RJP

The Richmond Journal of **Philosophy**

Volume **One** Issue **Three** **Spring 2003**

In this issue...

Frank Jackson
on the knowledge argument

Michael Luntley
on experience and reality

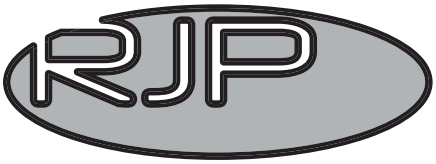
Stephen Grant
on the ontological argument

James Hill
on Hume and causation

Pierre Cruse
on scientific realism

Christopher Norris
on realism and possible worlds

Richmond upon Thames College



The Richmond Journal of Philosophy

Issue three
Spring 2003

Philosophy Department

Richmond upon Thames College

Egerton Road

Twickenham

Middlesex

TW2 7SJ

United Kingdom

email: rjp@rutc.ac.uk

www.rutc.ac.uk/rjp

Contents

[Editorial]

Welcome to the third issue of the Richmond Journal of Philosophy. In the first paper Frank Jackson considers his now classic thought experiment in the philosophy of mind on 'What Mary didn't Know'. Michael Luntley discusses our experience of the world while Stephen Grant undertakes an analysis of the ontological argument for God's existence. Next Hume's account of causation is the topic of James Hill's paper. In our final two papers, Pierre Cruse and Christopher Norris present ways of defending realism, with a particular emphasis on issues in the philosophy of science.

We are pleased to report that the first annual RJP conference took place on February 28th at the University of London Institute of Education. Speakers included Simon Critchley, David Papineau and Jo Wolff whose talks ranged over continental philosophy, the philosophy of mind and equality. We hope to present versions of their papers in future editions of the journal. Michael Luntley also spoke at the conference, basing his presentation on his contribution to this edition of the journal.

Purpose of the Journal

The motivation for and ambition of the journal is to provide serious philosophy for students who are at an early stage in their philosophical studies. The style and content of the papers will be accessible to students who have yet to become hardened to the more technical and specialised journals of professional philosophy.

