ALAN MUSGRAVE

RESPONSES

"Gee, Mum, my name is in lights!" So said David Stove, when I used his name in the title of a paper of mine. Stove deserved the little spotlight I shone on him. I surely do not deserve the floodlights that my friends now shine on me. I am dazzled, like a rabbit transfixed in headlights. Heartfelt thanks to those in the driving seat, for shedding so much light on my worthless carcass!

I have been anxious to defend two pretty commonsensical positions - critical rationalism and critical realism. I learned both from Karl Popper. But most 'Popperians' think I learned badly and got it all wrong. So as to avoid exegetical issues, which are pretty unimportant anyway, I will speak of them as 'my positions'. I take no credit for the good bits, but all the blame for the bad bits.

The papers collected in this volume fall into two groups. There are papers about critical rationalism - the role it gives to observation and testimony (Greg Currie, Colin Cheyne), severe testing (John Worrall, Deborah Mayo), and other critical methods (Volker Gadenne, Howard Sankey, Stathis Psillos), Then there are papers about critical realism - the metaphysics appropriate to it (Michael Redhead, Alan Chalmers, Robert Nola, Mark Colyvan), antirealist views that stand opposed to it (Noretta Koertge, Graham Oddie), its impact on historiography (Hans Albert) and on our understanding of the early history of astronomy (Andrew Barker). I shall organise my responses accordingly, and in that order. My friends will forgive me if, from now on and in deference to academic proprieties, they are usually 'Cheyne' not 'Colin', 'Koertge' not 'Noretta', and so forth.

1. CRITICAL RATIONALISM

Critical rationalism is the view that the best method for trying to understand the world and our place in it is a critical method – propose views and try to criticise them. What do critical methods tell us about truth and belief? If we criticise a view and show it to be false, then obviously we should not believe it. What if we try but fail to show that a view is false? That does not show it to be true. So should we still not believe it? Here critical rationalists distinguish acts of belief (believings) from the things believed (beliefs). They think there are reasons for believings that are not reasons for beliefs. Failing to show that a view is false does not show it to be true, but is a reason to think it true – for the time being anyway. Thus, it may be reasonable to believe a falsehood, if we have sought but failed to find reasons to

C. Cheyne & J. Worrall (eds.), Rationality and Reality: Conversations with Alan Musgrave, 293–333.

think it false. If we later find reason to think a view false, we should no longer believe it. Then we should say that what we previously believed was false – not that it was unreasonable for us ever to have believed it.

This is just common sense. The trouble is that philosophical tradition denies it. Philosophical tradition says that a reason for believing something must also be a reason for what is believed, it must show that what is believed is true, or at least more likely true than not. I call this 'justificationism', and reject it. I think we can justify (give reasons for) believings without justifying the things believed. The chief bone of contention between me and the Popperians concerns this point – they reject all justification, and think critical rationalism has no need of a theory of justified or reasonable believing. I am particularly grateful for Volker Gadenne's support on this point. As he says, critical rationalism bereft of such a theory is really no different from scepticism:

> The rejection of any kind of justification means that, for every proposition P, it is equally justified to believe P as to believe non-P; and this is not rationality, it is Pyrrhonian scepticism. It doesn't help to call criticism rationality as long as one does not make clear how criticism contributes to bringing about situations in which some beliefs turn out to be more acceptable than others with respect to truth. (108)

In fact, philosophical scepticism is underpinned by justificationism. How might we set about establishing the rational credentials of some belief? Justificationism says that we must show that what is believed is true. If we try to do this, by invoking something else that we believe, the sceptic demands that we show this to be true as well. Off we go on an infinite regress, which can only be stopped by invoking certainly true 'first principles' of some kind – 'observation statements' if you are a classical empiricist, 'self-evident axioms' if you are a classical rationalist. The rejection of justificationism enables the critical rationalist to drive a wedge between scepticism about certainty (which is correct) and scepticism about rationality (which is not). Failure to show that a belief is false does not show it to be true, but does show it to be reasonable. But do not sceptics show that our beliefs are false? No, sceptics produce no criticisms of our beliefs – they only produce excellent criticisms of attempts to prove that our beliefs are true.

Justificationism also lies behind inductivism. Given justificationism, empiricists need 'inductive' or 'ampliative' reasoning to show that some evidence-transcending views are true or more likely true than not, and hence reasonably believed. And they need inductive logic to show that inductive reasoning is valid or 'cogent'. Critical rationalists reject justificationism, and hence have no need of inductive reasoning or inductive logic. Deductive reasoning is enough for them, and deductive logic is the only logic they have or need.

Nobody can get by just with reasoning or argument, and critical rationalists are no exception. All arguments, whether deductive or non-deductive, rest upon premises. Not all the premises of our arguments can be conclusions of previous arguments, on pain of infinite regress. Or, putting the same point in terms of beliefs, not all our beliefs can be obtained by inference from previous beliefs, on pain of infinite regress. Our arguments must start somewhere, with premises that are not themselves reached by argument. Or, putting the same point in terms of beliefs, if inquiry is to get started we must have some non-inferential beliefs.

But are any non-inferential beliefs *reasonable* beliefs? If not, and if an inferential belief is reasonable only if the beliefs from which it was obtained are reasonable, then no belief is reasonable. *Logomania*, the view that any reasonable belief must be obtained by reasoning from reasonably-believed premises, is a royal road to wholesale irrationalism.

2. OBSERVATION AND TESTIMONY

These general reflections raise the following questions. What are the sources of non-inferential beliefs? Can non-inferential beliefs be reasonable beliefs, and if so, when? The answer – or part of it – to the first question is obvious. Sense-experience and testimony are obvious sources of non-inferential belief. (A paragraph back I spoke of 'evidence-transcending views' – the evidence that such views transcend is the evidence of the senses.) The answer to the second question – whether beliefs obtained from sense-experience or testimony can be reasonable – is perhaps less obvious.

Greg Currie takes up these questions as regards sense-experience. He is anxious to defend the idea that experience has a role to play in epistemology. I agree. We also agree, I take it, that experience cannot yield an absolutely secure or infallible 'empirical basis' against which theories can be tested, because observation statements transcend the experience that prompts them, and are themselves 'theory-laden'.

Currie rebukes me for having suggested in one place that (as he puts it) "theory-ladenness [is] an essentially linguistic phenomenon" (8). On the contrary, he claims, it makes perfect sense to speak of observation or perception itself as being 'theory-laden' or at least, 'concept-laden'.

I never meant to deny this. I was anxious, first of all, to defend that basic sense of 'see' (more generally, 'perceive') whereby, for example, a cat can see a typewriter without possessing the concept < typewriter >. My cat sees the typewriter, for she does not bump into it when the mouse she is chasing runs under it. Cats (or people) can see an X in that basic sense without possessing the concept < X >, let alone any theory or belief about Xs. There are philosophers who, bemused by Kant, deny this. [PROOF: I once met a German philosopher who said that cats cannot see typewriters because they lack the concept < typewriter >. I said that my cat frequently saw my typewriter. She replied that Musgrave's cat could do impossible things – just like Schrödinger's cat, which manages to be both alive and dead until somebody sees it. She even speculated that Musgrave's cat might become as famous as Schrödinger's cat. I should be so lucky!]

What we need here is, of course, a familiar distinction. There is another sense of 'see', seeing-that, which is clearly conceptual. My cat sees the typewriter when her mouse runs under it, but she cannot see *that* the mouse has run under the typewriter, since she lacks the concept < typewriter >. Seeing-that is propositional, hence conceptual.

In between seeing and seeing-that, there is seeing-as. This is also conceptual – you cannot see X *as a Y* without possessing the concept < Y >. The cat that sees the typewriter cannot see it *as* a typewriter. Perhaps the cat sees the typewriter as something else. Perhaps she sees the mouse as food, and the typewriter as non-food. Perhaps all seeing is conceptual in the sense that whenever A sees B, A sees B *as a C* for some concept C.

Despite my incautious formulations, which I shall not defend here, I never meant to deny the distinctions between seeing, seeing-as, and seeing-that, distinctions that I have used and defended in other places. Nor did I mean to deny that animals bereft of spoken language can see-as and even see-that (more generally, perceive-as or perceive-that). Both of these are 'concept-laden'. But having concepts is one thing, having beliefs or theories that use concepts is another thing, and having words to express those beliefs or theories is yet another thing. Or so I believe.

Which brings me to Greg's own interesting discussion. The old seeing, seeing-as and seeing-that distinctions are notable for their absence from it. He talks of the "content of perception" and says it is "a matter of the way perception represents the world as being" (7). So perception or perceptual experience already has content, already represents the world as being a certain way. As he says: "Believing that owls fly requires that I have the concepts *owl* and *flying,* and having a perception with the content *there is a flying owl* requires this also" (8). Put in terms of the old distinction, seeing-that is the whole focus of his attention. The reason is plain. He focuses on seeing-that because he wants to give the representational content of experiences a justificatory role in epistemology, contrary to what he (and most others) take to be Popper's view of the matter:

> It is the content of experience that matters to epistemology. It is this content which creates the possibility that an experience may provide a rational basis for the assertion of a statement describing some state of affairs. (9)

My view of the matter is this. (I think it Popper's view, too, but I shall not argue the exegetical point here.) Currie speaks of an experience providing a rational basis for *asserting* a statement (or, we might add, for *adopting* or *forming* a belief). Must it, in order to do this, also provide a rational basis for the *statement itself* (or for the *content* of the belief)? Justificationism says YES: a reason for asserting (or believing) that P must be a reason for P itself. We need to reject justificationism in the epistemology of perception, as we do elsewhere. Does Currie reject it?

Suppose (to use his example) that I have a perception with the content *there is a flying owl*. Obviously, the content of my experience is a logically conclusive reason for the (content of the) belief that there is a flying owl, and an equally conclusive reason for the (content of the) assertion that there is a flying owl. After all, the three contents are identical, and *C* logically implies *C*. It is equally obvious that the content C of my experience is no reason at all for forming a perceptual belief with content C, let alone for asserting an observation statement with content C. Forming a belief or asserting a statement is an action that we perform. Reasons for actions are causes of them, and contents or propositions are not causes.

Currie will perhaps agree. At least, he says explicitly that "what matters is not content alone":

> I am claiming that experience is capable of playing a justificatory role in epistemology because of its content, and hence that some particular experiences – namely those with

> the right kinds of contents – do justify some assertions. What matters is not content alone, but the content's being the content of an experience. (9)

An experience with the content *there is a flying owl* presumably has "the right kind of content" for a belief or assertion with exactly the same content. So does the experience justify forming the belief or making the assertion?

The experience does not show that the (content of the) belief or assertion is true. Suppose that what I see is not an owl, but a pigeon – and suppose it is a stuffed pigeon, that is not flying but has been thrown. In this case, the assertion is false, and the perceptual belief is false, and *the experience is false as well*, for the same reasons. Admittedly, the last is linguistically odd. It seems odd to say that I can see that there is a flying owl without there being a flying owl. 'Seeing' is a success-word, like 'knowing'. As ordinarily used, "A sees that P" entails P, just as "A knows that P" entails P. But once we endow perception with content, we must allow that perception might have false content, and we must rid 'seeing-that' of any success connotations that it might carry in ordinary speech. Thus, the fallibility of observation statements or of perceptual beliefs cannot be evaded by endowing experiences with statement-like or belief-like contents. That just makes experiences fallible as well.

Moreover, it is not to be assumed that having an experience with content C invariably issues in a perceptual belief with content C, let alone in an assertion with content C. The latter is obvious – the perceiver may lack spoken language. The former is obvious, too. Seeing is not always believing. I may have an experience with the content *there is a flying owl*, yet not come to believe that there is a flying owl – perhaps because I am also possessed of the mistaken belief that owls are flightless birds. In this sense, perceptual belief is obviously 'cognitively penetrated' (to use Currie's expression).

Currie's discussion of 'cognitive penetration' is puzzling. He seems to have become a 'concept-monger', conflating concept-possession with belief-possession. He runs together the question of whether perceptual content requires the appropriate concepts (it surely does), with the question of whether it requires the appropriate belief. He goes off into a side-issue, conceding that we can *imagine* that there is a flying owl without believing it. He then insists that we can only imagine things if we have a suitable stock of beliefs and belief-generated concepts – "a creature with imaginings but no beliefs is not possible" (11). Not *possible?* As for belief-*generated* concepts, does the belief which *generates* the concept C also contain C, which means that we must already possess C to form the belief? Presumably not. So the belief that generates the concept C does not itself contain C – how then does this generating work? But never mind this. Currie says that "The case where perception and belief have identical contents … is an obvious case where the content of the belief renders intelligible the perception" (11). But identical contents are not required for this. I can believe that there are no black swans and then see one. The content of my perception is rendered intelligible by my belief, if you like, in that both contain the concepts <black> and <swan>. But the contents of belief and perception are not identical – they contradict one another! Currie says that it is hard to specify "the point at which perceptual systems deliver their outputs and belief takes over" (13). But what if belief never 'takes over' the output of the perceptual system? What if I do not accept the 'evidence of my senses', because of other beliefs that I possess? I can have a perception with a certain representational content, yet not form the belief with the same content. I can perceive that there is a flying owl without believing it. To take another example, anybody who is not fooled by the Muller-Lyre illusion is rightly correcting the 'evidence of the senses' in the light of other beliefs.

