relevance condition on the process—something to tell us which process is being used. To evade this difficulty, reliabilists only need assert that there must be *some* process or other occurring which is either reliable or not. If it is reliable the belief is justified, if not, then it is not.

Translate this into the case with our inferential rule. When we use IBE we are inferring to the best explanation. This means that we take as input a set of explanations for some phenomenon, we evaluate them according to our criteria for what makes a best explanation, and we infer to the truth of that explanation.[20] Just as with our perceptual beliefs, we derive some output from a process that operates on some input. The input are the explanations, the output is our belief in one of those explanations. So, where is the Generality Problem here? Well, it at least arises in the sense that we find multiple different function types that are instantiated for any given token processing of an input/output pair. That is, for any set of explanations under consideration as inputs, there are multiple types of processes that get us to the same output belief.

[19] Strictly speaking this is analogous to the No Distinction Problem since there is no concern in MGP over actual frequencies of IBE—or even counterfactual or modal accounts of potential IBE's.

[20] If we incorporate the notion of a minimum threshold an explanation has to reach for it to be a good explanation, we might just be looking at whether or not we should believe a theory even without competitors. This would still be an IBE, it's just that there is only one explanation.

To make this more concrete, imagine we are trying to decide whether the adaptation of life to its environment is explained best by the theory of evolution or by creationism. Our inputs are the theory of evolution and the theory of creationism. The phenomena being explained are all the cases around the world where life seems to be well suited to its environment—giraffes being able to reach their food high in trees; polar bears having remarkably warm coats that prevent the otherwise life threatening temperatures they live in from wiping them out; parrots having peculiarly strong beaks with which to break nuts and seeds that are abundant in their habitat, etc.

The output will be the theory one decides is the best explanation for these phenomena—presumably evolution. The function between input and output is our friend IBE. But the question is, how is one to specify the *unique* type of process for this instance of inference? This is not the epistemic question of how we might justify the belief that IBE is reliable. This is the metaphysical question of whether there even exists a unique type of inference rule we can call IBE. If for any given instance of its use, there is no unique type that is instantiated by that token, then it is incoherent to suggest that IBE is a reliable rule of inference.

The argument I am making then, is that there is a generality problem for IBE. There are three steps to securing this thesis. First, I must show that there really are different types of IBE, and I appeal to the history of science to establish this premise. Second, it must be shown that these different types of IBE employed in the sciences are not equally reliable. Again, I appeal to the historical record to establish this point, but will go further and suggest that this evidence also compels the conclusion that even for a single type of IBE in the history of science, the reliability of that type itself varies across time. This makes the idea of a fixed reliability for any single type metaphysically problematic. The third step in establishing the plausibility of MGP is the most important one—showing that for any given token use of IBE, it satisfies potentially many different types of IBE, and these are likely to be of varying reliability.

What I hope to show in the next section then is that the required metaphysical assumption made by the realist (that there is only one unique, reliable type actually operative in each instance of IBE) is an assumption that viciously begs the question when interpreting the NMA as an IBE. There is also a potential response to my claims, and this response is found in Psillos (2002, 2007). I address Psillos' response in Sects. 3.3 and 3.4.

3.2.1 Historical evidence for a plurality of IBE's

Recall that IBE says 'If H explains a set of surprising data D, better than any other hypothesis, then infer that H is probably true.' To establish that this rule has been used in many different ways throughout the history of science we might worry about a host of issues. How is one to understand 'surprising data'? Does this have to be surprising to everyone, to the scientific community at hand, or just the person making the inference? How surprising does this data have to be, and can such a measure be made in an objective way? What is it to say H is probably true? Is it highly likely to be true on some subjective measure, by probability assignment on conditional updating, or is it perhaps that only parts of H are likely to be true?

Although the above issues are important for a full-blown account of IBE, I will avoid these difficulties because they are less helpful for making my modest point, which is merely that IBE is heterogeneous in the history of science. This can be done by showing how in scientific practice successful hypotheses have supposedly *explained better* than their competitors by appealing to very different underlying methodological principles. The diversity of kinds of explanation used in these cases of IBE illustrates the plurality of kinds of IBE in scientific practice. This will establish that different versions of IBE have been at work in science.

Instead of trudging through an exhaustive list of successful theories and categorizing the kind of explanations they provide, we can work the other way around, starting with commonly held explanatory strategies and drawing on cases from science instantiating them. Here is a non-exhaustive list of what *might* count as a best explanation, with some examples.

Causal mechanism: this is one of the most common forms of explanation in the sciences and examples range from electrostatic repulsion causing atoms to scatter when fired at one another, to genetic recombination explaining inherited characteristics. Establishing the mechanism by which events cause effects is a particularly compelling reason to accept a hypothesis (cf. Salmon 1998).

Unifying the phenomenon with others: The unification of apparently disparate phenomena under a single explanatory schema is used by scientists sometimes to show how something that looks unusual or odd, is in fact a case of something else, perhaps less unfamiliar. For example, Kitcher (1981) has convincingly argued that the appeal of Newton's theory was its generalized argumentative pattern for searching-out force laws that promised to unify phenomena as distinct as particle motion, light propagation, and chemical combination. Kitcher also illustrates how Darwin's theory of evolution promised the unification of a host of biological phenomena in terms of natural selection, inheritance, and variation of traits. These unifications are taken to be explanatorily compelling properties of theories.

Reliable methods for investigating the unobservable: using random experimental designs in laboratory experiments, following double-blind test procedures, and independently testing potential causal variables have all played an important role in establishing scientific theories. Such methods make for better explanations by ruling-out potential biases or confounding variables. Some examples include recent studies linking high levels of red meat consumption to colorectal cancer, establishing the effectiveness of vaccination against the infection of human papilloma virus, and rejection of the link between Vitamin C ingestion and shortening duration time for the common cold. Establishing a limited role for bias or confounders in experiments is a good reason to accept the integrity of explanations based on them (Giere et al. 2006).