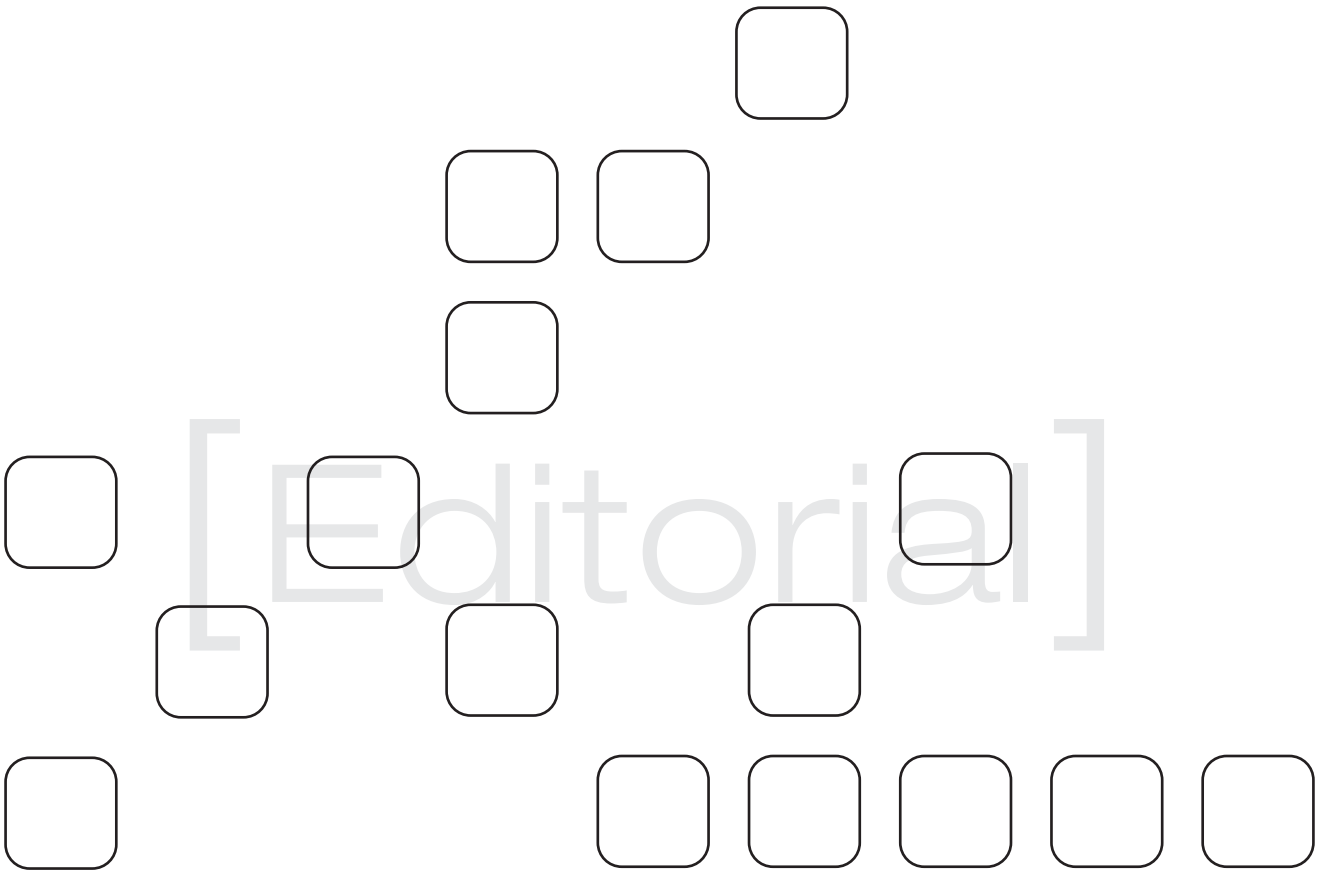
What do we mean by 'serious' philosophy? First, the content of the journal is not constrained by a remit to appeal to or reach the interested general public. Whilst the papers must speak to the needs of students who are relatively inexperienced in philosophy, they presuppose that their audience is actively engaged in philosophy. Second, the content is serious in its focus on the central areas of philosophy. One must beware of the dangers of trying to impose more precision on a subject than its nature will allow. Therefore, some degree of caution is called for in talking of the central areas of philosophy. Nonetheless, the big or traditional questions of metaphysics, epistemology, and ethics will provide the journal's centre of gravity. The third way in which the philosophy is serious is through the scope, variety and depth of analysis that can be achieved by the accumulation of papers over time. Moreover, each paper is not simply an introduction to one of the main topics on A-level, IB or degree courses. Such papers will indeed have a role in the journal, but they will not be the only kind. Our contributors will be offering original papers based on their own research. The journal will be a forum for the kind of critical engagement and debate that characterise the practice of philosophy. The fourth way in which the philosophy is serious is in the contributors themselves. The vast bulk of the papers will be written by professional philosophers engaged in both research and teaching.

About the [Editorial] Board

Stephen Grant is a full-time lecturer in philosophy at Richmond upon Thames College. He has also taught at King's College London where he is completing his doctorate on the emotions. His main interests are in the emotions, ethics and political philosophy.

Dr Paul Sheehy teaches philosophy at Richmond upon Thames College and King's College London. His main areas of interest are in metaphysics, political and moral philosophy and the philosophy of the social sciences. His doctoral thesis was on the ontological and moral status of social groups, and he has published papers on social groups, voting and explanation and realism.

Paul Sperring is head of the philosophy department at Richmond upon Thames College and an A-level examiner in philosophy. He completed his undergraduate and masters studies at Warwick University, studying both analytic and continental philosophy. He has recently become the inaugural teacher fellow at King's College London philosophy department.



Frank Jackson

The [Knowledge] Argument

I

The physical sciences tell us a great deal about what our world is like. They also tell us a great deal about what we are like. They tell us, for example, that our bodies are made up of the stuff that the physical sciences — physics, chemistry and biology — talk about. We can think of this as our physical nature, our nature as revealed by the physical sciences, or maybe by certain future developments of the physical sciences. A perennial question is whether our physical nature is our total nature. Is it the case that the physical account of us captures what we are like without remainder, or is there something more to us and, in particular, are our minds or aspects of our minds that something more?

This is one way of asking the dualism versus materialism question. Dualism says yes, there is more to us than our physical nature; materialism says that's all there is. Or, more precisely, the kind of materialism we will be concerned with says that's all there is. There is a weak kind of materialism which holds that each and every part of us is material — there is, for example, no soul as traditionally conceived—but grants that we have special properties different in kind from those inventoried in the physical sciences, and different in kind from those we can think of as constructions out of properties so inventoried. Weak materialism denies rather than affirms that our nature is exhausted by our physical nature and is really a kind of dualism, a dual attribute kind of dualism. Materialism, as we are understanding it, is the real McCoy and is often called physicalism when it is important to keep this in mind.