Where are we? Seeing that P is not always believing that P, let alone saying that P. And seeing that P, believing that P, and saying that P, might all involve a false P. Can no more be said about the epistemological role of experience? Seeing may not always be believing, but it often is. Having an experience with content C often causes a belief with content C. What is caused is not, of course, the content C - contents or propositions have no causes. What is caused is the formation or adoption of a belief with content C. The epistemological question is whether a perceptual cause of a believing is also some kind of reason or justification for that believing. Critical rationalism rejects justficationism and proposes that it is. If reasons for actions are causes, then causes of actions may sometimes be reasons for them. Critical rationalism proposes that when seeing that P causes a believing that P, then the seeing is a (defeasible) reason for the believing. If I have no reason to think P false, then seeing that P is a reason for believing that P. This holds even when P is false, when both my seeing and my believing are mistaken. Reasonable beliefs may be false beliefs, quite generally. Reasonable perceptual beliefs may be false beliefs, too. Still, sense-experience delivers us evidence, particular beliefs or statements about the world against which we can test other beliefs and statements. It is 'foundational' not in the sense that 'the evidence of the senses' is infallible, but just in the sense that it is non-inferential. Or so my critical rationalism maintains. I suspect that Currie's own view of the justificatory role of experience in epistemology is not much different from this.

He saddles Popper with the view that "experience lies outside the space of reasons" (Sellars, McDowell), that "nothing can count as a reason for holding a belief other than another belief" (Davidson). What Popper actually said was that *"statements can be logically justified only by statements" (Logic of Scientific Discovery*, p. 43, italics in the original), which is quite different. Davidson's slogan "nothing can count as a reason for holding a belief other than another belief" is ambiguous. Holding a belief, like forming or acquiring a belief, is an action, something we do. So is the slogan "Nothing can count as a reason for holding a belief other than [holding] another belief"? Or is it "Nothing can count as a reason for holding a belief [with content P] other than another belief-content [different from P]". The latter is ridiculous: belief-contents or propositions are not reasons for actions. The former is implausible: it means that all foundational believings, believings that do not arise by inference from other believings, are unreasonable.

In Section 4 of his paper, Currie tells us that McDowell sought to "bring experience into the space of reasons by seeing it as possessing … conceptual content". Currie thinks this is a mistake, because 'conceptual content' is a misleading term (8). True, an experience can have exactly the same content as a judgement or belief, and be a (conclusive) reason for it. But neither content is 'conceptual' in the sense that concepts are 'constituents' of the content. The difference between perception and belief is that the subject needs no concepts to perceive that P, but does need concepts to believe that P. It seems that my cat can see that the mouse has run under the typewriter after all. It is just that, lacking the requisite concepts, she cannot bring herself to believe it!

Is sense-experience the only source of such 'foundational' or non-inferential beliefs? No, testimony is another, arguably more important, source. I have written little about testimony. I extended the pretty commonsensical critical rationalist view of sense-experience to testimony as well. I said that it is reasonable to "Trust what other folk tell you, unless you have a specific reason not to". But that was barely scratching the surface.

I am grateful to **Colin Cheyne** for digging deeper. I agree wholeheartedly with most of what he has uncovered. We agree that many, perhaps most, of our beliefs are acquired from testimony, and that if we never believed what other folk told us our belief sets would be extremely meagre. We agree that if it is reasonable to accept the testimony of others, then the problem of induction is solved. Testimony can provide you with reasonable, evidence-transcending beliefs (believings). Cheyne occasionally writes as if this is a *criticism* of my critical rationalist attitude to testimony. I regard it, rather, as vindicating that attitude, and pointing up the absurdity of the traditional empiricist doctrine that all beliefs, or at least all reasonable beliefs, arise from personal experience. All of us do have lots of reasonable evidence-transcending beliefs that we acquired from other folk, some of which will, no doubt, turn out to be wrong.

Cheyne correctly reports me as maintaining that even a contradictory belief may be reasonably acquired through testimony. And, in pursuit of a *reductio*, he points out that according to me testimony might also yield reasonable belief in the validity of affirming the consequent, or of enumerative induction! Well, as they say, one person's *reductio* is the next person's derivation of interesting conclusions. The unsuspecting logic student who has the misfortune to have a very bad teacher may well come reasonably to believe that affirming the consequent is OK. As for the widespread belief in the validity (or 'cogency') of induction, that belief is not necessarily unreasonable, either – it may have been inculcated in those who possess it by bad teachers, heirs to a bad philosophical tradition. To paraphrase Cheyne (25), people who are surrounded by inductivists, pay close attention to them, perhaps even attend their religious services (seminars and conferences on inductive logic), may well acquire a reasonable belief in the validity of induction. Cheyne thinks this absurd: "The problem of induction is not so much solved as blown away! … a belief that inductive reasoning is reasonable may not be unreasonable, from which it appears to follow that inductive reasoning may be reasonable" (23).

Once testimony is admitted as a source of reasonable belief, it must be admitted that some folk may acquire from their elders and betters a reasonable belief that inductive reasoning is valid. That does not mean, of course, that inductive reasoning *is* valid - one may reasonably believe a falsehood, according to critical rationalism. Furthermore, folk who reasonably believe that inductive arguments are valid may also reasonably act in accordance with their false belief and reason inductively.

Is the problem of induction "blown away" by this? The problem is to avoid the irrationalist conclusion of the following Humean argument (19):

> We do, and must, reason inductively.
> Inductive reasoning is logically invalid.
> To reason in a logically invalid way is unreasonable or irrational.
> Therefore, we are, and must be, unreasonable or irrational.

What Cheyne has shown is that, once we admit testimony as a source of reasonable belief, we can avoid the irrationalist conclusion by rejecting the third premise. People who reasonably yet falsely believe that induction is valid, may reasonably act on their false belief. Just as children who reasonably believe in Santa Claus, because their elders and betters told them so, reasonably put Santa's supper in the hearth on Christmas Eve in accordance with their false belief. It is crazy to deny that the children do reasonably believe in Santa Claus. Just as it is crazy to say that the countless generations who were taught by their elders and betters that the earth stood still were unreasonable to believe this, just because it is false. What goes for Santa Claus or the stationary earth now goes for inductive logic. Or so deductivists like Popper and me believe.

Cheyne's discussion reinforces the point that words like 'reasonable', 'rational' and their cognates should be reserved for believings –beliefs are true or false, rather than reasonable or unreasonable. For once we admit testimony as a source of reasonable believing in certain circumstances, it will be impossible to say of any belief that it is unreasonable, meaning that it would be unreasonable in any circumstances for anyone to adopt that belief. That goes for belief in Santa Claus, God, a stationary earth, inductive validity, whatever.

Still, my formulation of Principle T (for testimony) was not careful enough. Cheyne objects that "as long as you refrain from criticising what you are told, your testimonial beliefs are reasonable. That cannot be right."(23). Indeed, it cannot. If you are in a position to criticise what you are told or to cast doubt on the veracity of your informant, and you refrain from doing either, you cannot be described as reasonably believing what your informant tells you. I accept Cheyne's more careful formulations of principles governing testimony.

I would add only one further point. There is, in these matters, an age of epistemic responsibility. Children reasonably believe in Santa Claus because Mum and Dad tell them so. They are in no position either to criticise what they are told or to doubt the veracity of their informants. It is different with grown-ups. And what goes for the children of today also goes for earlier generations of grown-ups. What used to be called 'the ethics of belief' is a neglected subject, chiefly because of the misguided empiricist notion that all reasonable evidence-transcending beliefs must arise by so-called 'inductive inference' from the so-called 'evidence of the senses'. Critical rationalism rejects this notion, and its 'ethics of belief' is the better for it – as well as being closer to common sense.

3. SEVERE TESTING

Sense experience and testimony are ways of getting started. They are sources of non-inferential reasonable beliefs, against which we can criticise and test other candidate beliefs. They are not infallible sources – some of the reasonable beliefs acquired from them will be false. Still, critical rationalism says that it is reasonable to believe (adopt, prefer) that evidence-transcending hypothesis, if there is one, which has best withstood serious criticism from these or other sources. One way to criticise a view is to subject it to the 'tribunal of sense experience'. (This does not just encompass our personal experience – it includes the experiences of others, transmitted to us through their testimony.) In the sciences this turns into the method of experimental testing. Clearly, critical rationalism owes us a story about what counts as a serious empirical criticism, or a severe experimental test, the result of which might genuinely confirm or corroborate a theory. John Worrall and Deborah Mayo both revisit this issue, and disagree sharply about it. I was tempted to let them fight it out – but I cannot resist entering the fray.

John Worrall and I agree that whether evidence **e** confirms theory **T** is to be assessed, not merely by considering the logical relations between **e** and **T** (as a 'purely logical' theory requires), but also by considering a third thing, 'background knowledge'. We also agree that a 'strictly temporal' view of background knowledge is no good. That leaves what I called the 'heuristic view' and the 'background (or touchstone) theory view'. Worrall favours the former, I once tentatively favoured the latter.

Worrall's first objection to the background theory view is that it yields the result that evidence that confirms the background theory **B** to a new theory **T** cannot also confirm **T**. He says "this is surely an extraordinarily counterintuitive judgement" (36). As he sees it, scientists will say that such evidence confirms *both* theories – though he concedes that scientists will be especially interested in whether the new theory is *better* supported than the old, a question to which evidence that supports both of them is irrelevant.

What this objection makes clear is that the background theory view makes evidential support an irredeemably *comparative* affair. On this view we cannot ask "Does **e** support **T**?", but only "Does **e** support **T** as against **B**?". (That evidential support is irredeemably comparative has also been argued by Larry Laudan and Elliot Sober. An early anticipation of it is 'Refutation or Comparison' by Archibald.) The background theory view is also irredeemably *historical*, because it is history that determines what the background theory **B** to **T** actually is. Worrall seems to ignore this historical dimension when he invites us to view the General Theory of Relativity as the background theory to Classical Physics, as well as viewing Classical Physics as the background theory to the General Theory of Relativity (32). You cannot do the former, if you take the *historical* character of the theory seriously.

Worrall finds it absurd to say that while there are phenomena that support Relativity but not Classical Physics, there are also phenomena that support Classical Physics but not Relativity (35). But these oddities fall away if you take the *comparative* nature of the theory seriously. There is nothing odd about saying that there are phenomena that support Classical Physics as against its rival (whatever that was), and other phenomena that support Relativity as against its rival, Classical Physics. Worrall insists that the former phenomena also support Relativity (provided that Relativity yields them in a non ad hoc way) "in the non-comparative sense of support" (38). But on the background theory view, there *is* no "non-comparative sense of support".

Is such a historico-comparative view of evidential support acceptable? It introduces a historical relativity into the issue, which seems unacceptable. I drew attention to this myself in my original 1974 paper: "… because Einstein had the misfortune to be preceded by Newton, his theory cannot be confirmed by all the evidence which it predicts, but which is also predicted by Newton's theory" (ERR, p. 246). If the 'Newtonian interlude' had never existed in the history of science, and Aristotle had been succeeded by Einstein, then Einstein's theory would have been much better supported (comparatively speaking) than it actually was! This is Worrall's chief worry, yet again.

But Worrall's own preferred theory confronted a similar worry. Worrall prefers the 'heuristic' view of novelty: evidence **e** is novel for theory **T** and supports it if **T** entails **e** but **e** was not used to construct **T**. My original worry about this was that if scientist **A** uses **e** to construct **T**, and scientist **B** constructs **T** without using **e**, then **e** supports **T** as proposed by **B** but does not support **T** as constructed by **A** (ERR, p. 241; cited by Worrall, p. 42)). Should theoreticians lie down on their couches and forget about the available data, if they want well-supported theories? Worrall's reply to this was "Science is not like that". Two independent considerations suggest that this is right.

The first is that scientists do not typically arrive at hypotheses randomly or through mystical flashes of intuition, but rather (as Newton said) by 'deducing them from the phenomena'. Newton was right that 'deduction from the phenomena' is deduction (not induction, abduction, or any other ampliative process of inference). Newton was wrong that its premises are just observed phenomena - scientists also need general 'heuristic principles' of one kind or another as well. Scientists would be crazy not to use known facts to help construct their theories, and there is nothing wrong with doing so. But you cannot use the same fact twice, as a premise from which you deduce your hypothesis, and as support for it. Worrall argues that the real problem here is the 'prediction *versus* accommodation' problem, a.k.a. the '*adhocness* problem', a.k.a. the 'independent testability' problem. A theory is not confirmed by evidence that it entails if it merely accommodates it, or is *ad hoc* with respect to it, or if it did not result from an independent test – where 'accommodates', '*ad hoc',* and 'not independent' are all cashed out in terms of 'used to construct'.

The second independent consideration is more philosophical. The Miracle Argument says (roughly) that the success of a theory would be miraculous if that theory were not true. The success spoken of is predictive success. But some predictive success is not miraculous at all. It is no miracle that a theory successfully predicts facts used to construct it.

Baby examples can illustrate both considerations. Suppose we do not know and want to find out what colour emeralds are. Should we lie on our couch, dream up hypotheses, and subject them to test? No, we should find an emerald, note its colour, and run through a so-called ‘demonstrative induction’ (actually a deduction): “Emeralds all share a colour, this one is green, so emeralds are all green”. Again, suppose we do not know and want to find out what the relationship is between two measurable quantities, P and Q, and we have a hunch that it might be a linear relationship. Should we lie on our couch, dream up linear hypotheses, and subject them to test? No, we should measure two pairs of values of P and Q, and do some ‘curve-fitting’ (actually a deduction): “P = aQ + b, for some values of a and b; when Q = 0, P = 3; when Q = 1, P = 10; so P = 7Q + 3”. It is no miracle that the hypotheses in these baby examples yield the observed facts used to construct them - neither do those facts support the hypotheses.

Of course, real science is not like these baby examples. In real scientific cases, the ‘heuristic principles’ that figure as premises in ‘deductions from the phenomena’ are specific ‘hard core’ principles of particular research programmes. Confirmation is not relative to persons, on the heuristic view, but relative to research programmes. And there is nothing undesirably subjectivist about that. Scientists in different programmes cannot come up with the same hypothesis (44). Scientists in the same programme can come up with the same hypothesis in different ways, but this is quite benign. Either one of them has read a parameter off the data when there was no need to do so, or one of them has mistakenly fixed a parameter from theory when it really could only be read off from the data. In the first case the data support the specific hypothesis and the programme in which it is embedded, in the second case they do not (44-47).