Maximal coherence of propositions that entail the phenomenon: Hess' theory of seafloor spreading was accepted over competitors largely on empirical confirmation of predicted alternating magnetic properties in seafloor strata. However, this mobilist theory was far more explanatorily satisfying than its competitor theory developed by Wegener on the grounds that the theory had a more coherent and integrated set of commitments—notably that the rising molten material found in oceanic ridges solidifies after orienting itself to the magnetic field of the earth, and such orientations will change in correspondence with the change in earth's polarization (cf. Giere 1988).

Simplicity: often, and especially in mathematically expressed theories, the methodological advice is to chose the theory that is simplest—the one that has the least number of adjustable parameters. For example, it was a virtue of General Relativity that it introduced a curved space–time using non-Euclidean geometry, rather than adopting changes in many physical parameters in flat space–time (Kosso 1992).

Fruitfulness: this is the idea that a theory is preferable if it explains new phenomena or explains a growing base of data, or points to extensions of itself (McMullin 1976). For example, Bohr's theory of the hydrogen atom can be said to be fruitful in terms of pointing to Sommerfeld's extension of it with elliptical orbits (Losee 2001).

The intelligibility or understandability of a theory: examples abound here, sometimes relying on causal or nomological explanatory relations, but sometimes on more abstract components such as are found in iconic and analogical reasoning. For example, analogies used by Faraday and Maxwell, comparing lines of magnetic force to the transmission of a fluid, or the billiard ball analogy for kinetic theory, made these theories more acceptable (Dear 2006).

Novel predictive power of theories: Mendeleev's predictions for properties of undiscovered elements; Maxwell's prediction of the viscosity of a gas being independent of its density; LaPlace's prediction of the speed of sound; general relativity's prediction of curvature of light around a massive object; and of course the Poisson white spot, are all taken to be compelling examples of how novel predictive success can lend explanatory power to a theory (cf. Zahar 1973).

Correspondence between relations in past and successor theory: the fact that there are correspondence relations between classical and quantum mechanics, as well as classical dynamics and the special theory of relativity, plausibly contributed to the explanatory merit of these successor theories (Post 1971).

Explaining past successes and failures of a predecessor theory: this is a quality held by many theories that has contributed to their being selected. One excellent example is general relativity, where the theory was capable of illustrating why Newton's gravitational theory was able to predict the tides, yet failed to account for the perihelion of Mercury (Sellars 1963, Chap. 4).

Now if one accepts these examples as illustrating the diversity of explanatory considerations which go into the selection of one theory over others, then it is plausible to think that they are also principles which play a role in the pattern with which we are primarily concerned—IBE. In fact, we don't ourselves have to accept that each of these examples is a case of IBE, but the scientific realist with whom we are concerned uses many of these examples himself as instances of IBE. Where the NMA is a general argument that takes as its data points successful theories in the history of science, it takes the inference to those theories as applications of IBE (we saw this in the Boyd/Psillos two-step interpretation of the NMA). So, for the realist, these examples of successful science *have* to be cases of IBE. I take it that this therefore establishes the heterogeneity of types of IBE in scientific practice.

3.2.2 These principles are not reliable

Having established that there is a significant heterogeneity in principles implemented in IBE reasoning through the history of science, it is now relevant to ask whether these principles are reliable. If all these different types of IBE are reliable then the realist really has nothing to worry about. This is because even though MGP points to multiple types for a given token, its devastating consequences only follow if these types are of differing reliability—otherwise the realist can remain agnostic about the specific type of IBE instantiated in a given token, and simply appeal to its reliability. This is, I argue, not an option for the realist since these types of explanatory virtue do in fact differ in reliability, and in two important ways: contextually and diachronically. They differ in reliability dependent upon the context in which they are implemented, and they differ in their own individual reliability within a given context over time. I will use examples to support both of these claims.

Divergence in reliability across as well as within a single context is easily shown through the following examples, each of which represents the use of one type of IBE with clearly differing levels of reliability.

Causal mechanical modeling: It was famously Lord Kelvin's criterion for belief that one be capable of providing a mechanical model which generates the phenomenon to be explained. This is a popular view even today, and not surprisingly so since it has been a compelling method for establishing the credentials of theories such as the vibratory theory of heat and the kinetic theory of gases (which explain among other things the diffusion of gas through a room and the compression or 'springiness' of the air). However, causal mechanisms have also led us far astray in the history of science. Classical mechanical stories were given for a myriad of phenomena, each of which turned out to be false, illustrating the limited reliability of mechanical modeling. Examples of such mechanical failures include Descartes' vortex theory of the solar system, Huygen's mechanical explanation of gravity, Newton's corpuscular theory of light propagation, and of course ether theories of electric and magnetic phenomena (especially theories of the luminiferous ether). In all of these cases, attempted causal-mechanical modeling led us to accept (at least temporarily) theories that later turned out to be false, thus revealing the failure of such methodology to be reliable.

Unifying the phenomena may be reliable in Kitcher's examples (Newtonian mechanics and Darwinian evolution) but has actually failed rather dramatically in other cases. For example, one could plausibly argue that unification was an important component in the acceptance of both phlogiston and caloric theories. Phlogiston provided a unifying explanation not merely for processes of combustion, but also for calcination, respiration, and smelting. Similarly, caloric explained not merely the transfer of heat, but was also used to derive the adiabatic gas laws, and provided more accurate predictions for the speed of sound in air. Such unifications are compelling on the surface, but since these examples are cases of radically false theories, the realist in particular must accept the varying reliability of unifying principles through the history of science.