In the first issue of this journal, Alan Thomas, 'Is Your Mind Your Brain?', canvasses some of the arguments for and against the view that our physical nature exhausts our nature. We will focus on an especially thorny part of the debate between dualists and materialists. It concerns the 'feely' side of psychology, the mental states with a phenomenology, the mental states for which there is something it is like to be in them. These phrasings are different ways of getting at the same general idea: the idea that there is a feel to states like pain, itches, experiences of colour, feelings of heat and so on that is missing in the case of mental states like the belief that oxygen is essential to life or the desire that it will rain soon. Such beliefs and desires may be accompanied by various feelings but in themselves are not feelings and lack any distinctive phenomenology.

Physicalism has a special problem with the mental states with a phenomenology. We often think of cognitive states like belief and conative states like desire in functional terms. Belief is a state typically induced in us by the environment, which carries putative information about the environment, and desire is a state that works with belief qua informational state to make our bodies move in such a way that the environment is changed in various ways—the ways that would satisfy our desires in cases where our beliefs are true. This is, of course, far too crude an account of belief and desire but one gets a glimmering of how something like it might work, and if something like it could be made to work, there would be no threat to physicalism from belief and desire. Functional states are part and parcel of the materialist cum physicalist conception of what our world is like. The situation is very different, it would seem, in the case of itches, heard sounds, sensings of blue and the like. They appear to have an intrinsic feel connected with our consciousness of them that is left out of account by any functional story. A way to bring out the contrast between belief per se and perceptual experience is to think of what happens when you shut your eyes. You will likely retain some sort of belief about

the location, size and colour of the objects around you. But, it seems, 'something' goes when you shut your eyes. What goes is the phenomenological aspect of perceptual experience.

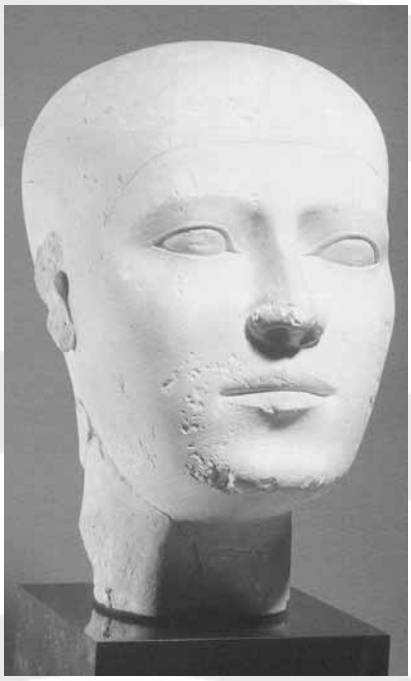
The knowledge argument is one way of seeking to turn the intuition in play in the remarks immediately above into an argument from premises even confirmed materialists find it hard to deny. The argument has a number of forms and has been advanced by many writers. Some references are given at the end. From my possibly biased perspective, I like the following version.

We suppose that we have a brilliant physical scientist, Mary, who is confined in a black and white room. There are no windows. She herself is, we may suppose, painted white all over and dressed in black. All her information about the world and its workings comes from black and white sources like books without coloured pictures and black and white television. However, the lectures she receives over the black and white television and the books she reads are amazing feats of exposition in physics, chemistry, biology and cognitive science, and she has extraordinary powers of comprehension and retention. In consequence, she is, despite the artificial restrictions in which she works, extraordinarily knowledgeable about the physical nature of our world, the neurophysiology of human beings and sentient creatures in general, and how their neurophysiology underpins their interactions with their surroundings including for instance the fact that on many occasions they produce words like 'red' and 'yellow' (if they speak English) when in front of blood and buttercups, respectively.