Can this be the whole story? There is a type of hypothesis that is constructed, not from particular observed facts, but from another hypothesis. This is the *surrealist* hypothesis T*: “The phenomena are *as if* T were true”. Surrealist hypotheses are constructed, not by scientists, but by antirealist philosophers of science. They are constructed by simple deduction from T: “The phenomena are as if T were true” is a fancy way of saying “T is empirically adequate”, and truth entails empirical adequacy. T* is, by design, empirically equivalent with T. But is it evidentially equivalent with T, is it equally well-supported by the evidence? On the ‘background theory’ view, clearly not: since T is obviously the ‘background theory’ to T*, there is no independent evidence at all for T*. Worrall’s heuristic view is less clear about the case. But he can say, perhaps, that we implicitly use all the ‘phenomena’ that T entails to construct “The phenomena are as if T were true”. (This would be in line with his analysis of the ‘Gosse dodge’. Armed with the general principle that God created the Universe in 4004 BC as if the teachings of geology and evolutionary biology were true, the creationist finds out that geology and evolutionary biology entail fossils in the rocks, and straightway formulates the specific hypothesis that God decided to install ‘fossils’ in the rocks at the creation.)

Worrall considers another objection to his view, which my baby examples will make clear. *Given that* emeralds all share a colour, the observation of one green emerald *deductively entails* that all emeralds are green - what better evidence could there be? *Given that* the relationship between P and Q is linear, the results of a couple of measurements *deductively entail* a specific linear equation – what better evidence could there be? (This assumes, of course, that the observation is correct, and that the precise values of P and Q obtained from the imprecise measurements are correct as well. That is a different issue.)

In response to this objection, Worrall develops a dual view, according to which there are two types of confirmation or evidential support. First, e supports T' *relative to T* if T and e entail T' - in this case, e does not also support T. Second, e supports T' absolutely or unconditionally if T' predicts e and this prediction is experimentally verified - in this case e also supports the general T of which T' is a special case (in the baby examples the general Ts are that emeralds share a colour and that the relationship between P and Q is linear).

Worrall says he is a residual Popperian who thinks that "a test of a theory must surely be able to refute that theory" (58). I am more of a residual Popperian than Worrall. I also think that evidence for a theory must come from testing that theory. Worrall does not accept this. I would not call e 'evidence' for T', just because there is some T which together with e is deductively conclusive reason for T'. If we are not fussy, there will always be a T which together with e entails any T' – "If e then T'" will do. Nor, for similar reasons, would I grant that if e is evidence for T', then it is evidence for any T entailed by T'. Suppose that observing swans in Europe gives us evidence that all swans are white (T'). Does observing European swans give us evidence that *Australasian* swans are white (T), just because it is evidence for T' and T' entails T? (I had to get this example in – after all, having black swans in it is Australasia's chief contribution to the philosophy of science.)

I wish Worrall had stuck to his guns, and not developed his dual theory of confirmation. I wish he had said instead that observation and experiment, and the data yielded by them, play two roles in science. First, they help us construct theories. Second, they enable us to test theories and, if we are lucky, confirm or support them. Worrall should grant his critics that e can be a conclusive *reason* for T' *given T*, but not grant that e is any kind of *evidence* for T'. As we will see, Worrall says something very close to this against Mayo's theory of the severity of tests.

I turn to **Deborah Mayo**'s wide-ranging and combative piece, which puzzled me at first. As explained already, critical rationalism owes us a theory about what a serious (or severe) empirical test is. Deborah Mayo is best known to the world for her resolute defence of a particular theory about this. What puzzled me initially was that there would seem to be nothing to stop the critical rationalist from adopting Mayo's theory. She formulates the general principle of critical rationalism thus: "CR: It is reasonable to adopt or believe a claim or theory P which *best survives* serious criticism" (64). Why cannot we expand this to "which *best survives* serious criticism in Mayo's sense"?

Mayo obviously sees things differently. She thinks that her theory of severity of tests is quite at odds with critical rationalism. She says repeatedly that a hypothesis can be the "best-tested" or "best survive serious empirical criticism" according to CR without having been severely tested or seriously criticised at all (e.g. 70, 71, 72). She asks "But why should it be reasonable to believe in the first hypothesis put forward …?"(71). She talks about "the critical rationalist's problem: being unable to say what is so good about the theory that (by historical accident) happens to be best-tested so far" (93). Critical rationalism has obvious answers to all this. If the 'first hypothesis put forward' has not survived serious criticism or severe tests, then CR will not say that it is reasonable to believe it. A hypothesis that has not been tested at all cannot have been tested better than any other hypothesis. What is good about the best-tested theory just *is* that it is the best-tested theory (so far).

What is going on? Mayo's own theory of severity is, in brief, that a test is a severe test of some hypothesis *h* if its outcome would be highly improbable if *h* were false. So the severity of a test with outcome **x** depends not just on p(**x**, h), but also on p(**x**, not-h). In the simplest case, where h entails **x**, the former is 1. To find out whether the test is severe we need to estimate p(**x**, not-h). How to do this?

At one point, while bashing Popper, Mayo says:

> "P is false" includes the disjunction of all possible hypotheses or claims other than P that would also "fit" or accord with **x** – the so-called 'catchall hypothesis' – including those not even thought of. Existing data **x** would be just as probable were one of the catchalls true, and P false. (71)

In what sense does "P is false" (equivalently "not-P") *include* this disjunction? It certainly does not logically include or entail it. The disjunction had better not include **x** itself, which fits **x** like a glove. If it did, "one of the catchalls" – which I take to mean one of the disjuncts in the catchall – would also entail **x**, and **x** would always be "just as probable were one of the catchalls true, and P false".

(By the way, critical rationalism does not traffic in the undreamt-of possibilities of the catch-all. It need not traffic in them, because it does not seek to justify any hypothesis. Its question is, which of the *available* competing theories is it reasonable to adopt or prefer or believe? You cannot believe an undreamt-of hypothesis.)

For Mayo, *'H is false' is Not the So-called Catchall Factor* (as she tells us in a section heading on 92). Mayo's view is that "mere accordance between **x** and P – mere survival of P - is insufficient for taking **x** as genuine evidence for P. Such survival must be something *that is very difficult to achieve* if in fact P deviates from the truth (about the phenomena in question)." (71). Well, what does P say about the phenomena in question? In the simplest case, P says (entails) **x**. So if P "deviates from the truth (about the phenomena in question)", we have **not-x**. Now "mere accordance between **x** and P – mere survival of P – is *very difficult to achieve* if in fact P deviates from the truth (about the phenomenon in question)", since p(**x**, **not-x**) = 0. On this reading, then, all tests of deductively entailed predictions have Mayo-severity 1.

This second reading is the right reading of simple cases like this, if Worrall is right. He complains that for Mayo, the process of adding up the SAT scores of the students in her class and dividing the total by the number of students to arrive at the number 1121, is a maximally severe 'test' of the 'hypothesis' that the average score is 1121. That is because the chance of the number arrived at being 1121 *if the 'hypothesis' were false* (that is, if the average score was not 1121) is zero. Like Worrall, I find it bizarre to talk of a 'test' or a 'hypothesis' in this case. The procedure is a demonstration that the average score is 1121, not a 'test' of the 'hypothesis' that it is. If we have done our sums correctly, the procedure is completely reliable 'error probe' and we can infer from its results that the 'hypothesis' is true. In this case, the hypothesis is deduced from phenomena alone (the definition of what an average score is, being true by definition, is a redundant premise). In other cases, the hypothesis is deduced from phenomena and other premises. In all cases, we gather the data not to test the hypothesis that is to be deduced from the data, but to figure out what the hypothesis is.

Suppose the hypothesis that we are interested in is the crazy hypothesis that the average SAT score in *all* classes is 1121. What does this crazy hypothesis say about the 'phenomenon in question', namely Mayo's class? It says that the average score is 1121 in that class. And the sums are a severe test of that hypothesis, too – for the chance of the sums yielding the answer 1121 if what the crazy hypothesis says about Mayo's class is false is zero.

(I said just now that the chance of arriving at the number 1121 for the average score, if the average score were not 1121, is zero. That was not strictly true: there is a small chance that I made an error in my sums. But this chance is the same whether or not the average score is 1121. The question of whether the evidence is reliable is not the same as the question of whether it represents a severe test of some hypothesis.)

As well as speaking of P deviating from the truth "about the phenomena in question", Mayo repeatedly cashes out the supposition that H is false as "a specified flaw in H is present" or "a specified discrepancy from H is present" (82). Again, she says that "H is false" refers to a "specific error that hypothesis H may be seen to be denying" (92). What 'specific error' is this? The specific error denied by H is, in the simplest case, the denial that its prediction about the case is mistaken. Mayo is quite open about this:

> What enables this account of severity to work is that the hypothesis H under test by means of data x is designed to be a specific and local claim, e.g., about parameter values, about causes, about the reliability of an effect, or about experimental assumptions. 'H is false' is not is disjunction of all possible rival explanations of x … This is true, even if H is part of some large scale theory T: the condition 'given H is false' always means 'given H is false with respect to what it says about *this particular* effect or phenomenon'. If a hypothesis T(H) passes a severe test we can infer something positive: that the theory T gets it right about the specific claim H, that severely passes.
>
> The price of this localisation is that one is not entitled to regard full or large-scale theories as having passed severe tests as long as they contain hypotheses and predictions that have not been well probed. (92)

The upshot is that Mayo has nothing to say about which "full or large-scale theories" should be believed or accepted or preferred. Do not be fooled by the phrase "full or large-scale theories". Let T be any theory that entails but is not entailed by H, so that it might have another testable consequence H'. A Mayo-severe test of H does not allow us to say that T has been Mayo-severely tested: we are "*not* allowed to say that the entire theory is severely probed as a whole" (93). All we are allowed to say is that the particular testable consequence H has been 'severely probed'. In effect, she denies that we test a theory by testing its consequences, insisting that all we have *really* tested are the consequences! No wonder she rejects the 'comparativist' view that we should tentatively believe or accept or prefer that theory (if there is one) that has been Mayo-well-tested. No "full or large scale" theory can be Mayo-well-tested. It is not for nothing that her book is called *Error* *and the Growth of* ***Experimental*** *Knowledge.* She denies that "full or large-scale" theoretical knowledge grows.

Critical rationalism goes further than this. It proposes that it is reasonable to believe (adopt, prefer) that theory, if there is one, whose consequences have been best tested. Why is Mayo reluctant to take this further step? Basically, because this proposal about which theories we should (tentatively) believe has not been shown to be *reliable.* Can we show that following this proposal, adopting this belief-producing stratagem, will lead us to believe more truths than falsehoods? Unless we can, the proposal is to be rejected. Mayo rejects critical rationalist methods because they have not been shown to be reliable. Not that she has up her sleeve an alternative method of choosing between evidence-transcending theories that she thinks reliable. Rather, she thinks that there is no such method. That is why we should not believe any evidence-transcending theory. We should stick to reliable experimental methods and the experimental beliefs they licence.

But must a rationally adopted method be a reliable method? Consider the parallel question: must a rationally adopted belief be a true belief? Critical rationalism answers NO to the parallel question – we can rationally believe falsehoods, if we have tried and failed to show them to be false. All this is part and parcel of the rejection of justificationism. But when it comes to the meta-level or methodological level, Mayo is a justificationist. She is in good company. Other contributors to this volume, such as Sankey, Psillos, and even (in one small place) Gadenne, are meta-level justificationists, even though they all explicitly reject lower-level justificationism.

Perhaps, on this point, I may be allowed to quote myself (in English – 'BPS' stands for a belief-producing stratagem or method):

> Must a rational BPS be a reliable BPS?
>
> It might seem obvious that rationality requires reliability. After all, to believe is to think true. *If I find out* that something I believe is false, then it is no longer rational for me to believe it. Quite so. But the words 'if I find out' are crucial … After all, this is what makes room for rational beliefs which happen to be false, though I have no reason to think them so.
>
> Similarly with reliability. If I find out that BPS on which I have relied is not reliable, then it is no longer rational for me to rely upon it. But I do not need to show that a way of acquiring beliefs [is reliable] in order for it to count as rational. Of course, once I show that a way of acquiring beliefs is not [reliable], then I am epistemically at fault in persisting in that general strategy of acquiring beliefs. There is a kind of asymmetry here …
>
> Reliability is a desideratum on BPSs, just as truth is a desideratum on beliefs. As with beliefs and truth, so with BPSs and reliability. A belief does not have to be true to be reasonable … But if you find out that a belief is false then it is unreasonable to persist in it. A BPS does not have to be reliable to be reasonable … But if you find out that a BPS is unreliable then it is unreasonable to persist in it. (Musgrave 2001, pp.111-112)

If this is right, then whether it is reasonable to adopt some method or BPS depends on whether it has been criticised and shown to be unreliable.

Here I should confess an error. Part of Mayo's trouble is of my making. She says that critical rationalists "deny the reliability of the method they espouse" (64), "feel bound to deny the reliability of the methods they espouse" (64), "deny that tests which are severe in the critical rationalist's sense are reliable tools for uncovering errors"(64). When I read these statements, I wondered where Mayo got them from. Then I found that she got them from me! I did once say "Critical rationalists deny that the process they commend is reliable" (ERR, p. 346; cited by Mayo on 73). That was a mistake. What I should have said is that critical rationalists need not assert or prove that their method is reliable before they can rationally adopt it. However, if the method can be criticised by showing to be *un*reliable, then they should cease to employ it.

Never mind my mistake. What is more interesting is Mayo's claim that the critical rationalist rule CR is "demonstrably unreliable" (64). That claim, if it could be made out, should give the critical rationalist pause. But has she made it out? Has she shown that to adopt or believe claims which best survive serious criticism is to adopt false claims more often than true ones? I do not think so.