Using 'reliable' methods for carrying-out experimental studies is by definition not going to cause the realist any problem *within* a context. However, even here it must be recognized that context plays an important role. Performing a random experimental study on rats has the drawbacks that not only are humans (about which conclusions are usually drawn from such studies) very different from rats, but the size of doses used in such studies are frequently very different from those we humans are usually exposed to in our daily lives. Thus, for example, claims that ingestion of saccharin causes bladder cancer in humans has to be seriously qualified before we accept it on the basis of laboratory studies on rats. Similarly, when it comes to double-blind studies we have to be careful that our generalized conclusions are based on cases where selection, representation, recall, and interviewer biases are kept to a minimum. Importantly, how this is actually achieved in any given context is going to depend on background assumptions already prevalent in that context, and it is these that have led to erroneous theories being accepted. Examples include the acceptance that arthritis is related to the weather, the 'Mozart Effect' (classical music improves learning), the belief that sugar contributes to hyperactivity in children, the effectiveness of the D.A.R.E. program in U.S. schools, the recovery of repressed memories, the effectiveness of polygraph testing, and the prevalence of autism and multiple-personality disorder in the twentieth century (Lawson 2007). All of these topics were taken to be well established due initially to reliable test results. However, what we see is often the use of alternative testing methods, or conceptual concerns, raising serious problems for each.

Coherence: this principle seems to have worked well with sets of propositions interpreted as internal to theories, such as with Hess' theory of seafloor spreading, but the principle would impose an unnecessarily conservative constraint on science if it were imposed broadly to include coherence between current background beliefs and propositions within new theories. For example, Newton's gravitational action at a distance, Maxwell's electromagnetic vector field, Einstein's curved space–time, and the discrete jumps in quantum systems would all have been rejected if intertheoretic coherence had been the sole criterion for theory choice. So, although a plausible explanatory trait in some instances, coherence fails as a reliable indicator of correct explanations.

Simplicity: although on the surface this appears a straightforward property of explanations, evaluating competing theories for simplicity is extremely complicated. Is evolution really simpler than creation-science? Is Special relativity simpler than Newtonian dynamics? It seems the notion of simplicity itself has to be given a complex account in order to save these apparent counter-examples to good inference using simplicity as a measure (cf. Kosso 1992).

Fruitfulness: in fact the example above of Bohr's theory of the Hydrogen atom is a useful counterexample to the reliability of explanatory properties. It is clear that the model Bohr provides is not entirely accurate for Hydrogen, and certainly the assumption of electrons in discrete orbits around a central nucleus does not generalize to other elements. So, even here with this famous example, we see fruitfulness is not a reliable indicator of a good explanation.

Intelligibility: this property has led us astray most clearly in all the classic cases that form the basis for the PMI. Phlogiston, caloric, and the luminiferous ether all appealed strongly to the inherent understandability of particles moving as fluids, or elastic solids vibrating, to generate their respective phenomena. These are clearly unreliable indicators, although at the time taken to be marks of intelligibility.[21]

Novel predictions: although a very compelling property for explanations, there have been cases where theories that made novel predictions turned out false. Fresnel's theory is a good example of how a very surprising result (Poisson's white spot) was a confirmed prediction, yet the theory turned out to be entirely wrong about the nature

21 See Newman (2009) for an extended historical argument that these cases should not be counted in the PMI.

of light. The wave theory of light is defunct, and although white spots still appear they are explained now by very different mechanisms than those given in Fresnel's theory. Newton's gravitational mechanics also made the surprising prediction that a new planet inhabited the solar system given observations of Uranus' orbit. The new planet Neptune was found, but the theory is strictly speaking incorrect. (It has been argued by realists that these cases are not counter-examples to good explanatory theories because both Fresnel's wave optics and Newton's gravitational mechanics are approximately true. Attempts to specify exactly what this means are ongoing).

The correspondence between past and new theories seems to add to the explanatory power of a theory but even with our strongest case, quantum mechanics, there are good reasons to doubt this an explanatory virtue. Primarily concerns come from trying to specify exactly the nature of the correspondence at issue. Radder (1991) has illustrated the difficulties with assuming this case to be a simple and straightforward matter, suggesting that the transitions that occur between classical and quantum formalisms are multiple and logically strained. Hartmann (2002) provides further reason to suspect correspondence holds even in our best cases by arguing that the notion of correspondence itself has many meanings, and this suggests the realist cannot hang any justification for his position on a single interpretation of this relation.

Explaining the successes and failures of a predecessor theory seems reliable in many cases, but again there are exceptions. One particularly compelling example comes from the history of optics where Fresnel's wave theory seemed to explain the failures of its Newtonian corpuscular predecessor. However, as previously indicated, the wave theory is not now taken to be an adequate explanation of light propagation, so again we have the failure of a potentially reliable explanatory virtue.

The above examples provide reason to think that although common principles indicating better explanations can sometimes be good guides to theory choice, they are not always so. The lack of reliability for even the most promising types of IBE, such as causal-mechanical, or unifying forms, is a sign that IBE is not itself of a singular reliable type in any given token instance. Furthermore, these examples show not only that reliability of different types of IBE differ from one context to another, but also across different time periods even within the same domain of science. For example, causal-mechanical explanations seem to vary in their success across the sciences—think of trying to give a causal-mechanical explanation of a probabilistic event such as one finds in particle physics—but also across time scales within a single domain. We see this illustrated clearly again with causal-mechanical as well as unifying explanations. Where causes in physics helped to lend explanatory power to the vibratory theory of heat, they also lent plausibility to the Caloric theory—similarly for oxygen and phlogiston theories of combustion, and for corpuscular and wave theories of light. Unification also varies in reliability within a domain but over time. For instance, we've seen the unification of Galilean and Keplerian physics under Newtonian gravitational mechanics as a significant aid in explanatory power, but unification also supported the caloric theory of heat. So, it is reasonable from the above examples to conclude not only that principles indicating different types of IBE as used in the practice of science have varied across disciplines and over time, but also that in both dimensions these types are of varying degrees of reliability.