Can she in principle deduce from all this physical information what it is like to see, say, red? It seems that she cannot. Despite her vast knowledge of the physical facts, there is something about our world and especially about persons' colour experiences she is ignorant of. This conclusion is reinforced by reflecting on what would happen should she be released from her room. Assuming that there is nothing wrong with her colour vision despite its lack of exercise during her imprisonment, she would learn what it is like to see red, and it is plausible that this would be learning something about the nature of our world, including especially the nature of colour experiences. Surely, runs the argument, she could not have predicted this in advance, and surely she would come to realise that her conception of the mental lives of others had been seriously impoverished. It follows that she did not know while in the room all there was to know about our world. But ex hypothesi she did know all there was to know physically. Therefore, there is more to know than all there is to know physically. Physicalism is false

This argument has attracted some strong supporters (but ones who have typically sought to make one or another improvement to the argument, as is the way of philosophers), and some strong critics. As you would expect given the current and understandable presumption in favour of materialist views of mind, the second group has been larger than the first. The criticisms have been very various; so various as to constitute close to an empirical refutation of the idea sometimes floated that it is obvious where the knowledge argument goes wrong.

The objections can be usefully categorised in terms of which of the two main claims in the argument is targeted. One claim in the knowledge argument is that complete physical knowledge is not complete knowledge tout court (or anyway not as far as the mind is concerned). The other claim is that if physicalism were true, it would be. From these two claims it follows that physicalism is false by Modus Tollens[1]. Let's call these two claims the incompleteness claim and the deduction claim, respectively. The incompleteness claim is supported by the plausibility of the contention that Mary would learn something on her release. The deduction claim is that were physicalism true, there would be nothing Mary could not in principle work out about what our world is like.



II

Let's look at some objections to the deduction claim. Critics of this claim urge that it is consistent to hold a) that complete physical knowledge is incomplete knowledge of our world, with b) that the physical account of what our world is like is complete.

Sometimes the critics spell this out by pointing out that no amount of knowledge of what one's world is like amounts to knowledge of, for example, who or where one is in one's world. Suppose that some high-powered physics demonstrates that our world will go through two exactly similar cycles and that we know this. In that case we could not possibly know which cycle we were in. There would be no way to tell the difference between being in the first cycle and being in the second cycle. Our surroundings and our bodies, for example, would be indistinguishable whether we were in the first or the second cycle. In particular, Mary would not know which one she was in. She would know that there were two exactly alike people called 'Mary', each living in exactly alike black and white rooms, but would not know which one she was. The moral is that complete knowledge of what our world is like does not necessarily deliver knowledge of where or who one is. A special case will be where complete physical knowledge fails to deliver knowledge of where or who one is. So the idea that physicalism is committed to complete knowledge of the physical delivering complete knowledge simpliciter is a mistake.

This point about the inadequacy of knowledge of what our world is like to deliver knowledge of who and where we are is correct and important. But it is far from clear that it goes to the heart of the knowledge argument. Mary's lack of knowledge seems at least in large part to concern what her world is like, not her or anyone's identity or location in it. She is ignorant before her release, it seems, of what certain experiences are like. That is the point of the argument.

The other main line of attack on the deduction claim starts from the point that the very same things, facts, events and so on can be known under many guises. The FBI may know Jones under the guise of the main suspect in a mail fraud; you may know him as your next door neighbour; I may know him as the person who has just bought an expensive car at my dealership. The same happening may be Jones's arrest; the disturbance next door; the event that means the car is never paid off. This suggests that we could grant that Mary's knowledge of what her world is like is incomplete without being forced to the conclusion that what is not known is non-physical. Her ignorance is a matter of there being features or categorisations of certain happenings in the world, especially those involving colour experience, that elude her while she is inside the room. All the same, the happenings in question are purely physical ones. When Mary leaves the room, she acquires knowledge but entirely through knowing about the very same

physical things, facts, events and so on under different guises or under different categorisations. She gets new ways of categorising happenings around her and thereby acquires new knowledge, but it is, all the same, knowledge of the purely physical and so no threat to physicalism. Consider, for example, someone knows the Cartesian coordinates of a series of points that all lie on a circle without realising that the points lie on a circle. It is not until they graph the points, or do the calculations that reveal that the points satisfy the relevant equation, that they make the discovery. They will learn something, but it is plausible that they do not learn about a new feature—the circularity was 'there' in what they knew from the beginning—the learning was a matter of its becoming revealed when they saw that the points could be categorised in a certain way.