What she does instead is discuss how we might defend epistemic principles or methods such as CR. She describes an argument of mine as "subtle and interesting", and then amuses herself by trying to show that it is neither subtle nor interesting. The argument was:

> Any general epistemic principle is either acceptable by its own lights (circularity), acceptable by other lights (hence irrational by its own lights …), or not rationally acceptable at all (irrational again). So even though the rational adoption of CR involves circularity, this cannot be used to discriminate against it and in favour of some rival theory of rationality." (ERR, p. 331: cited by Mayo on p. 74)

By a 'general epistemic principle' I meant a *comprehensive* principle, which says that a belief is rational if it has some feature F and that having feature F is the *only* way that a belief can be rational. The context of my discussion was two-fold: Bartley's comprehensively critical rationalism, developed in opposition to Popper's view that belief in reason must be based on irrational commitment; and Nozick's view that "justificatory principles … deep enough to subsume themselves" are "a triumph". Mayo says that the form of a "general epistemological principle" is "EM: a claim P is acceptable if it is classified as acceptable or beliefworthy by belief-classification method M" (74). She supposes that EM is acceptable by its own lights. She then says "this does not yet entail that the *only* warrant for EM is EM" (74). Perhaps not. But what EM does entail, if it is a 'general epistemological principle' in my sense, is that this other 'warrant' for EM is unacceptable so that beliefs 'warranted' by it (including belief in EM) are not thereby shown to be acceptable. (By a general epistemic principle I mean what Mayo calls a 'self-sealing' principle. If a method is not self-sealing, then it is not the only method, and we can ask why we should accept this further method, and be off on a regress.)

This is all terribly abstract. Should we care about 'general epistemic principles'? Do we not always have a *multiplicity* of principles, or methods, or 'warrants'? Perhaps. But let the multiplicity of methods be M_1, …, M_n, form their disjunction M*, and let the general principle EM* be that a belief is rational if and only if it is classified as such by M*. (This assumes, non-trivially, that the different methods will yield consistent results.) Now ask whether belief in EM* is rational. This is the question of whether reason can be defended by reason, which Popper, Bartley, Nozick and I were discussing. (Related issues are whether scientific methods can be defended by scientific methods, whether logical principles can be defended by logic, and so on.) I stick by my argument.

Mayo discusses something different, but I am not sure what. She formulates a principle saying (in effect) that we should accept the claim that some book is in print if the book is listed in the *Handbook of Books in Print* (BIP). This principle clearly specifies a 'method' which applies only to a restricted class of claims. It is also clearly a 'meta-principle' about that restricted class of claims – it is not itself a claim that some book is in print. Mayo imagines that the principle is printed on the first page of BIP, and so is "acceptable by its own lights". No, since it is not itself a claim that some book is in print, it is not acceptable by its own lights at all. (If BIP listed BIP, the claim that BIP is itself in print would be acceptable according to the principle.) Obviously, we need to assess the acceptability of this principle in other ways, by trying to criticise the claim that BIP is a comprehensive list. As Mayo says, we reject Sloppy Joe's Books in Print because we could "adduce many reasons for regarding its listing as unreliable, out of date, and so on" (75). So, I would invoke CR to decide between Sloppy Joe and BIP. Mayo is wrong that "According to Musgrave, even "Sloppy Joe's Books in Print" would be as acceptable as the authoritative BIP!" (75).

Mayo also discusses another epistemic principle that I invented to show, contrary to Nozick, that "self-subsumption is not a virtue" (ERR, p. 331). That was "It is reasonable to believe anything said in a paper by Alan Musgrave". (Lest I be accused of hubris, I have never *said* this in any paper of mine. I merely *mentioned* it as an example of a crazy principle! It was a use-mention confusion for me to call it 'self-subsuming'.) Mayo asks me on what grounds I think this principle crazy. Critical rationalist grounds, of course: the principle could not withstand obvious criticism. Mayo seems to think that I have committed myself to the view that all self-subsuming principles are on a par and equally acceptable. This is precisely what I was concerned to deny: saying that two principles are alike in being self-subsuming is not saying that they are on a par or equally acceptable.

4. OTHER CRITICAL METHODS

Returning to the central issue, must a method for determining what to believe be shown to be reliable, before it can be reasonably adopted? As explained, critical rationalism denies this. **Volker Gadenne** is a critical rationalist, but he leaves it behind when he ascends to the meta-level, and asks about the rationality of methods themselves (including critical rationalist methods). He writes:

> But let us assume that A is the goal of science and our task is to decide whether procedure M should be recommended with respect to A or not. In this case, a rational person will recommend M if and only if he or she *believes that M contributes to A,* or that *M gives us a greater chance to achieve A.* And this belief is reasonable if there are good arguments in favour of the hypothesis that M is the best we can do to achieve A. (101)

My quarrel is only with the last sentence here, which suggests that rational belief in the meta-hyothesis MH ("M is the best we can do to achieve A") requires good arguments in favour of MH. This is meta-level justificationism. If we reject it, what is needed for rational belief in MH is that MH has withstood criticism, that there are no good arguments against it. Or so critical rationalism, applied at the meta-level or methodological level, maintains.

I suspect that Gadenne would be sympathetic with this. He goes on to say, rightly in my view, that any attempt to 'guarantee' that a method will be successful is "too strong for a methodology that is committed to fallibilism", whereas merely to hope that a method will be successful is not enough. We need something "less than guarantee but stronger than mere hope" (101). I suggest that having withstood criticism, there being no good arguments against it, is precisely what Gadenne needs here – it is more than mere hope, yet no guarantee. Here, by the way, I agree with Gadenne and others who urge that methods or methodological rules are not mere conventions, if by 'convention' we mean something that cannot be rationally assessed. As Gadenne says, "We can argue for [the adoption of] methodological rules, in the same way we argue for [the adoption of] other … assumptions" (102).

Like Gadenne, **Howard Sankey** seems to reject justificationism at the level of first-order belief, only to insist upon it at the level of method. Sankey correctly reports that the rejection of justificationism lies at the heart of critical rationalism (116-118). He concedes that:

> The critical rationalist account of theory acceptance is … of clear relevance to the problem of method and truth. For the critical rationalist asserts that survival of critical scrutiny provides the basis for rational belief in the truth of scientific theories. (118).

And yet Sankey insists that:

> An explanation is therefore required on the part of the [critical] realist of why certification by method provides warrant with respect to truth. I will refer to the need to provide such an explanation as the problem of method and truth. (109)
>
> By itself, the rejection of justificationism does not suffice to solve the problem of method and truth. If truth is non-epistemic, and the critical method is the basis of theory acceptance, the connection between method and belief in the truth is left entirely unexplained. (119)

But the connection between critical method and (rational) believing is precisely what critical rationalism explains. What is going on? What exactly is the problem of method and truth, left unsolved even if we reject justificationism? Are we being asked for an explanation of why theories certified by the critical method are true? Or are we being asked for an explanation of why theories certified by the critical method are rationally believed? Sometimes it is the latter: "Musgrave … must confront the question of why it is rational to believe theories certified by the methods of science" (109). Sometimes it is the former: "What bearing does method have on truth? Why should the use of method lead to theories that are … true …? This is the problem of method and truth." (109)

Sankey's formulations systematically conflate beliefs and believings, in line with justificationism. One example will suffice, although many could be given:

> But if it does not follow from survival of criticism that a theory is true [this concerns beliefs], then neither does it follow that the theory is to be accepted as true [this concerns believings, or 'acceptings'] (119)

In fact, Sankey does not reject justificationism, but assumes it, both at the level of particular belief and at the level of method.

Sankey has another worry about critical rationalism: "The trouble is that nothing has been done to secure belief in truth as the unique mode of theory acceptance" (119). He considers inference to the best explanation (IBE) and its meta-instance, the Miracle Argument for Realism (MA). Having displayed my deductivist formulation of IBE, he says:

> The question is why it is reasonable to accept the best explanation *as true*. Might it not be equally reasonable to accept the best explanation as empirically adequate …? (116)

My answer to this question is NO. Suppose that H is the best explanation that we have of some phenomena. The T-scheme says: H if and only if it is true that H. Given the T-scheme, to believe that H and to believe that H is true are the same. Given the T-scheme, to accept that H and to accept that H is true are the same. So what is it to "accept H as empirically adequate"? It is not to accept H, for this is the same as accepting that H is true. Rather, it is to accept a meta-claim about H, namely the meta-claim "H is empirically adequate" or equivalently "The phenomena are as if H were true". Call this meta-claim H*. Now, and crucially, H* is no explanation at all of the phenomena that (we are assuming) H is an explanation of. "It is raining" explains why the streets are wet, but "The phenomena are as if it is raining" does not. At least, H is a better explanation than H* is. So according to IBE, H* should not be accepted as true. That is, according to IBE, H should not be accepted as empirically adequate.

I wonder what part of this argument Sankey and others who think like him will reject. Not IBE – at least, they pretend to accept IBE. Not, presumably, the T-scheme. Not, presumably, its consequence, that to accept H and to accept H as true are the same thing. Not, presumably, the equivalence of "H is empirically adequate" and "The phenomena are as if H were true". Not, presumably, the claim that H is a better explanation of the phenomena than "The phenomena are as if H were true".

This does not refute constructive empiricism. But it does refute the claim that constructive empiricists can happily accept IBE and give a constructive empiricist reading of it. If you try to recast IBE in terms of empirical adequacy rather than truth, you end up with something quite incoherent. Your major premise says that it is reasonable to accept the best explanation as empirically adequate. But this is to accept something that is no explanation at all! The fact is that realism and explanation go hand-in-hand. It is no accident that down the ages antirealists have pooh-poohed the idea that science explains things.

The same applies to the Miracle Argument (MA), a meta-instance of IBE where the fact to be explained is not a fact about the world but a fact about science, the fact that some scientific theories have (novel) predictive success. The truth of a theory T is a better explanation of T's (novel) predictive success than "T is empirically adequate" or "The phenomena are as if T were true". Or, putting the point in terms of empirical adequacy, truth explains empirical adequacy, but empirical adequacy does not explain itself. Given MA, a meta-instance of IBE, it follows that it is reasonable to accept that T is true.

Sankey comes up with a "constructive empiricist version of critical rationalism" according to which "theories which survive criticism are to be accepted as empirically adequate" (120). This is a possible view. What is not possible is to combine this view with IBE. But the trouble with IBE, as I construe it, is that the best explanation is not shown to be true (just reasonably accepted as true). And the trouble with critical rationalism is that theories which survive criticism are not thereby shown to be true either (just reasonably accepted as true). Notoriously, the same trouble arises for Sankey's critical constructive empiricism: theories which survive criticism are not thereby shown to be empirically adequate (just reasonably accepted as such). Which brings me back to Sankey's problem of method and truth.

How does Sankey solve his problem? He says that epistemology must rest on metaphysics (111), and wheels in some metaphysics. Well, resting epistemology on metaphysics is a welcome change from resting metaphysics on epistemology. That was the basic mistake of idealists down the ages – including that version of idealism known as 'internal realism', which Sankey discusses and joins me in rejecting. As he rightly sees, the internalist "closes the gap between method and truth" (113) by going for an epistemic theory of truth. But such a theory, if combined with the T-scheme, closes the gap between method and reality as well. The world comes to "depend on our methods of inquiry or our theories in idealist fashion" (114). But, Sankey rightly says, "that is evidently not something that a realist can accept" (114).

I wish that Sankey would complete the separation of epistemology from metaphysics. That is precisely what critical rationalist epistemology does, with its rejection of justificationism. As he himself puts it, when introducing his problem of method and truth:

> But matters of method and rationality are separate matters from those of reality and truth. This is especially the case from the perspective of realism. …Reality is not subject to determination by human thought. This remains the case even if the belief that the world is a given way is a belief that is rationally justified. For one may rationally believe what is false. The point applies with equal force to scientific theories certified by the norms of scientific method. A theory that is certified by the norms of scientific method is not thereby shown to be true. A theory which satisfies methodological norms may yet be false. (108)

The critical rationalist proposal is that a theory that is 'certified by the norms of scientific method' is rationally believed, despite the fact that it is not thereby shown to be true. Of course, from the point of view of justificationism, this is an absurd proposal – any reason for believing must be a reason for what is believed. But IF justificationism is wrong (a big 'if', of course), THEN Sankey's 'problem of method and truth' is solved – or to be more precise, the problem of method and rational belief (in truth) is solved. In fact, if we reject justificationism at the meta-level, Sankey's 'problem of method and truth' actually becomes two problems. First, there is the problem of whether 'certification by critical methods' shows truth. Second, there is the problem of whether 'certification by critical methods' shows rational belief (in truth). Critical rationalism solves the second problem, but not the first. According to critical rationalism, we cannot show of any method that it yields truths infallibly, or even that it yields truths reliably (yields more truths than falsehoods). At least, we cannot show this unless we argue in a circle, or set off on an infinite regress. For what we have here is, basically, the problem of induction yet again.

We can try to solve the problem of induction by wheeling in a metaphysical principle like "Unobserved cases resemble observed cases". But where does this come from – inductive reasoning (the circle), or some other metaphysical principle (the regress)? Similarly, we can solve the problem of method and truth by wheeling in a metaphysical principle like "Scientific methods yield truths". But where does this come from – scientific methods (the circle), or some other metaphysical principle (the regress)? Nothing is changed if we weaken the metaphysics, and say "Unobserved cases resemble observed cases more often than not" or "Scientific methods yield more truths than falsehoods".

Something like the last principle is Sankey's own solution to his 'problem of method and truth'. He claims that "the rules of method are reliable means of promoting the realist aim of truth" (122). This means, I take it, that most theories certified by the rules of method are true. How do we know this? From empirical or scientific inquiry, "For it is an empirical matter whether use of a particular method reliably conduces to a given cognitive goal" (122). Now suppose we could somehow show that most theories certified by the rules of method *up until now* have been true. Would this show that the rules of method are reliable? Not without inductive reasoning, not without invoking precisely the 'rules of method' Sankey is supposed to be justifying. Would it show that it is reasonable to believe that the rules of method are reliable? Yes, provided we abandon justificationism and adopt critical rationalism.