3.2.3 Each token of IBE falls under multiple types

Having established that there are many types of IBE employed in the sciences, and having shown that each has varying reliability, it still does not follow that MGP is correct. The necessary third step in establishing the plausibility of MGP is illustrating that for any given token use of IBE it satisfies potentially many different types. Since these types are of potentially varying reliability, it follows that there is no plausibility to the realist's metaphysical assumption that IBE is a singular natural kind type.

To establish that any given token of IBE potentially satisfies multiple types I adopt the following simple strategy for identifying IBE types: IBE's are of different types according to the properties they have that satisfy usage of the notion 'best explanation' in scientific practice. That is to say, we ascribe an instance of IBE in science to a 'best explanation type' category, determined by what was taken in that instance to be the relevant best explanation predicates. Thus, if caloric was preferred for causal reasons, then that instance of IBE falls into the IBCE category (inference to the best causal explanation). If caloric was however preferred for unifying reasons, then it falls into the IBUE category (inference to the best unifying explanation). And so on. Thus, a token of IBE falls into a single type if the theory selected was chosen based on the predicates naming that type. This approach gives us a plausible naturalistic approach to naming types of IBE and identifying tokens with such types by looking to the history of science.

The problem is of course that for any given token of IBE, the practice of science indicates there are multiple plausible types under which it falls. Furthermore, I have shown in the examples above that each of these types under which a token may fall can vary in reliability across context and over time.

For instance, are we supposed to categorize the IBE token of selecting Newtonian mechanics over competitors on the basis of its causal-mechanical properties, its unifying properties, its coherence, simplicity, or its fruitfulness? Which of these was it, or *should* it have been? Could it not have been a complex combination of all of these? In which case, isn't this just another type, albeit one with a far more detailed description? Even if we are confident that the theory was adopted on grounds of, let us say, causal attributes, what kind of causes were being appealed to? Was it causal contact action, instantaneous action at a distance, some kind of local field action, or what? The ambiguities are multiple.

The same goes for other theories we've considered. How should we categorize the IBE types in the instances of selecting Darwin's evolution theory, Maxwell's electrodynamics, Einstein's special and general theories, or the revolution in quantum mechanics? What rule do we have at hand to identify the relevant type of IBE in each of these instances? The answer is that we don't have one, and because of this the generality problem genuinely poses a difficulty for the scientific realist who wishes to interpret the NMA as an IBE. Without specifying for any given token instance of IBE the unique type under which it falls, the varying reliability of types undermines the appeal to such a token use. There is no non-question-begging rule of relevance to determine under which type to place any given token. Therefore, the scientific realist is unable to adopt an interpretation of the NMA in a non-viciously circular manner.

3.3 A possible escape by appeal to coherence?

There may be a means of escape for the realist from the accusation I am making—which remember is just that IBE is subject to the MGP and as such one cannot legitimately interpret the NMA as an IBE until this problem is solved. This potential route can be constructed from some further work by Psillos (2002, 2007). I will first sketch this potential escape from the MGP, then raise several objections which I take to show the proposal ineffective.

I have suggested above that not only is IBE heterogeneous, but also that because of variations in reliability of any given type of IBE, the inference rule falls prey to a generality problem. Importantly, the examples I have used to show this also illustrate the contextual nature of IBE—how it depends heavily on context for its reliability. Now although this would seem to undermine the objectivity of IBE, Psillos has suggested such contextuality can be accommodated on a realist construal of science by arguing for two important claims. First, the diverse nature of IBE (illustrated in my examples above) is, he says, only apparent. When we look more closely at scientific practice, we see that IBE's all have a common ampliative character. Second, when we appreciate the *role* played by explanations in science, we are justified in thinking IBE leads us to approximately true beliefs in virtue of it enhancing coherence in our belief corpus. Combined, these two claims provide an escape from my generality problem because they deny both the heterogeneity of IBE and the circularity of interpreting the NMA as an IBE. To make more sense of his claims so we can analyze them, I will briefly unpack each.

The first claim Psillos makes—that IBE is not heterogenous after all—requires a rejection of thinking about IBE in abstract terms. When we resist the temptation to provide a logical analysis of the inference rule, instead paying attention to the fine contextual details behind any given instance of IBE, we see, he claims, that there is a general schematic character to IBE. This character is one of a *genus* (rather than a species) of a general, ampliative, context-sensitive rule of inference. Instead of thinking of defeasible reasoning in abstract logical terms, Psillos encourages us to think of it as having a fine context-sensitive structure that appreciates the fact that explanations are stories ranked according to relevant background assumptions and knowledge. It is our background assumptions that determine the possible explanations available, as well as how to rank them. This is essential because for Psillos IBE has the following abstract template structure: it is a two-step explanatory quality test. We first use background assumptions (context) to evaluate the intrinsic explanatory quality of a potential explanation. We then compare those explanations that pass the first step. This is achieved by ranking the remaining competitors in terms of their structural features (completeness, importance, parsimony, unity, precision) and selecting the one most highly ranked. Importantly, IBE itself makes no commitment to any particular type of explanatory relation (causal, mechanical, nomological, unifying, etc.). It is context that determines which of these relations is explanatorily worthy. Thus, IBE is to be seen as a placeholder, a general overarching category for types of explanatory inference, all of which fall subject to the same two-step evaluative procedure.