The interest of this suggestion is clear but again it seems that the knowledge argument survives. For the guises, ways of categorising, must all be consistent with physicalism if physicalism is true. But then, it seems, Mary could know about them when inside the room. It is hard to see how given physicalism, there could be ways of grouping things into categories that are, in principle, unavailable to her while in the room. Of course, the ways of grouping may not be easy ones to latch onto. It is easy to miss the fact that a series of points lie on a circle and it can be much harder in more complex cases. But it should be possible in principle to spot the relevant groupings if only one is smart enough and can put the data together aright. However, no amount of cleverness in assembling data and spotting patterns will in itself tell Mary in the black and white room what it is like to see red, or so it seems.

III

Attacks on the incompleteness claim, the claim that as Mary learns (would learn) something new about what the world is like when she leaves the black and white room and so that her knowledge while in the room is incomplete, fall into two broad categories. The first can be introduced by reference to the example of hard to spot patterns that we have just been discussing. The difference between, on the one hand, being in a situation where patterns in, or ways of classifying, data are very hard to grasp although available in principle and, on the other hand, being in one where it is impossible to make the classifications is not always transparent. In consequence, we should insist that Mary can know what it is like to see red while in the room. The strong intuition to the contrary is the result of the fact that it would be extremely hard for her to spot the relevant patterns, along with wrongly conflating the extremely hard with the impossible. This reply is sometimes coupled with the view that the way our brains enable us to see colour goes via the way our brains and optical systems pick up on very unobvious patterns in the effects coloured objects have on light.

Mary's practical problem inside the room will be that her pattern detector, her optical system, is not being allowed to do its job of making sense of data that looks, on the face of it, a complete mess. But if we suppose, as we should when evaluating the knowledge argument, that Mary has worked out everything she could know in principle (though not in practice) from her vast data bank of physical information, then, runs this reply, we give her knowledge of what it is like to see red. She will not learn what it is like to see red on her release; she will already know.

The second line of attack on the incompleteness claim offers a different diagnosis of the appearance that she acquires knowledge on her release. Instead of the explanation being a slide from the extremely hard to the impossible, the explanation is that we confuse knowledge how with knowledge that. Mary gains knowledge how, not knowledge that.

'Knowledge that' is propositional knowledge, knowledge of the kind of world we live in, of how things are. 'Knowledge how' is knowledge of how to do something: ride a bike, capture a likeness in oils, or recognise a dog that is about to attack. It is an ability. Exercising and acquiring an ability may well require propositional knowledge. Painting classes help us acquire and exercise the ability to capture a likeness in oils, and a lot of propositional knowledge is imparted at these classes: the right paints to buy, good ways of getting skin colour and relative proportions right, and so on. But no amount of propositional knowledge is in itself enough to enable one to capture a likeness in oils. If the knowledge Mary acquires on her release is knowledge how, abilities, there is no problem for physicalism in the knowledge argument. The argument will merely have demonstrated her lack of certain abilities while in the room, not a gap in her knowledge concerning the nature of our world. It will not show that there are features of our world she knows nothing of when in the room despite knowing everything physical there is to know; it will only show that there are things she cannot do.

What abilities does Mary acquire on leaving the room? The usual suggestion is that she acquires the ability to summon up in memory what it is like to see red, to imagine how well a new colour scheme will go, to recognise colours (as opposed to having to ask someone else what something's colour is) and the like.

Everyone agrees that this is part of what happens, would happen, to Mary on her release but there is a persistent intuition that in addition she learns more about how things are. Isn't part of the explanation of her new abilities the fact that she has more knowledge that? Imagining seeing a rhomboid is greatly helped by knowing what a rhomboid is. In the same way, it seems that Mary's new abilities will rest in part on her new knowledge of what it is that she's exercising her abilities on.

Conclusion

The overall situation with the knowledge argument seems to be that the objections to it all make important points but somehow leave one unsatisfied. However the reasons that favour physicalism are very strong. Many feel that the case for physicalism is so strong that one or more of the objections to the knowledge argument must be right, and that the task before us is to find a way of putting the successful objection or objections that removes the feeling of dissatisfaction, or maybe to find an explanation of why there will always be a feeling of dissatisfaction which allows us to discount the significance of the feeling. The latter would be an explaining away of why we are in the grip of the argument.