Do not mistake me. I do not object to the empirical or scientific study of our cognitive apparatus (perception, rules of scientific method, or whatever). And I do not object to pointing out ways in which the results of such studies may cohere with or mutually support the results of studying the world using that cognitive apparatus. I once did a bit of this myself, when I made the simple-minded suggestion that if the theory of evolution is to be believed, then perception is a reliable process. This suggestion may be criticised in all sorts of ways – never mind that. The relevant point here, as I immediately pointed out when I made the suggestion, is that circularity looms. If the theory of evolution is to be believed, perception is reliable. And why is the theory of evolution to be believed? Because of the evidence in its favour, got through reliable perception.

Sankey discusses the views of Kornblith about natural kinds, which come from the same stable as my simple-minded suggestion. But what does Kornblith's suggestion come to? If science is to be believed, then there are natural kinds in nature. And if there are natural kinds in nature, then science is to be believed. More generally, given some metaphysical M, science is true or mostly true. And how do we know M? Why, science teaches us that M is so. M is not some science-transcending metaphysical 'sky-hook' (to borrow Dennett's term), more in need of justification than the science it is supposed to justify. The only escape from this justificationist circle is to free epistemology from metaphysics, and adopt critical rationalism. The problem of method and rational belief (in truth) is solved by critical rationalism. The problem of method and truth is insoluble, for much the same reason as the traditional problem of induction is insoluble.

But what if the empirical or scientific study of science were to show that most theories so far 'certified by the rules of scientific method' were false? What if it could be shown that inference to the best explanation (IBE) has so far led folk to believe more falsehoods than truths? What if it could be shown that believing the hypothesis that best withstands criticism has so far more often than not been believing falsehoods? Would such results not constitute severe criticisms of the rules of scientific method, IBE, and critical rationalism in general? Yes, they would. As I already said, if a method for forming beliefs is shown to be unreliable, it is no longer reasonable to persist with it. But the critics of critical rationalism and critical realism have got nowhere near showing any of these things, as we will see later.

If somebody can propose non-critical non-realist methods that withstand criticism better, then we should adopt them. But the principles advocated here are very general, and I can think of no better alternative to them:

> Unless you have a specific reason not to, trust your senses.
> Unless you have a specific reason not to, believe what other folk tell you.
> Prefer that evidence-transcending hypothesis that best withstands criticism.
> Accept the best available explanation of any puzzling phenomenon.

Having no beliefs is not an alternative. Having beliefs but thinking them irrational is just irrationalism. Saying that you believe things but do not think them true is just a linguistic confusion.

Stathis Psillos says that the rejection of justificationism "is exactly right" (140). Yet he joins Sankey in thinking that it is incumbent upon the scientific realist to show that scientific methods are reliable. For Psillos, these methods are "ampliative-abductive methods" (more on this in a moment). The realist must show that "the ampliative-abductive methods employed by scientists … are reliable: they tend to generate approximately true beliefs and theories" (134). How is the realist to show this? Psillos says that "The reliability of abductive reasoning is an empirical claim, and if true, it is contingently so." (135). Presumably, then, we are to employ the ampliative-abductive methods of science to show that the ampliative-abductive methods of science are reliable.

This is obviously circular. But in an attempt to disarm this objection, Psillos introduces a distinction between premise-circularity, which is vicious, and rule-circularity, which is not (135; see also his 1999, pp. 81-90). My problem with this is that non-vicious rule-circularity turns into vicious premise-circularity, given Psillos's own admission that an invalid ampliative inference can always be turned into a valid deductive inference by the addition of suitable missing premises (140). Somebody notes that all observed As were Bs, and infers that all As are Bs. We protest that this is invalid. To disarm this objection, the extra premise that unobserved cases resemble observed cases is wheeled in. If we ask what justifies this extra premise, we are told that it is inferred from the fact that unobserved cases have resembled observed cases so far. Enter Psillos. Though we *can* wheel in the extra premise, whereupon vicious premise-circularity befalls us, we *should not* wheel it in. Instead, we should say that inductive generalisation is not deduction at all, but an ampliative method, which proceeds according to an ampliative rule whose use can be vindicated by employing that very rule. This rule circularity is not vicious. Similarly with other ampliative-inductive methods of science. We use them to show that they are reliable, in a non-vicious rule-circular way. How is a problem solved by refusing to state explicitly what you admit might be stated explicitly?

In the course of his earlier discussion of these matters, Psillos wrote something with which I entirely agree:

> If one knew that a rule of inference was unreliable, one would be foolish to use it. But this does not imply that one should first be able to prove that the rule is reliable before one [non-foolishly, wisely, reasonably?] uses it. All that is required is that one should have no reason to doubt the reliability of the rule … . But we have no such reason. (1999, p. 85)

Having no reason to doubt P does not show P to be true – but it does show that it is reasonable to think that it is. Having no reason to doubt that R is reliable does not show that R is reliable – but it does show that is reasonable to think that it is. Or so critical rationalism assumes.

Chief among the ‘ampliative-abductive methods of science’, according to Psillos, is IBE. I regard IBE as a valid deductive scheme, instances of which might be sound as well. First, taking a leaf out of Peirce’s book, I add an epistemic modifier to its conclusion. Then I add an epistemic principle to its premises so that it becomes a valid deductive scheme. This deductivist reconstruction simply makes explicit what is being implicitly assumed.

Psillos has no problem with this, but he does object to the deductivism that motivates it (132). His objections are familiar ones. He says that human reasoning is content-increasing, while deductive reasoning is not. But the undoubted fact that human thought is content-increasing, that I may think more today than I thought yesterday, does not mean that yesterday’s thoughts were the premises from which I reasoned to today’s thoughts. There is more to thinking than reasoning or arguing.

Again, Psillos says that human reasoning is defeasible, sensitive to new information, evidence and reasons, while deductive reasoning is not. The defeasibility of ampliative reasoning means that so-called ‘ampliative logic’ becomes an empirical science – new information, evidence and reasons can show you that an argument you thought valid (cogent) was not. Besides, valid deductive reasoning is also defeasible: new information, evidence and reasons may show that some explicitly-stated premise is false.

Finally, Psillos complains that deductive reasoning cannot establish the truth of its premises – for that we need ampliative or non-deductive reasoning. Deductive reasoning must stop somewhere, on pain of infinite regress – whereupon ampliative reasoning takes over. But ampliative arguments have premises, too, so there is no stopping the regress with their help. Besides, the assumption here seems to be that everything we assert is the conclusion of some argument or other. This logomaniac assumption is absurd: human reasoning requires premises that cannot themselves be conclusions of arguments, on pain of infinite regress.

The bulk of Psillos's paper contains an attempt to rebut Colin Howson's criticism of the No Miracles Argument (NMA), to the effect that it commits the 'base-rate fallacy'. The version of NMA that Howson criticises is the traditional one, where it is the truth or high probability of the best explanation that is supposed to be established. Since my version of NMA concludes neither that the best explanation is true nor that it is probably true, my version is immune to this criticism. Moreover, my version is deductively valid, and involves no fallacy.

Still, the question remains whether NMA (or IBE) are reliable methods of forming beliefs. Psillos thinks it incumbent upon realists to show that they are (which I deny), and he thinks that they can show it. He regards the following thesis as constitutive of realism:

> *The Epistemic Thesis*: Mature and predictively successful scientific theories are well-confirmed and approximately true. So entities postulated by them, or, at any rate entities very similar to those postulated, inhabit the world. (133)

This is not an epistemic thesis at all – it is a metaphysical thesis. It says (cutting out the complications) that mature science is true so that the entities it postulates exist. Next he says "What is worth stressing is that Musgrave takes NMA to aim to tell in favour of the *Epistemic Thesis* …" (138). It is worth stressing that this is wrong. As I see it, "NMA makes it reasonable to accept that truth has been achieved" (138), but it does not show that truth has been achieved (which is what the so-called 'Epistemic Thesis' says). Justificationists assume that a reason for accepting something as true must show that it is true. That is precisely what I deny.

Nor do I claim that most theories whose acceptance is licensed by NMA are true – which is where the base-rate fallacy comes in. I do not even claim that most theories *up until now* whose acceptance was licensed by NMA were true. (To get from the latter to the former, to the reliability of NMA, would involve, of course, a simple inductive leap. How *could* a critical rationalist assert the reliability of CR, without engaging in inductive reasoning? How could a critic of critical rationalism assert the unreliability of CR, without engaging in pessimistic inductive reasoning?)

Never mind this. I agree with much of what Psillos says in defence of NMA. I read his arguments as showing, not that NMA is reliable, but that its critics have not shown it to be unreliable. More generally, have the critics of critical methods shown that they are an unreliable way of forming beliefs? I do not think so. Some critics have claimed that all scientific theories are false, usually by arguing according to the pessimistic induction – "All past theories were false, therefore all theories are false". But the premise of this argument is preposterous, and the conclusion does not follow. Nor does confining attention to 'big theories' ('global theories', 'paradigms', 'world-views', 'research programmes', etc.) help. Saying "Aristotle was wrong, Galileo was wrong, Kepler was wrong, Newton was wrong, Einstein is likely wrong" gets nowhere near to showing that all the 'little theories' (including those that follow from the big ones) are wrong. Positivists argue that all theories that postulate unobservables are false, because there are no unobservables. But why should we accept the positivist, human-chauvinistic premise of this argument? But I have discussed all this elsewhere, and will say no more here.

Save for one point. Antirealists like to point to past theories that were partially successful, but which then were falsified. Obviously, the partial success of falsified theories cannot be explained by their truth. How then to explain it? And if partial success is explained by something other than truth, why cannot total success be explained by something other then truth as well? Here Psillos goes in for truthlikeness or verisimilitude, which is (notoriously) problematic. Why not go in for partial truth instead? This is not the same as truthlikeness. Despite its falsity, "All swans are white" is predictively successful in Europe, and bird-watchers find it useful to employ it there. I do not know how close to the (whole) truth "All swans are white" is, and none of the captains of the verisimilitude industry can tell me in less than 100 pages of formulas. I do know that it has a true part, "All European swans are white", whose simple truth explains the success European bird-watchers have. Psillos qualifies NMA as follows:

> … realists should *refine* the explanatory connection between … predictive success, on the one hand, and truthlikeness, on the other. They should assert that these successes are best explained by the fact that theories which enjoyed them have had *truthlike theoretical constituents* (i.e., truthlike descriptions of causal mechanisms, entities and laws). The theoretical constituents whose truthlikeness can best explain empirical successes are precisely those that are essentially and ineliminably involved in the generation of predictions … (135)

The point is important, I agree. But why cannot we replace 'truthlikeness' with 'truth', and 'truthlike' with 'true'? Then NMA could be qualified in the following very simple way: truth is the best explanation of total empirical success; partial truth is the best explanation of partial empirical success.

Science grows out of commonsense, and scientific methods grow out of commonsense methods. We use critical methods, and IBE, all the time, and most of the time they serve us very well indeed, in everyday life and in science. Armed with these methods, we have come up with innumerable rational beliefs that are true (and some that turned out to be false).

5. THE METAPHYSICS OF CRITICAL REALISM

I have so far been resisting the idea that critical rationalist epistemology must be *underpinned* by some metaphysical assumption. Still, that leaves open the question of what metaphysics or ontology a critical rationalist/realist will adopt. Obviously, scientific realists will have no truck with the positivist idea that only things that happen to be observable by humans exist. Scientific realists will be realists about the things postulated by our best scientific theories, whether they happen to be observable by humans or not. Scientific realists believe in what **Michael Redhead** calls 'the Unseen World'. I agree with pretty well everything Redhead has to say about this. I have a few nit-picks. I wish Redhead had not floated the old, discredited idea that it is not the table we see, but "light reflected off the table, or … electrical stimulation in the retina caused by the light, or … " (135). We do not see light, or electrical stimulations somewhere in our heads caused by light. These things are part of the Unseen World. The science of vision posits them in order to explain how we see the table. To say that we observe things "by the *effects* they produce" in us (157) is not to say that we only observe the effects.

Several of my critics want me to go further. Alan Chalmers wants me to be a realist about *essences*, Robert Nola wants me to be a realist about *structures*, and Mark Colyvan wants me to be a realist about *numbers and other abstract platonic objects*. I am not so sure. My realist metaphysics is explanatory or more generally problem-solving. I believe in things that do explanatory or problem-solving work for us. So my question is, what explanatory or problem-solving work is done with essences, structures, and platonic objects?

Alan Chalmers wants me to become an essentialist. He says there are two problems I cannot solve that essentialism can solve – "straightforwardly" (163). First, there is the problem of explanatory asymmetries: some scientific deductions are explanatory and some not. Second, there is the problem of distinguishing genuine laws from accidentally true generalisations. The problems seem connected. One appealing thought is that a deduction is explanatory if a genuine law figures in its premises, non-explanatory if an accidentally true generalisation figures in its premises. Thus, if we can solve the second problem, we can solve the first problem as well. This appealing idea is hinted at by Chalmers when he says:

> Appeal to the law governing the expansion of metals can help explain why the bottle top is loosened when held under the hot water tap, whereas 'all the coins in my pocket are silver' cannot help to explain why any one of them is silver. (164)

But this will not do – laws figure in non-explanatory deductions as well as explanatory ones. Chalmers' own example proves the point:

> (The deduction of the range of a projectile from Galileo's laws of motion plus initial conditions explains why the projectile has that range, but the deduction of the height of a cliff from those laws plus the time of fall of a stone from top to bottom does not explain why the cliff has that height.) (163)

The example is quite typical. Scientific laws are typically functional dependencies, rather than universal generalisations of the form "All As are Bs". And a functional dependency can always be manipulated to yield a deduction that is intuitively non-explanatory. The loosening of the bottle top does not explain why it was held under the hot water tap.