The second claim Psillos makes is that this procedure (IBE) is truth-conducive. His reasoning is as follows: when it comes to IBE what really matters is not the particular type of inference made, be it causal, lawlike, or whatever. What matters is the *role* being played by explanation in our inferential practices. Explaining something incorporates it into the reasoner's background knowledge by linking it with other hypotheses already held by the reasoner. Explaining therefore has a coherence-enhancing *role*. Furthermore, explanatory coherence is what holds our belief corpus together, and aside from making the world more understandable, such coherence is an epistemically probative virtue: "Explanatory coherence is a cognitive virtue because it is a prime way to confer justification on a belief or corpus of beliefs" (2007, p. 445). To underscore the epistemic faith he has in the property of coherence he says,

> In the end, what IBE does is to enhance the explanatory coherence of a background corpus of belief by choosing a hypothesis which brings certain pieces of evidence into line with this corpus. And it is obviously reasonable to do this enhancement by means of the best available hypotheses. This coherence-enhancing role of IBE, which has been repeatedly stressed by Harman…Lycan…and Thagard…is ultimately the warrant-conferring element of IBE. (2002, p. 619)

To summarize the challenge posed by Psillos: IBE is not heterogeneous, as MGP requires, therefore the claim that NMA begs the question is not justified. Additionally, one can justifiably take IBE to be reliable in virtue of its ability to select coherence-enhancing explanations since coherence is a truth-conducive desideratum.

3.4 Response to Psillos

I wish to challenge this potential escape from MGP on four grounds. The first two concern Psillos' first claim—that IBE is not a heterogeneous type. The last two objections will address Psillos' heavy reliance on coherence as truth-conducive.

Is step one of IBE internal or external? In step one of IBE Psillos suggests we evaluate explanations on their own merits. To explicate the evaluative process he appeals to Pollock (1986) and his account of 'prima facie warrant', which he generalizes to scientific methodology:

> The presence or absence of defeaters is directly linked with the degree to which an ampliative method can confer epistemic warrant on an outcome, that is, the degree to which it can be epistemically probative. So, to say that S is prima facie warranted to accept the outcome Q of an ampliative method is to say that although it is possible that there are defeaters of the outcome Q, such defeaters are not actual. In particular, it is to say that S has considered several possible defeaters of the reasons offered for this outcome Q and has shown that they are not present. If this is done, we can say that there are no *specific* doubts about the outcome of the method and, that belief in this outcome is prima facie warranted. (2002, p. 609)

In this quote it looks very much like Psillos is characterizing the practice of IBE—of evaluating hypotheses—as an *internal* affair: we evaluate the plausibility of a hypothesis by considering ways in which it might be wrong (defeaters). If no defeaters either contradict or undermine the derivation of the thing to be explained, then we have reasons to believe the explanation. All of this is internal to the subject performing the evaluation (or community of subjects if we want to remain more in accord with actual scientific practice). Our justification for believing an explanation, based on this step, is grounded in our conclusions regarding the possibility of defeaters. This is to say that our justification is subject to our internal cognitive limitations—our ability to think-up and test possible defeaters—and not whether there actually are defeaters out there in the world, of which we may be ignorant. An important property of this evaluative process then is its internal nature. However, characterizing IBE in this way, as Psillos does, seems in tension with the whole project of interpreting the NMA as an IBE on an externalist, reliabilist epistemology. Yet that is precisely what the realist has been trying to do in the first place. It is exactly because reliabilism is being used to interpret the NMA, yet also requires an answer to the generality problem, that we ended up in this mess to begin with. If IBE is really an internalist methodology, then we are back to square one trying to answer the accusation of premise circularity.

There seem to be a couple of responses available to the realist here. First, one might suggest the account given by Psillos of IBE is indeed internalist, but that is because it is a characterization of *scientific* uses of IBE. When it comes to *philosophical* uses, the form of IBE used in the NMA is actually externalist.

However, introducing this distinction seems problematic for two reasons. First, even if we concede these different uses, what makes these different epistemologies relevant to the different instances? What relevance rule dictates when one form of IBE, let's say internalist, rather than the other is to be used? Second, conceding two forms of IBE, internalist and externalist, gives the game away against MGP—which claims that there are multiple types of IBE. Even if only internalist and externalist forms exist, that is enough to secure MGP.

The second response available to a realist is to concede that although Psillos' language indicates otherwise, IBE really is only to be given an externalist interpretation. On this view, consistency with the original rule-circular interpretation of the NMA is maintained, however the price is high: it is simply implausible to think that scientists evaluate the virtues of explanations with the help of background beliefs in an *external* fashion. In fact it is hard to imagine what such an evaluative process might look like. How could one evaluate the plausibility of instantaneous action at a distance in Newtonian gravitational mechanics without having awareness of accessing one's background beliefs about it? Given the internalist/externalist ambiguity of Psillos' characterization of step one of IBE we should therefore reject this line of escape for the realist.

The structure/content split is not convincing: In the second step of IBE Psillos suggests we make a structural comparative evaluation of explanations:

> To a certain extent, there is room for a structural specification of the best explanation of a certain event (or piece of evidence). That is, there are structural standards of explanatory merit which mark the explanatory power of a hypothesis and which, when applied to a certain situation, rank competing explanations in terms of their explanatory power. (2002, p. 615)

His list of these structural explanatory desiderata includes: completeness; importance; parsimony; unification; precision. These are to be applied after step one has isolated plausible explanations in light of background beliefs. As an example,[22] Psillos himself points to the fact that for Fresnel, after deriving important experimental results on polarization, it was still underdetermined based on background assumptions whether light waves should be taken to be solely translational or also have a longitudinal component. The latter factor was not necessary, and so the explanation which appealed to both components was less simple. Simplicity then, Psillos argues, is a structural feature which aids in theory selection, and hence is a component in IBE.