Reading

There have been many statements of the knowledge argument, or of arguments close to the knowledge argument in one way or another. The statement above is closest to those in Frank Jackson, 'Epiphenomenal Qualia', and 'What Mary didn't Know'. A recent reprinting of 'Epiphenomenal Qualia' is in David Chalmers, ed., *Philosophy of Mind: Selected Classic and Contemporary Readings* (Oxford: OUP, 2002). Recent reprintings of 'What Mary Didn't Know' are in Frank Jackson ed., *Consciousness* (Dartmouth: Ashgate, 1998), and John Perry and Michael Bratman, eds, *Introduction to Philosophy: Classical and Contemporary Readings*, third edition (Oxford: OUP, 1999). These collections contain many articles discussing the knowledge argument and related matters.

Notes

1 If P then Q, not Q, therefore not P. Put more simply, if physicalism is true (P), then physical knowledge provides a complete account of knowledge (Q). But physical knowledge doesn't provide a complete account of knowledge (not Q), therefore physicalism is not true (not P).

Michael Luntley

The World [in View]

I

In experience the world comes into view. Experience can be a direct experience of the world. On such a view, what we receive in experience is not a facsimile of the world, it is the real thing. This is a direct realist model of experience. It is an attractive model. For the direct realist, what it is to be a conscious subject is to be engaged with the world. Our consciousness does not leave us semi-detached from things but, often enough, in direct presence of things. I shall call this the unitary model of experience.

To say that the world is directly present in experience is to challenge the prevailing view, articulated in detail by Kant, but common from earlier philosophers like Locke and Hume. The prevailing view is indirect realism. It embodies a binary model of experience in which experience has two components: the raw given - sense-data, intuitions - plus concepts that organise the given. The given is supplied by the world, concepts are supplied by the mind. It takes two to make an experience.

In this essay I want to examine one aspect of the unitary model of experience. I want to examine the question of the boundary between self and world. The unitary model of experience is attractive and, I suspect, correct. It is, however, problematic, for in jettisonning the binary model it is, prima facie, unclear where the world ends and the self begins in experience. The idea of a unitary model of experience has acquired renewed interest in contemporary philosophy. The idea is central to the renewed interest in Sellars' seminal critique of the 'myth of the given',[1] and to contemporary work on singular thoughts.[2] The latter provides a simple illustration of the central idea.

[The World]

II

Suppose you have a perceptual demonstrative thought about a piece of chalk. You think,

This is white.

How could you have the 'this-ness' part of the thought without the object? How could there be a 'this' way of thinking of an object without the object to which the 'this' way of thinking is orientated? If the answer to these questions is that you cannot make sense of the 'this' way of thinking independently of the object to which the thought is orientated, then the individuation of the thought is object-dependent. This is a way of thinking that is only available if the object is present in experience. Having the experience of thinking

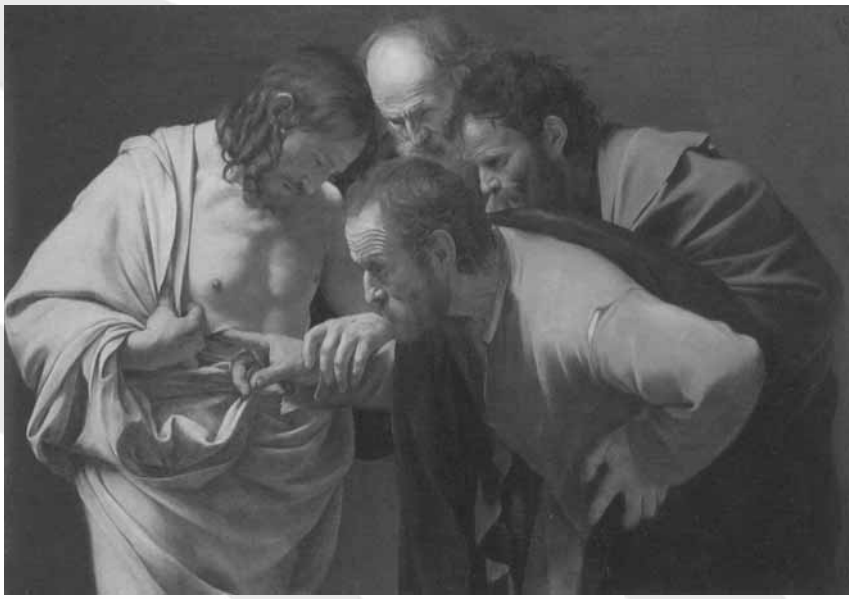
This is white

is not possible without the object that provides the focus to the experience. This account of perceptual demonstrative thoughts is not uncontentious, although I think it is correct. Whether or not it is correct is not, however, relevant.[3] This idea illustrates the unitary account of experience and what I am interested in is the unitary account of experience. I want to know what it means to say that experience is unitary.