That is why I suggested a different answer. A deduction is explanatory if its initial conditions specify the cause of the event to be explained, non-explanatory if they do not. But to make this answer good, we need an independent account of causality, one that will make good our intuitions that the time it took the stone to fall did not cause the cliff to have a certain height, that the loosening of the bottle top did not cause it to be held under the hot water tap (though the antecedent desire to loosen it might have), and that being in my pocket does not cause a coin to be silver. The principle that an effect cannot temporally precede its cause suffices in these cases.

What about the second problem, distinguishing genuine laws from accidentally true generalisations? Chalmers says that laws are "necessarily true, as opposed to accidentally true" (164). Now the 'necessity' spoken of here is not logical necessity, truth in all possible worlds, but rather physical necessity, truth in all physically possible worlds. And what are physically possible worlds? They are worlds in which the laws of nature obtain. The circle is complete: laws as opposed to accidental generalisations are physically necessary, true in all physically possible worlds, and a physically possible world is a world in which the laws are (accidentally?) true. Possible worlds are no help, as Chalmers says (164).

Can counterfactuals help? What about the idea that genuine laws support counterfactuals, whereas accidental generalisations do not? To make this idea work, we need accounts of the truth-conditions of counterfactuals and of the 'support' relation. The favourite account of the former is David Lewis's. To find out whether "If it were the case that A, then it would be the case that B" is true, inspect the closest possible world(s) to ours in which A is true and see if B is true in that world as well. And what world(s) are closest to ours? Why, worlds in which our laws of nature are true. The circle has got bigger, but it is still a circle. Possible worlds are no help once more.

I favour the idea that counterfactuals are elliptical statements of logical consequence: "If it were the case that A, then it would be the case that B" is true just in case A, together with some unstated premise(s) C that are true, entails B. Given this, to say that a generalisation G supports the counterfactual "If it were the case that A, then it would be the case that B" is just to say that G and A entail B. But this simple (simple-minded?) view means that accidentally true generalisations support counterfactuals just as well as laws do. We are still stuck.

Still, this theory of counterfactuals can illuminate the status of 'ideal laws', laws about 'ideal entities' that do not exist. These are a problem for realists, who seem forced to say that any such 'law' is false if construed as having existential import or vacuously true if construed as having no existential import. Neither option is satisfactory: the ideal gas laws are important, non-vacuous (and non-accidental) truths. I suggested that the ideal gas laws are true counterfactuals, thought neither vacuously nor accidentally true because they are supported by the kinetic theory of gases. Chalmers objects (167) that the ideal gas laws were empirically supported, independently of their derivation from the kinetic theory, by regarding ideal gases as the limit to which real gases tend as pressure is reduced. Quite so. I agree that experiments convinced physicists that the ideal gas laws were not mere vacuous truths. But it took the derivation from kinetic theory to convince physicists that they were not just accidentally true. The kinetic theory told them that if any gas were to be ideal, then it would obey the ideal gas laws. But what about the laws of the kinetic theory itself? If these are only true by accident, so are the counterfactual ideal laws that they support.

Can essences help us to break out of the circle? Chalmers says that "there is something about the nature of metals that makes it physically necessary that they expand when heated" (164). Metals by their nature have ontologically basic dispositional properties, powers and capacities, and precise statements of these dispositional properties or modes of acting and interacting are genuine laws of nature. If you ask why metals *must* expand when heated, the answer is that it is in their nature to do so. They would not be metals if they did not do so.

Chalmers correctly reports me as having two reservations about essences. First, are explanations in terms of essences to be regarded as *ultimate*? Is there no explaining why metals expand when heated, except to say they would not be metals if they did not? Second, is knowledge of the essences of things a priori, as is suggested by talk of *defining* things by their essences? Chalmers says that neither reservation applies to scientific essentialism, as defended by Brian Ellis. I am not so sure.

Chalmers insists that essences "needed to be discovered, not merely defined" (170), and that the "adequacy of our essentialist definitions needs to be established empirically" (170). But how do we find out empirically that a metal *must* expand when heated, because this is part of its essence, as opposed to finding out that all metals *do* expand when heated? "Metals must expand when heated" entails, but is not entailed by, "Metals expand when heated". There is no independent *evidence* for the stronger statement. Nor is the stronger statement more refutable than the weaker one – indeed, it does not seem refutable at all! Chalmers says that our characterisation of the essential properties of metals "may or may not correspond to what they actually are" (170). But he also says that anything that lacks an essential property of Xs is not an X (168). So we cannot discover a *metal* that fails to expand when heated – anything we find that fails to expand when heated was not a metal in the first place. Empirical methods seem powerless here. The most they can establish is that all metals do expand when heated, not that they must do so as a matter of physical necessity. Nor could we ever empirically refute "It is an essential property of metals that they expand when heated". Essentialism is unconfirmable and irrefutable metaphysics, not physics.

Something has gone wrong here. On the one hand, "Metals must expand when heated" seems irrefutable. On the other hand, it entails "Metals expand when heated" which is refutable. But if the latter is refutable, so is the former. Here is a diagnosis of what has gone wrong. If we stipulate that the term 'bachelor' means unmarried man, this cannot be refuted by finding a married bachelor. Similarly, if scientists stipulate that the term 'metal' is to be (partially) defined as something that expands when heated, this cannot be refuted by finding a metal that does not expand when heated. Statements of the essences of things (real or essentialist definitions) are irrefutable because they are really verbal stipulations (nominal definitions).

How plausible is this diagnosis? Scientists found out empirically that all metals expand when heated. They came to take this for granted in their future researches, and they sought to explain it. Suppose they went further, and decided that henceforth the term 'metal' was to be (partially) defined as something that expands when heated. This verbal stipulation marked a change in the meaning of the term 'metal'. To mark that change pedantically, we might say that tin was once called a 'metal', now it is called a 'metal*'. The new stipulation makes nonsense of the old empirical researches that established that metals expand when heated – those researches must be described using the old term 'metal', not the new term 'metal*'. The stipulation seems to turn a contingent truth into a necessary one. But this is an illusion that stems from overlooking the change in meaning. There is not one truth here, once thought merely contingent, now discovered to be 'physically necessary'. "Metals expand when heated" was and remains contingent. "Metal*s expand when heated" was and remains necessary, but only verbally or conceptually so.

Now I do not object to changes in language of this kind, brought about by new verbal stipulations. I do object to reading essentialist metaphysics off them. Words are one thing, things another. Our stipulations about what our words shall mean yield necessary truths – but the necessities they yield are logical or conceptual or verbal necessities, not real or natural or physical necessities. Nor can these verbal necessities do any explanatory work for us. The necessary truth "Metal*s expand when heated" does not entail the contingent truth "Metals expand when heated", let alone explain it or bestow necessary status upon it.

I said that I do not object to (partially) defining a metal as something that expands when heated. Or more precisely, since we do not define things but rather words, I do not object to (partially) defining the word 'metal*' to stand for things that expand when heated. This is just to make the word 'metal*' a *dispositional* word. Dispositional words have necessities built into them from the start – but the necessities are logical or conceptual. The meaning of an ordinary dispositional word involves from the outset a generalisation about behaviour. Something is brittle if it will break when struck by something else, something is soluble in water if it will dissolve when placed in water, and so on. The generalisations are, of course, vague and rough-and-ready bits of 'folk science', which need specification and refinement. No matter. The key point here is that such statements as "Something is brittle if it will break when struck by something else" are conceptual truths, matters of logical or conceptual necessity, not matters of fact. We do not discover empirically that they are true, by inspecting brittle things and checking whether they break. Just as we do not discover empirically that "Bachelors are unmarried men" is true, by inspecting bachelors and checking whether they are unmarried. Nor can we explain why something broke by saying that it was brittle. Just as we do not explain why some bloke is unmarried by saying that he is a bachelor.

Chalmers wants me to adopt Brian Ellis's version of essentialism. This claims that the essences of things need not be monadic properties (how the things are in themselves, without relation to other things), but may also be dispositions, powers, capacities, "how they are disposed to act and interact with other objects" (168). This is a welcome change from the usual essentialism, which insists that essential properties must be intrinsic or monadic properties. Now suppose that essences are dispositions, and consider the following:

> A charged body will attract or repel other charged bodies, give rise to a magnetic field when moving and radiate when accelerating because it is in the nature of charged bodies to do such things. Precise statements of these modes of acting, such as Coulomb's law or the Lorentz force law, describe the laws of nature. They are not something imposed on charged bodies because they are already implied in what it is to be a charged body. So charged bodies necessarily obey the laws that they do. (168-9)

If 'charged' is a dispositional term (which I will not dispute), then the laws here specified are *logically or conceptually* implied by describing something as 'charged'. It is logically or conceptually necessary that charged bodies obey those laws. (Which is not the same as saying that "charged bodies necessarily obey the laws" – see below). Nor do we explain why some body obeys the laws by saying that it is charged.

Why do we think that "Metals expand when heated" is not just 'accidentally true', like "All the coins in my pocket are silver"? The answer lies, I believe, in their causal explanations. We can explain why metals expand when heated in terms of deeper regularities, which reveal to us how heating up a piece of metal causes it to expand. There is no such explanation in the case of the coins in my pocket – we know that being put in my pocket does not cause a coin to be silver. But given what we know about metals, they *must* expand when heated.

But there is a fallacy here. Suppose there is an explanation, and a deductive explanation to boot, of why metals expand when heated. Let its explanans (whatever it is) be E, and its explanandum G. If the deduction of G from E is valid, we can say "Necessarily, if E then G". Here, of course, the necessity spoken of is logical. We can also misplace the word 'necessarily' and say "If E, then necessarily G". Now the necessity spoken of cannot be logical, since G is not a logical or conceptual truth. So is the necessity of G non-logical or 'natural' or 'physical'? No: this route to physical necessity is just misplaced talk of logical necessity, where the misplacing stems from Stove's fallacy of misconditionalisation. "Given what we know about metals, they must expand when heated" misconditionalises on logical necessity (Necessarily, if E then G) to yield physical necessity (If E, then necessarily G).

What about the coins in my pocket? I said that putting a coin in my pocket does not cause it to be silver, which is why we think "All the coins in my pocket are silver" is just 'accidentally true'. It is not that there is no explaining why all the coins in my pocket happen to be silver. But the explanation of it involves a rare confluence of independent causal chains. It happens 'by accident' in Aristotle's sense. We pronounce a verdict of 'accidental death' on a person killed by a loose brick falling from a building as he walks beneath. We can explain the death, by explaining why he walked there when he did and why the brick worked loose just then and fell. The accidental death is not uncaused or random, it arises out of a rare confluence of independent causal chains. Similarly with the coins in my pocket all being silver.

Of course, no misconditionalisation is involved in saying "Metals must expand when heated" if the deeper explanatory principles E are themselves natural necessities. Then the explanatory deduction does establish (as a matter of logical necessity) that "Metals expand when heated" is also a natural necessity. Now we are saying that what distinguishes genuine laws from accidental generalisations is that genuine laws follow from genuine laws, whereas accidental generalisations do not. This is hardly illuminating!

Does it become more illuminating if we can make out independently that the *ultimate* explanatory principles, the places where the explanatory buck stops, are not just accidentally true? There might, I grant, be ultimate laws – the 'Russian Doll' model of the universe might be false. But why cannot the ultimate laws be ultimate contingencies? There might, I grant, be fundamental particles. But why must fundamental particles have all their properties essentially?

I would quite like to be an essentialist and to believe in necessities in nature. My problem is to find a version of these doctrines that does not stem from two sources – the unholy alliance with definition, and the fallacy of misconditionalisation. My problem is to find a version of these doctrines that does not project necessities of language and/or logic onto the world. My problem is to get over the 'positivist' principle (prejudice?) that the only necessity is logical necessity.

I confess that I cannot understand Ellis's claim that "physical necessity is a species of logical necessity" (177). Physical necessity is in the world (if it is anywhere), logical necessity in our world-representations. To say that the former is a species of the latter just seems to be projecting necessities of language and/or logic onto the world, in the ways I have tediously explained. I do not object to a 'modest essentialism' whereby an essential property of an object is not one that it must possess, so that it would not be the object that it is without that property, but rather a particularly important property possession of which determines what the object does (and identification of which helps us to give causal explanations of its actions and interactions). 'Essentialness' in this modest sense obviously comes in degrees. And it has more to do with dispositional properties than monadic ones. Why cannot objects be 'bare particulars', not in the sense that they have no properties, but just in the sense that they have none of their properties essentially? Why cannot we stop asking "What is X?" (Popper called these 'essentialist questions') and ask instead "What does X do?"?

Which brings me to structures. What are they? Should realists believe in them? And what about *structural realism*, which is (roughly) the view that structures are the only things that realist should believe in? **Robert Nola** discusses these questions, and I agree with pretty well everything he says. I would say some of it more bluntly. Like Redhead, I think it bizarre to say that "it is only structure which really exists" (157). Nola calls this loopy view *Platonistic ontological structural realism* – "all that exists are mathematical structures … there do not exist placeholders, such as objects, within the structures" (182). To my mind, a structure must be a structure *of* something. My house is a structure made of house-bricks, and it is loopy to say that the structure exists but the bricks don't. Similarly, the bricks are made of molecules, and it is loopy to say that the bricks exist but the molecules don't. Reverse loopiness is just as bad. It is also loopy to say that the house-bricks exist but the house doesn't, or that the molecules exist but the bricks don't. Eddington started all this with his tale about the 'table-of-physics' and the 'table-of-common-sense', and his silly question "Which table is the real table?" I wish Redhead had not mentioned Eddington with approval (155). I have discussed all this elsewhere, and will say no more here.