However, the adoption of these structural desiderata themselves relies upon substantive background assumptions, so they really cannot be separated from analysis of content in the clean way Psillos suggests. For example, simplicity, aside from being a rather vague concept (as we saw above) is only justifiably adopted as a realist desiderata on the assumption that simpler explanations are more likely to be true. This assumption itself relies upon the background belief that the world metaphysically and ontologically has a preference for simplicity. Now, this may be true, but establishing reasons for this belief will rely upon either *a priori* or empirical considerations. Naturalists will of course prefer the latter, but if appeals are made to the simplicity of the world based on background theories revealing it to be so, then the set of explanatory properties Psillos takes to be *structural* are nothing of the sort.

Taking coherence to be truth-conducive itself requires an IBE, and hence suggests circularity: Calling it 'Cartwright's Challenge', Psillos is worried about the following question: why should the information that a hypothesis is the best explanation of the evidence be a prima facie *reason* to believe that this hypothesis is true (or likely to be true)? In answering this challenge Psillos points to the relationship between background knowledge and explanatory coherence:

> It is this explanatory connection [between hypothesis and evidence] which makes the acceptance of H prima facie reasonable since it enhances the coherence of our total belief corpus. By incorporating H in our belief corpus BC as the best explanation of the evidence we enhance the capacity of BC to deal with new information and we improve our understanding not just of why the evidence is that way it is but also of how this evidence gets embedded in our belief corpus. (2002, p. 619)

As we have seen above, it is not just Psillos that takes the coherence-enhancing nature of IBE to indicate the method is truth-conducive—Harman (1986), Lycan (1988), Lehrer (1990), Pollock (1986), and Thagard (1988) are all in on the act (and Bonjour used to be). However, coherentist accounts of justification are heavily criticized in the literature, and it is extremely controversial to hang one's philosophy of science on such a suspicious epistemology. Still, even if coherentism turns out to be a correct account of epistemic justification, there is a tension in using coherence to justify IBE. In fact there are two tensions. First, coherentism is traditionally taken to be an *internalist* position, and although adopting it would be consistent with an

[22] Psillos (1999, pp. 217–219).

internalist description of the scientific practice of IBE, it would be in conflict with the realist's required externalist reading of the NMA—we have seen this issue already. The second tension is that the claim that coherence tends to produce justified beliefs is itself dependent on an ampliative, and hence defeasible inference. This means that the move from coherence to truth is inductive, and as a consequence it threatens to beg the question.

One response to this last concern might be to suggest that the inference from coherence to truth is a different kind of inference from IBE, and as a consequence no question is begged. However, this move would sacrifice the metaphysical uniqueness of IBE required by realists, and it would also run counter to Psillos' own argument that other forms of inductive inference really are just forms of IBE (2002).

Coherence has counterexamples: As we have seen, if we take step one of IBE to be an evaluation of the intrinsic explanatory qualities of a potential explanation then the analysis depends on background assumptions, and is hence contextual. For example, one is going to find Newton's gravitational mechanics far more explanatory if one's background beliefs lead one to be suspicious of causal-mechanical interactions. Now this does not entail that the notion of a good explanation need become entirely relative in the sense that there are no persistently good reasons to accept astronomy but not astrology. However, if one accepts the contextual relativism of good reasons for scientific theories and then suggest that these reasons are truth conducive in virtue of being coherent, as Psillos does, then one runs into the following problem:

1. Coherence of explanations indicates truth of explanations
2. PMI cases such as Phlogiston, Caloric, and the Luminiferous ether gave false explanations
3. PMI cases offered coherent explanations

According to the coherence account of explanations these three propositions cannot all be true. The motivation behind the realist project of interpreting the NMA as an IBE was driven by accepting (2) and (3). If these latter two propositions are accepted by the realist, as they seem to be, then (1) must be false.

The realist will probably respond that the notions of truth and falsity in (1) and (2) need to be modified to include the concept of 'approximate'. This would make the three statements consistent, but at a significant price: the idea that coherence leads to approximate truth does allow cases such as those from the PMI to arise, and this is acceptable to the realist; however, coherence does not accommodate radical conceptual revolutions such as those we find in the work of Newton, Maxwell, Einstein, Bohr, etc. This entails that a coherence account of explanatory veracity epistemically privileges those defunct theories from the PMI over the most successful scientific theories we have achieved. No plausible realist can tolerate this reversal of priority.

I take a combination of the above arguments to be conclusive. The realist must commit to IBE falling under multiple type descriptions if he is to accommodate its successful application through the history of science. Yet, it is precisely this point which makes IBE susceptible to a methodological version of the Generality Problem. That probably wouldn't bother the realist except that we have seen he needs to solve this MGP in order to solve the Generality Problem for cognitive processes. He needs to solve the Generality Problem for cognitive processes in order to justify his metaphysical assumption that such cognitive processes are even possible. This metaphysical assumption itself is necessary in order to justify the adoption of psychological realism, which we have also seen is necessary to secure reliabilist epistemology. And finally, of course, reliabilist epistemology is required because of its central role in this realist defense of the NMA. So, by committing to the multiple types under which any token of IBE may fall the realist is falling prey to a critical problem underlying his epistemology. For the realist to assume the legitimacy of adopting reliabilism without explicit justification, he is assuming an answer to the MGP. This begs the question in a premise, not a rule circular manner, because such an answer is not going to come in the form of a rule of inference. It is a metaphysical issue, not an epistemic one. The realist has to assume that IBE is a unique natural kind rule of inference, and although the rule is an epistemic process, the metaphysical status of the rule is certainly not. Not only is the metaphysical assumption therefore premise circular, the realist also actually denies it when accommodating the many different kinds of virtues implemented in IBE in the history of science. This is double-trouble for the realist.

I suggest, given the above reasoning, that if the reliabilist wishes to appeal to IBE to ground his local realism as a means of grounding his defense of IBE he has to viciously presuppose in the premises of his argument the metaphysical possibility of IBE being a unique type of inference rule. To do otherwise is simply to pull from thin air the assumption of its reliability.

4 Scientific realism as a philosophical package

I have argued that the scientific realist who adopts an externalist reading of the NMA is in fact viciously begging the question. Before leaving with this conclusion, one last realist response ought to be considered, and it is one to which Richard Boyd drew attention quite some time ago. In his (1980, 1985) Boyd points to just the concerns I have been raising for the realist. He notes for example that the realist inference is necessarily dependent upon an epistemology that itself might find justification from a realist interpretation—what I have been referring to as the local realist reading of mind/brain science. This problem is just the one we are dealing with, and his realist response to this question is both a concession and an advance. He answers the challenge by broadening the issue. He accepts that if our epistemology relies upon a realist interpretation and this alone is how one defends realism, then indeed the question is begged. All is not lost though, he thinks, because scientific realism ought to be viewed in a more holistic manner:

> The defense of realism, however, depends not upon the theory of epistemic contact *alone* but upon the ability of realists to incorporate suitably elaborated versions of it into an epistemological, semantic and metaphysical conception of the theory or tradition in question (a *philosophical package*) that is superior to those available to defenders of the various anti-realist conceptions. (1996, p. 250)

This notion of a realist 'philosophical package' is somewhat vague, but surely at least covers the three realist theses outlined on the second page of this paper: epistemic, semantic, and metaphysical. The important point is that the realist approach be more plausible (on general philosophical grounds) than alternative empiricist or constructivist packages. But is it? I have been arguing that the realist approach faces difficulties if defended with an externalist epistemology. How can this package be expected to win-out over its competitors?

Boyd's answer is to argue that the realist's approach is the only one that can provide justification for *even merely instrumental* findings in science—findings that the empiricist and constructivist also acknowledge as a form of knowledge. That is, the anti-realist of an empiricist stripe might argue that we can *accept* the reasoning methods (inductive rules) of realistic interpretations of science, although we need not believe these methods deliver the truth—they are merely useful instruments for arriving at reliable empirical truths—especially those that are projectible. Boyd's response is that such an inductive move itself requires an empiricist inductive justification, and this in turn will require a further inductive justification, …etc. The justification for adopting an anti-realist methodology for arriving at merely empiricist goals is thus reduced to an absurd infinite ascension though higher levels of justification. He concludes that the empiricist package is therefore only slightly plausible, and easily overwhelmed by the reasonableness of the realist package. In fact he goes further. In support of the realist approach he says,

> Given that the realist's package already incorporates an alternative, less speculative, and independently justified naturalistic epistemology I predict that it will prove superior. (1996, p. 252)

In response to this general approach I have provided an argument that undermines the claim made here. The realist's package under consideration requires an externalist epistemology, which despite its advertised potential, fails to ground the global realist thesis. The 'independently justified naturalistic epistemology' doesn't yet exist, and as such cannot provide the realist with the non-vicious rule-circularity he seeks. As such, it is a stretch to claim the realist package 'superior'. In particular, as I have illustrated, the realist package seems to lack internal coherence—the advocate of IBE has to assume its unique type character in order to answer the Generality Problem, but also has to deny its unique type character in order to accommodate its successful application through the history of science.

But although this argument defeats the defense of NMA proposed by reliabilist scientific realists, perhaps I am being unfair to Boyd himself, for he never explicitly appeals to an externalist epistemology, even though reliability of scientific method is definitely his central focus. I will leave this point open, and merely point to a possibility which still remains for the scientific realist: to defend a Boyd-inspired account of realism as a philosophical package by some other means where a unique type rule of IBE is defensible.

Acknowledgements Many thanks for helpful feedback on earlier drafts of this paper to Craig Callender, Jason Ford, Tristram McPherson, P. Kyle Stanford, and Sean Walsh.

References

Alston, W. (1995). How to think about reliability. *Philosophical Topics, 23*, 1–29 (Reprinted in Sosa, E., & Kim, J. (Eds.), *Epistemology: An anthology*. Oxford: Blackwell).

Alston, W. (2005). *Beyond "justification": Dimensions of epistemic evaluation*. New York: Cornell University Press.

Armstrong, D. (1973). *Belief, truth and knowledge*. Cambridge: Cambridge University Press.

Bird, A. (1998). *Philosophy of science*. London: McGill-Queens University Press.

Black, M. (1958). Self-supporting inductive arguments. *Journal of Philosophy, 55*, 718–725.

Bonjour, L. (1986). *The structure of empirical knowledge*. Harvard: Harvard University Press.

Boyd, R. (1981). Scientific realism and naturalistic epistemology. In P. D. Asquith & T. Nickles (Eds.), *PSA 1980* (Vol. 2) East Lansing, MI: Philosophy of Science Association.

Boyd, R. (1984). The current status of the realism debate. In J. Leplin (Ed.), *Scientific realism*. Berkeley: University of California Press.

Boyd, R. (1985). *Lex Orendi est Lex Credendi*. In P. M. Churchland & C. A. Hooker (Eds.), *Images of science*. Chicago: The University of Chicago Press.

Boyd, R. (1996). Realism, approximate truth, and method. In Papineau, D. (Ed.), *The philosophy of science*. Oxford: Oxford University Press. Originally published in Savage, W. (Ed.). (1980). *Scientific theories: Minnesota studies in the philosophy of science, XIV*. Minneapolis: University of Minnesota Press.

Braithwaite, R. B. (1953). *Scientific explanation*. Cambridge: Cambridge University Press.

Busch, J. (2008). No new miracles, same old tricks. *Theoria, 74*, 102–114.