Redhead and Nola think it less easy to dispose of what Nola calls (182) *epistemic structural realism* – "we can have knowledge of structures but we cannot know the items that are placeholders in such structures (such as objects); they are a "something-we-know-not-what"" (182; also Redhead, 157). The obvious response to unknowable Kantian *ding-a-ling-an-sich* is that if we know a lot about a structure, then we *ipso facto* know a lot about how the placeholders in that structure relate to one another. Yes, comes the stock reply, but knowing how the placeholders relate to one another is not knowing what their *true natures* are (Redhead, 157), what they are *in themselves*, what they are *intrinsically*, what they are *essentially*. And we are deluged with a heady mixture of bad essentialist metaphysics and/or bad philosophy of language, much of the former fuelled by the latter.

There is supposed to be a difference between knowing truths about Xs, about what they do, about how they relate to or interact with other things, and knowing about the *nature* of Xs. But what is the difference? Must truths about the nature of Xs concern only intrinsic or monadic properties of Xs? That seems wrong – truths about how beams of light behave are truths about the nature of beams of light. Perhaps the thought is that truths about relational properties of Xs cannot tell us the essential nature of Xs. At which point the so-called 'dispositional essentialist' wonders why relational or dispositional properties of Xs cannot be 'essential' to them – whatever light is, it would not be light unless beams of it obeyed this or that law.

I think that Nola shows convincingly in his paper that the problem which structural realism is supposed to solve is actually a pseudo-problem. The problem is so-called 'ontological discontinuity': as theories change, ontology changes. Ontological discontinuity cannot mean that the contents of the world change as our theories about the world change. To credit thinking or talking with such transformative powers would be a "virulent form of human chauvinism" (184 – I wish I had coined the phrase!). Nola asks whether "there are two different objects, the Bohr-Rutherford electron and the mature-Bohr electron" (205). There are certainly two phrases, two hyphenated names. But they are both empty names, no 'hyphenated entities' are picked out by either of them. They are out of the same (Kantian) stable as Eddington's 'table-of-common-sense' and 'table-of-physics'. "The Moon-as-conceived-by-Aristotle was perfectly spherical, whereas the Moon-as-conceived-by-Galileo had mountains and oceans on it" is just philosopher's gobbledy-gook for "Aristotle thought that the Moon was perfectly spherical, whereas Galileo thought it had mountains and oceans on it".

If ontological discontinuity is not in the world, where is it? It is in our theories. It is discontinuity in what our theories *say* is in the world, in the 'ontological commitments' of our theories. Difference of theory yields, we are told, difference of ontological commitment. The false theories of our ancestors did not succeed in referring to objects at all. It is not that the world changed when Galileo got a different theory about the Moon than Aristotle's. Rather, we found out that Aristotle was not talking about any object in the world at all. The same applies to Galileo, of course, since his theory about the Moon was not quite right, either. And the same will apply to us, if our theory is not quite right.

But of course, this is just bad philosophy of language. Difference of theory does not imply difference of ontological commitment. Aristotle's theory, and Galileo's, and ours are different theories about the same object, the Moon. The expression 'the Moon' has the same referent in all these theories. People say that these simple-minded views are all very well for observational terms like 'the Moon', where we can point to the object to fix the referent, but will not work for theoretical terms. Nola shows convincingly that continuity of reference can be established for theoretical terms as well. As he says, of the example involving competing theories of light:

> Given the above account of reference determination for 'light' in the two theories, we can dispense with the claim that there is object discontinuity from Fresnel's theory to that of Maxwell. There is a "something" that both theories are about. And it is not just structure … We have not said very much about the intrinsic properties of light or even its nature or essence … All we have is a "something" which … obeys F- and M-equations. None of this forms the basis for an objection to the above account of referential continuity. If the Fresnel and Maxwell equations are correct, then there will be the same "something" that satisfies them … There is no need for extreme structuralists to deny the existence of "objects" … that stand in the structural relations. So there is no need for ontological flight from objects to structure. (219)

What Nola may not have noticed is that essentialism makes a mess of all this. He writes as if his story about referential continuity is compatible with any kind of metaphysical view. But essentialist metaphysics is incompatible with that story. As Nola well knows, Fresnel and Maxwell also had different views about the intrinsic nature of light. Fresnel thought light was intrinsically F, Maxwell thought it was intrinsically M. Suppose Fresnel had been an essentialist, who thought light not just intrinsically F but essentially so – light would not *be* light if it were not F. Suppose Maxwell had been an essentialist, too, who thought that light would not *be* light if it were not M. Then, despite the continuity of F-equations and M-equations, there is no one "something" that both theories are about. According to Maxwell, nothing answers to Fresnel's 'essentialist definition' of light – and vice versa. According to us, who think light is neither F nor M, neither Fresnel's nor Maxwell's theory is about anything at all. All that is left are common F-equations and M-equations, neither of them about anything. The way out of this morass is obvious enough – give up essentialism.

A few paragraphs back I uncharitably described the view that mathematical structures are *all* that exists as 'loopy'. It is not, of course, loopy to think that mathematical structures are *some* of the things that exist – as also, perhaps, are the mathematical objects that mathematical structures are structures of. Which brings me to **Mark Colyvan**'s paper. Colyvan correctly identifies an apparent tension in my views. I have argued, *contra* philosophical idealists of various kinds, for realism about (some of) the observable entities of commonsense. I have also argued, *contra* scientific antirealists of various kinds, for realism about (some of) the unobservable entities of science. But I am reluctant to extend realism to platonic entities, including the entities of mathematics. The reason for my reluctance is simple. Platonic entities are *queer* entities: not only are they unobservable, like the theoretical entities of science, but also they do not exist in space, or time, or space-time, and they have no causal powers. That being the case, why should we believe that such entities exist?

Colyvan and I agree that the only decent answer to this question is the Indispensability Argument. Mathematics is indispensable to our best theories about both the observable and the unobservable world. Furthermore, mathematics is to be taken at face-value, and the usual semantics applied to it. From which it follows that we should believe in numbers for the same kinds of reason as we believe in electrons. Numbers and electrons are in the same boat, epistemologically speaking. I agree with Colyvan and many others that this is the best, indeed the only decent, argument for numbers and other platonic entities. And I reject the argument.

It is important to see that there are *two* premises in the indispensability argument, one concerning the indispensability of mathematics, the other about the indispensability of numbers. Here a use/mention confusion pervades Colyvan's writings. He says "we ought to count as real any entity that plays an indispensable role in our best scientific theories" (225). But entities such as rocks, or electrons, or numbers, do not play any role in our *theories*. What play a role in our theories are *expressions* which, if they refer to anything at all, refer to entities such as rocks, or electrons, or numbers. In a telling footnote in his book, Colyvan writes:

> I often speak of certain entities being dispensable or indispensable to a given theory. Strictly speaking it's not the entities themselves that are dispensable or indispensable, but rather it's the *postulation of* or *reference to* the entities in question that may be so described. Having said this, though, for the most part I'll continue to talk about *entities* being dispensable or indispensable, eliminable or non-eliminable and occurring or not occurring. I do this for stylistic reasons, but I apologise in advance to any reader who is irritated by this. (Colyvan 2001, p. 10, fn. 18)

But use/mention confusion is not a matter of style. Consider Santa Claus theory, the stories we tell to small children at Christmas time. The name 'Santa Claus' occurs in these stories, and (let us grant) is not eliminable from them, and is indispensable to them. But Santa Claus does not occur in the stories – how could he, he does not exist? In this case, we all agree that the indispensability of the name to the theory does not carry with it commitment to the existence of the entity. We take the name at face-value, and say that it fails to name anything.

Acausal platonic entities are odd because they play no causal role in the world, unlike the unobservable yet causal entities that scientific realists are happy to believe in. How can they be indispensable if it makes no difference whether they exist or not? Armed with use/mention confusions, Colyvan disarms this worry: mathematical entities are indispensable to our theories, and "do not need to play causal roles in those theories (indeed, it is generally agreed that they do not play such roles)" (230). But mathematical entities play no role in theories, expressions for them do. The role played by mathematical expressions in our theories is, like the role played by all expressions, not causal but semantic.

Are acausal mathematical objects any more mysterious than stars and planets outside our light cone that do not causally interact with us? And do we not "accept the existence of stars and planets outside our light cone because they play an indispensable role in our best cosmological theories"? (232). Well, no. What Colyvan should say is that we "accept the existence of stars and planets outside our light cone because our best cosmological theories say that they play an indispensable causal role in the world". Acausal mathematical entities play no causal role in the world, and you cannot argue for their indispensability as we argue for that of stars and planets outside our light cone.

Use/mention confusions ease the transition from words to things. But in the case of numbers, they are not the only thing that does this. The transition proceeds without confusion if we insist (a) that the words are to be taken at face-value, and (b) that appropriate sentences containing them are true. Given (a) and (b), the indispensability of mathematics, of number-talk, gets us to numbers as well. Antiplatonists or nominalists about numbers must resist the transition by rejecting (a), or (b), or both – they must either refuse to take the talk about numbers at face-value and give it some sort of antiplatonist construal, or take the talk at face-value and say that it is false, though indispensable. The indispensability argument, then, only shows the indispensability of thought or talk about numbers (and other mathematical entities). It does not, by itself, show the indispensability of numbers, when these are viewed as platonic entities. To get the entities out of the indispensability of the talk, more is required, and the more that is required can be disputed.

Colyvan evidently finds it absurd to dispute (a) and (b). He complains that "the nominalist cannot employ the usual semantics to account for the truth of sentences such as 'there is a number smaller than 2'" (224, note 2). We might as well say that a nominalist about the creeps cannot employ the usual semantics to account for the truth of sentences such as 'Osama Bin Laden gives me the creeps'. We are all nominalists or antirealists about the creeps. We all say that if the 'usual semantics' is applied to this sentence, then it is false. Or, more plausibly perhaps, we say that if it is true, then the 'usual semantics' is not to be applied to it, either because it is just a colourful way of saying that Osama Bin Laden makes me nervous, or because what makes it true is just the fact that Osama Bin Laden makes me nervous. (These two alternatives are not quite the same: the first seems to involve some claim of 'sameness of meaning', the latter does not. But I ignore this complication here.)

The creeps is one thing, numbers another. The nominalist about numbers can simply deny that 'there is a number smaller than 2' is true, and set to work to explain the indispensability of number-talk. That is Hartry Field's programme. It is consistent with acknowledging the *logical* truth of "If Peano's axioms are true, then it is also true that there is a number smaller than 2". As for the easy move from the undeniable truth "I have five fingers on my right-hand" to "The number of fingers on my right-hand is five", the nominalist can say that the latter is just a long-winded way of saying the former.

Colyvan discusses my argument that "If we view [the indispensability argument] from a Popperian perspective, it begins to lose its charm" (1986: p. 90). By "a Popperian perspective" I simply meant a falsificationist perspective:

> Imagine that all the evidence that induces scientists to believe (tentatively) in the existence of electrons had turned out differently. Imagine that electron-theory turned out to be wrong and electrons went the way of phlogiston or the heavenly spheres. Popperians think that this *might* happen to any of the theoretical posits of science. But can we imagine natural numbers going the way of phlogiston, can we imagine evidence piling up to the effect that there are no natural numbers? This must be possible, if the indispensability argument is right and natural numbers are a theoretical posit in the same epistemological boat as electrons.
>
> But surely, if natural numbers do exist, they exist of necessity, in all possible worlds. If so, no empirical evidence concerning the nature of the actual world can tell against them. If so, no empirical evidence can tell in favour of them either. The indispensability argument for natural numbers is mistaken. (1986: pp. 90-91)

One thing is sure – one should never, in philosophy, say "Surely". Colyvan grants that this objection "presents serious difficulties for any defender of the indispensability argument who takes mathematical entities to exist of necessity" (230). His way out is contingent platonism. It is a contingent matter whether mathematical entities exist. So, for example, the number five exists in some possible worlds but not in others, and only empirical inquiry can tell us whether the actual world is one of the worlds in which the number five exists!

I confess that I overlooked this possibility. Before I turn to discuss it, it is worth dwelling on what Colyvan loses by advocating it. The charms of necessary Platonism are considerable. Its chief charm is that it enables us to disarm Benacerraf's epistemic worry about platonism. How can we find out that the number five exists? No problem, says the necessary Platonist. "The number five exists" is a necessary truth, true in all possible worlds. Necessary truths are knowable a priori. So we know a priori that the number five exists. This is the line of thought that many Platonists, early and late, have pursued. It has difficulties of its own, of course, concerning the very idea of necessary existence. Like many philosophers, I can make little sense of that idea.

What about contingent platonism? According to this view, there is a possible world in which the number five exists, and another one in which it does not. Now, of course, we cannot identify the 'possible worlds' spoken of here with standard set-theoretical interpretations of formulas. Nothing is easier than to produce an interpretation in which the sentence "The number 5 exists" is false – let the domain of that interpretation contain just Mark and me, viz. {Alan Musgrave, Mark Colyvan}. It is equally easy to produce another interpretation in which that sentence is true – let the domain be {5}. 'Possible worlds' are not such model-theoretic trivialities. 'Possible worlds' are heavy-duty metaphysics of some kind.

So the question recurs – is there a heavy-duty possible world in which the number five does not exist? A contingent Platonist like Colyvan evidently thinks that there is. So what will he make of the equivalence "The number of fingers on my right-hand is five if and only if I have five fingers on my right-hand"? The left-hand side is false in a world bereft of the number 5. The right-hand side might be true in a world bereft of the number 5. If the right-hand side *is* true in that world, and the left-hand side false, then the equivalence is false in that world. The equivalence is contingent as well, true in some heavy-duty worlds and false in others. (One can hardly avoid this by saying that in a world bereft of the number five, there cannot be five of anything, not even numerals to count up to five with!) I think that if the equivalence is true, then it is necessarily true. And that if numbers exist, then they exist necessarily.