Conee, E., & Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies, 89*(1), 1–29 (Reprinted in Sosa, E., & Kim, J. (Eds.), *Epistemology: An anthology*. Oxford: Blackwell).

Dear, P. (2006). *The intelligibility of nature*. Chicago: University of Chicago Press.

Douven, I. (2005). Wouldn't it be lovely: Explanation and scientific realism. *Metascience, 14*, 338–343.

Feldman, R. (1985). Reliability and justification. *The Monist, 68*, 159–174.

Fine, A. (1984). The natural ontological attitude. In J. Leplin (Ed.), *Scientific realism*. Berkeley: University of California Press.

Fine, A. (1986). Unnatural attitudes: Realist and instrumentalist attachments to science. *Mind, 95*, 149–179.

Fine, A. (1991). Piecemeal realism. *Philosophical Studies, 61*, 79–96.

Fine, A. (1996). *The shaky game* (2nd ed.). Chicago: University of Chicago Press.

Giere, R. (1988). *Explaining science*. Chicago: University of Chicago Press.

Giere, R., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning*. Belmont: Thomson-Wadsworth.

Goldman, A. (1976). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge*. Dordrecht: Reidel.

Goldman, A. (1986). *Epistemology and cognition*. Cambridge: Harvard University Press.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 74*, 88–95.

Harman, G. (1986). *Change in view*. Cambridge: MIT Press.

Hartmann, S. (2002). On correspondence. *Studies in History and Philosophy of Modern Physics, 33*, 79–94.

Hudson, R. (2004). The generality problem. *Southern Journal of Philosophy, 42*, 193–211.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48*, 507–531.

Kitcher, P. (1993). *The advancement of science*. Oxford: Oxford University Press.

Kosso, P. (1992). *Reading the book of nature*. Cambridge: Cambridge University Press.

Lawson, T. (2007). *Scientific perspectives on pseudoscience and the paranormal*. Upper Saddle River: Pearson Prentice-Hall.

Lehrer, K. (1990). *Theory of knowledge*. Boulder: Westview.

Leplin, J. (1997). *A novel defense of scientific realism*. Oxford: Oxford University Press.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.

Losee, J. (2001). *A historical introduction to the philosophy of science* (4th ed.). Oxford: Oxford University Press.

Lycan, W. (1988). *Judgment and justification*. Cambridge: Cambridge University Press.

McMullin, E. (1976). The fertility of theory and the unit for appraisal in science. In R. S. Cohen, P. K. Feyerabend, & M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science* (Vol. 39, pp. 400–424). Dordrecht: Reidel.

Musgrave, A. (1985). Realism vs constructive empiricism. In P. M. Churchland & C. A. Hooker (Eds.), *Images of science*. Chicago: The University of Chicago Press.

Musgrave, A. (1988). The ultimate argument for scientific realism. In R. Nola (Ed.), *Relativism and realism in science*. Dordrecht: Kluwer Academic Press.

Musgrave, A. (1996). Realism, truth and objectivity. In R. S. Cohen, et al. (Eds.), *Realism and anti-realism in the philosophy of science*. Dordrecht: Kluwer.

Newman, M. (2009). Beyond structural realism: Pluralist criteria for theory evaluation. *Synthese*. doi:10.1007/s11229-009-9463-6.

Nola, R., & Sankey, H. (2007). *Theories of scientific method*. Ithaca: McGill-Queens University Press.

Nozick, R. (1981). *Philosophical explanations*. Cambridge: Harvard University Press.

Papineau, D. (1987). *Reality and representation*. Oxford: Blackwell.

Papineau, D. (1993). *Philosophical naturalism*. Oxford: Blackwell.

Peacocke, C. (1986). *Thoughts, an essay on content*. Oxford: Blackwell.

Pollock, J. (1986). *Contemporary theories of knowledge*. Savage: Rowman & Littlefield.

Popper, K. (1956/1982). *Realism and the aim of science: From the postscript to the logic of scientific discovery* (W. W. Bartley III, Ed.). London: Hutchinson.

Popper, K. (1963). *Conjectures and refuations* (3rd rev. ed.) London: RKP.

Post, H. (1971). Correspondence, invariance, and heuristics. *Studies in History and Philosophy of Science, 2*, 213–255.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. New York: Routledge.

Psillos, S. (2002). Simply the best: A case for abduction. In A. C. Kakas & F. Sadri (Eds.), *Computational logic* (pp. 605–625). Berlin: Springer-Verlag.

Psillos, S. (2006). Thinking about the ultimate argument for realism. In C. Cheyne & J. Worrall (Eds.), *Rationality & reality: Essays in honour of Alan Musgrave* (pp. 133–156). Dordrecht: Springer.

Psillos, S. (2007). The fine structure of inference to the best explanation. *Philosophy and Phenomenological Research, LXXIV*(2), 441–448.

Putnam, H. (1975). What is mathematical truth? In H. Putnam (Ed.), *Mathematics, matter and method. Philosophical papers* (Vol. I). Cambridge: Cambridge University Press.

Putnam, H. (1978). *Meaning and the moral sciences*. London: RKP.

Radder, H. (1991). Heuristics and the generalized correspondence principle. *British Journal for the Philosophy of Science, 42*, 195–226.

Salmon, W. (1998). *Causality and explantion*. New York: Oxford University Press.

Sellars, W. (1963). *Science, perception, and reality*. Atascadaro: Ridgeview Publishing Company.

Smart, J. J. C. (1963). *Philosophy and scientific realism*. London: RKP.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge: MIT Press.

Van Cleve, J. (1984). Reliability, justification, and the problem of induction. *Midwest Studies in Philosophy, 9*, 555–567.

Zahar, E. (1973). Why did Einstein's programme supercede Lorentz's? *British Journal for the Philosophy of Science, 24*, 103.