Which brings us back to the indispensability argument. The chief burden of Colyvan's paper is that if you go in for what he calls 'confirmational holism' or 'justificatory holism', you will be led to realism about mathematical entities. 'Confirmational or justificatory holism' is just the Duhem-Quine thesis that whole systems of theory, and not isolated hypotheses, are required to obtain testable predictions about the world. If these predictions turn out to be correct, then it is the whole system that gets confirmed. If the whole system includes mathematical theories, that postulate mathematical entities, then confirmation of the whole system gives us evidence that the mathematical entities exist.

I am afraid that this misses exactly the 'Popperian perspective' from which my objection to the indispensability argument proceeded. Never mind 'confirmational holism' – what about 'refutational holism'? When a theoretical system gets refuted, scientists know that something is wrong somewhere, but they do not rest content with that. They try to pin the blame more narrowly, and there is an enormous literature on how they do that, beginning with Duhem and Quine themselves. My question was: can we imagine pinning the blame for a refutation of a theoretical system that contained arithmetic on the non-existence of the natural numbers? Or, to take a silly example, when we refute a system by finding out that one drop of water put together with another drop of water does not yield two drops of water, but rather one big drop, does it make sense to pin the blame on "1 + 1 = 2" and say that in our world the number two does not exist?

Never mind silly examples. When physicists found out that space is not Euclidean but rather non-Euclidean, did they find out that Euclidean geometry is *false*? A hard-won logical empiricist distinction is important here, the distinction between pure and applied geometry (or more generally, between pure and applied mathematics). What physicists found out is that real space-time is not adequately represented by Euclidean laws, so that applied Euclidean geometry did not work. But Euclidean geometry considered as a theory of pure mathematics remained intact.

Colyvan tries to convict me of a Popperian 'separatist' theory of confirmation or justification. He says that "The separatist wants a crucial experiment that identifies the causal roles of the entities in question" (231). He asks "What are the crucial experiments that establish the existence of [stars and planets outside our own light cone]?" (231). Of course, there are no such crucial experiments. No 'crucial experiment' can *establish the existence* of any entity. Since experiments test whole bodies of theory, one cannot say that any successful experiment confirms (let alone establishes the truth of) some particular existential assertion in that body of theory. Similarly, if an experiment contradicts a prediction drawn from a body of theory, it is muddle-headed to say that it contradicts some particular hypothesis in that body of theory. Still, scientists want to try to figure out which particular hypothesis is responsible for the failed prediction. And the question recurs: would it ever make sense to pin the blame on the hypothesis that some abstract mathematical entity happens to exist in our world?

I confess that I can make little sense of this. The reason is simple. Existence claims regarding acausal abstract entities form no part of testable theories about the (actual) world. As Cheyne has shown elsewhere, when Colyvan is defending his contingent Platonism he points out that it is neither obvious nor necessary nor knowable a priori that there are odd numbers greater than five. Rather, what is obvious, necessary and knowable a priori is the logical truth "If the axioms of arithmetic are true, then there are odd numbers greater than five". Conflating the two is committing the 'conditional fallacy'. But when Colyvan is trying to convince us that acausal objects can help explain things, he commits that same fallacy. We are to explain why a square peg of side length l will not fit into a round hole of diameter l. Colyvan says there is a non-causal explanation involving acausal squares and circles. Cheyne points out that there is only a causal explanation involving square pegs and circular holes, and the claim that these concrete objects satisfy the antecedent and consequent of the conditional geometrical claim. The conditional claim may be indispensable to the explanation, but it does not assert the existence of acausal objects. (For further details and examples, see Cheyne (unpub.).)

6. ANTIREALISMS

I turn, finally, to those contributions that discuss antirealist views of one kind or another. One currently fashionable antirealist view is the *semantic conception of theories* (hereafter SC). According to SC, the traditional 'statement view of theories' is wrong: theories are not true or false statements, but sets of models. Laws or generalisations are clauses in definitions of those sets of models. Now if anything is 'true by definition', a definition is. But definitions say nothing about the world, and neither do scientific theories construed as clauses in such definitions. Claims about the world only enter the picture when it is claimed that some definition is not empty, that it applies to some part of the world, that this or that set of data can be 'fitted' into one of the models defined. Antirealism is already to the fore. The aim of scientific theorising is to provide a toolkit for building models of the phenomena. The name of the game is 'saving the phenomena', as van Fraassen put it. The aim is empirical adequacy, not truth.

Noretta Koertge criticises SC on methodological grounds: its associated methodology is "either antithetical to commonly accepted norms of scientific inquiry or hopelessly ad hoc" (237). It is an appealing line of criticism. Popper launched a similar criticism of what he called 'instrumentalism', the view that theories are more or less useful instruments or tools for saving phenomena:

> "Instruments … cannot be refuted … . The instrumentalist interpretation will therefore be unable to account for real tests, which are attempted refutations, and will not go beyond the assertion that *different theories have different ranges of application*. But then it cannot possibly account for scientific progress. Instead of saying (as I should) that Newton's theory was falsified by crucial experiments which failed to falsify Einstein's, and that Einstein's theory is therefore better than Newton's, the consistent instrumentalist will have to say … 'Classical mechanics … is everywhere exactly "right" where its concepts can be applied'" [*Conjectures and Refutations*, p. 113. The quoted sentence is from Heisenberg.]

It is no wonder that Koertge's arguments against SC resemble Popper's. As she rightly sees, SC is yet another version of the instrumentalism that Popper was attacking. According to SC, recalcitrant phenomena do not refute a theory, they merely show that a definition fails to apply. Newton's theory was not refuted by data about Mercury's perihelion – rather, "scientists failed to find a Newtonian model for the trajectory of Mercury" (239). Similarly, Lavoisier's experiments with mercury merely showed that "the phlogiston model fails to represent the controlled combustion of mercury as carried out in this experiment" (Koertge, 239, quoting Giere, 1991, p. 76).

Koertge asks why the failure of scientists to find a Newtonian model for the trajectory of Mercury loomed so large in the history of science, and led to the replacement of Newton's theory by Einstein's. She suggests that SC adherents "must also posit that scientists value theories whose models fit more phenomena" (4). Well, they might posit that, but what methodological justification of such a 'posit' might they give? If we have a tool that saves many phenomena, why not keep it for those purposes, and only use a more refined tool when we need to? Hanson saw the point. (Hanson was defending yet another antirealist view of theories, the inference-licence view. No matter.) Writing about 'The Scientist's Toolbox', he warmed with characteristic verve to the analogy between theories and tools:

> There are those who, knowing something of modern physics, dismiss Newtonian mechanics with a snap of the fingers. I suggest that such people … are confusing the *purposes* of what are now seen as two distinct … methods of representation. It is no longer a question of Newton's laws being wrong and Einstein's laws being right. … because of its greater simplicity, the Newtonian formulation is greatly preferable to relativistic quantum laws … Only when our experiments absolutely require more refined representation do we place the Newtonian formulation to one side. But then there is no question of appealing to … the truer method of representation. Newtonian mechanics is simply *inappropriate* to the representation of relativistic and quantum phenomena. By exactly the same token, relativity theory and quantum mechanics are inappropriate to the representation of a good deal of the macro-physical world. (We do not distemper walls with water-colour brushes, nor do we repair watches with sledge-hammers.) [Hanson, 1969, 315-7]

This fits the semantic conception like a glove. As Hanson makes clear, if we take seriously the view that theories are tools for representing bodies of data, then failure to represent some data-set does not refute a theory, it merely shows that it is not a good tool for that particular job. The methodological demand for theories of broad scope makes as much, or as little, sense as the demand for multi-purpose tools. Nor does the methodological demand for theoretical unity make much sense. If theory A saves all its phenomena, and theory B saves all its phenomena, then all the A and B phenomena have been saved. Why search for a unified theory that will save A-phenomena and B-phenomena all at once? Scientists seek comprehensive and unified theories – but toolmakers do not seek multi-purpose or all-purpose tools. (Swiss Army knives are fun, but not many professional carpenters jettison all their special purpose tools in favour of them!)

Antirealists have, down the ages, invoked simplicity or 'economy of thought' at this point. But if simplicity is viewed, as it must be by instrumentalists, merely as a 'pragmatic virtue' (van Fraassen), then it is not clear that this will work. Modern physics is not pragmatically simpler than classical physics. Hanson took it for granted that the 'Newtonian formulation' is simpler and hence pragmatically preferable to 'relativistic quantum laws'. Relativistic mechanics is not 'economical of thought', as any physics student knows. The great Duhem saw the problem better than most:

> Why should we forbid the worker the successive employment of disparate instruments when he finds that each one of them is well adapted to a certain task and not well adapted to another job? [P. Duhem, *The Aim and Structure of Physical Theory*, p. 294.]

Duhem's answer to his question was not pragmatic simplicity or economy of thought. Duhem's answer was, in a word, metaphysical faith (*op. cit.,* pp. 296-7). When Giere talks about the world having "a single structure", he merely echoes Duhem. But professions of metaphysical faith are anathema to Duhem's latter-day incarnation, van Fraassen, who has managed to convince the world that it is realists who must invoke metaphysics, not constructive empiricists. Van Fraassen's constructive empiricist demands for comprehensive and unified theories are quite ad hoc.

Koertge is right. One can graft demands for comprehensiveness and unity onto SC, and make it seem that there is no methodological difference between realism and instrumentalism. But such grafts are ad hoc. They should be rejected by an SC that takes seriously the idea that theories are just tools for representing data-sets.

Another route to antirealism is to jettison the commonsense realist theory of truth in favour of an epistemic theory of truth. Are you obsessed with the sceptical nightmare that our best theories might be false and the entities they postulate non-existent? Then *define* truth as what ideal science, pursued to its limit, will throw up. Earlier I used a Fitch-style argument, made famous by Williamson, against epistemic theories of truth in general, more precisely, against the idea that all truths are knowable which is a common presupposition of such theories. As **Graham Oddie** explains, that argument can be disputed, and it would be nice to have a direct argument against Peircean idealism that does not get bogged down in those disputes. This Oddie has provided. He modestly confesses that there may be an error lurking in his knock-down-drag-out argument. I can only report that I have not been able to find one.

Nor can I find anything to disagree with in **Hans Albert**'s critique of what he calls 'methodological historism'. Like Albert, I find this doctrine unclear, to say the least. Is it the view that naturalistic methods cannot be applied in history, that history *qua* history is not amenable to them? That view seems to be refuted by the historical natural sciences, or the historical parts of natural science, such as the attempts by astronomers to explain the formation of the solar system, of geologists to explain current formations on the surface of the earth, or of evolutionary biologists to explain the current structure and distribution of animals and plants on the surface of the earth. One does not find astronomers or geologists or evolutionary biologists saying that they employ some special method of *verstehen*, that they have no need of the concept of cause, and such things. Or is it, more likely, that it is not history as such that is non-naturalistic, but rather the explanation of human action, past and present? That view seems to be refuted by the theoretical social sciences, such as psychology or economics or sociology, which try to formulate general hypotheses and give causal explanations of human behaviour and human society using them. All that remains, when confusion is set aside, is the distinctive focus of interest of historical as opposed to generalising scientists, a point that Weber stressed long ago. But we can see an historical event in all its particularity and uniqueness without thinking it uncaused or not governed by law. It may just arise from a unique confluence of independent causal chains, each of them law-governed.

And so, finally, to one such event, back at the beginnings of science. I can say little about **Andrew Barker**'s fascinating study of Ptolemy's *Harmonics,* except to thank him for it. In my amateur excursion into these matters, I criticised Duhem's instrumentalist interpretation of Ptolemy's geometrical models in astronomy, an interpretation that had become something of an orthodoxy amongst subsequent historians. Although the realism/instrumentalism issue is not the main focus of Barker's paper, I take comfort from what he says there. For what he says, as I read it, is that Ptolemy was as much a realist in his *Harmonics* as he was in astronomy. Ptolemy's harmonic ratios were a guide, not just to real musical consonances, but also to real psychological phenomena and to real astrological phenomena. In saying this, Ptolemy tapped into traditions foreign to our ears. But it is unhistorical to suppose that speculations which seem bizarre to us cannot have been seriously (that is, realistically) intended by their proponents. Besides, I am tempted to add, Ptolemy's harmonic speculations, while they may seem bizarre to us in their details, are not so bizarre in their general orientation. That reason (mathematical reason) is the key to understanding nature is a Greek legacy passed down to us through the scientific revolution, and still going strong today, as Redhead points out (pp. 8-11).

REFERENCES

Archibald, G. C. (1967) 'Refutation or Comparison?' *British Journal for the Philosophy of Science* 17: 279-296.

Cheyne, C. (unpub.) 'Getting an "Is" from an "If": Mathematical Realism and the Conditional Fallacy'.

Colyvan, M. (2001) *Indispensability of Mathematics.* New York: Oxford University Press.

Duhem, P. (1954) *The Aim and Structure of Physical Theory* (transl. by P. Wiener). Princeton: Princeton University Press.

Hanson, N.R. (1969) *Perception and Discovery*. San Francisco: Freeman, Cooper and Co.

Musgrave, A. (1986) 'Arithmetical Platonism: Is Wright Wrong or should Field Yield?', *Essays in Honour of Bob Durrant.* Otago University Philosophy Department.

Musgrave, A. (1999) *Essays on Realism and Rationalism.* Amsterdam; Atlanta: Rodopi.

Musgrave, A. (2001) 'Rationalität und Zuverlässigkeit.' ['Rationality and Reliability'] *Logos* 7: 94-114.

Popper, K. R. (1968) *The Logic of Scientific Discovery,* 2nd edn. London: Hutchinson.

Popper, K. R. (1969) *Conjectures and Refutations.* London: Routledge and Kegan Paul.

Psillos, S. (1999) *Scientific Realism: How Science tracks Truth.* London, New York: Routledge.