If the world, rather than merely sense-data, can be available in experience, then it is the world that is in view, not the given and what is in view directly engages with thought. The denial of the given means that concepts apply directly to the world. On this unitary model of experience, there is nothing else for concepts to apply to. On a binary model the given is causally produced and then rationally organised by concepts. On the unitary model, the world itself is rationally organised in thought. That is what is meant by saying that, on such a view, the world falls within the space of reasons.[4]

The problem with this view is if the world enters directly into the conceptual content of experience (the world is rationally engaged), what becomes of the idea of receptivity? What becomes of the idea that experience involves something that impinges on us? The attraction of the binary model is that it has a simple story about receptivity. Receptivity is explained by the first component, the causal receipt of the given. Experience also has a spontaneity to it as we organise the given with concepts. On the binary model there is, in principle, a clear divide between self and world. Of course, it is notoriously difficult to pin down precisely where that divide occurs, but the general idea is that there is a point in the causal sequence that produces experience where, as it were, causation leaves off and we take over. That is the point where receptivity ends and spontaneity begins. It is the point where the self meets the world. And it's the existence of this point that appears to make it easy for the binary model to explain error and false thoughts. One attraction of the unitary model is that it does not separate receptivity and spontaneity. In particular, spontaneity is integral to the account of what it is to experience the world. But that attraction is also what makes the model problematic, for if spontaneity is too apparent in experience, what becomes of receptivity, the idea that we are impinged upon? And if that is unclear then, what is missing in the unitary model is a clear account of how to individuate the point where the world ends and the self begins.

The binary model has considerable intuitive appeal. It accommodates familiar Cartesian fantasies about our detachment from things. In the film, The Matrix, the central character Neo starts off as a victim of the Matrix - the artificial intelligence that gives people the experience of leading



autonomous lives when, in fact, they are biological batteries for the Matrix. The film captures the intuitive appeal of the Cartesian binary model of experience, for it provides for the possibility that what we get in experience is less than the world as conceptualised. What Neo gets is a given (sense-data) that is then conceptually organised to produce an experience as of a real world. The experience is misleading, for the conceptual organisation is not faithful to the real causal source of the given.

It might seem that the intuitive plausibility of the Matrix scenario shows that the unitary model of experience cannot be right. Surely, so the thought might go, we understand the Matrix scenario and that understanding requires the binary model of experience? If so, the unitary model has a real problem in accounting for the intelligibility of that film. My response to this is to argue that, properly understood, the Matrix requires the unitary model. Furthermore, properly understood, the Matrix shows that the unitary model has a powerful account of the subject/world boundary, a boundary that, on closer examination, is deeply problematic on the binary model. In other words, much as in the Matrix, things ain't quite what they seem.

III

The binary model looks attractive. It is, however, deeply flawed. The attraction of the model is, as noted, its ready account of receptivity - the idea that the world impinges on us in experience. Receptivity is explained in terms of the causal receipt of the given. This apparent strength is the model's chief flaw.

The concept of the given has to achieve two things. It has to account for receptivity - experience impinges on us, plus it has to show that receptivity provides that against which we adjust belief. The latter is the idea that experience can give us reason to revise belief, for it can falsify our beliefs. Now, for experience to provide reason to revise belief, the source of this cannot be the second component - the application of concepts to the given. If that were so, we would only ever revise our beliefs (operations of concepts) against other beliefs (operations of concepts). That does not provide the notion of that against which we revise belief, for it only provides a model of revising a belief with respect to another belief. That is why, although we say it is experience that can give us reason to revise belief, it must be the first component - the given - that supplies the force to belief revision.

Note the problem here is not that if we only ever revise belief against another belief we can never know that the revision is correct. That epistemological problem, although real, is not my concern. The problem is that the idea of revising a belief in terms of another belief does not provide an adequate account of what revision consists in. The point against using the second component to achieve this is not an epistemological point, it is a metaphysical point. It is a point about what is constitutive of the idea of revision. The thought is simply that for genuine revision to be possible, belief must be measured against something independent of belief. And that cannot therefore be another belief. This means that the given has to capture both the idea of receptivity as bare impact and also as rational force to belief revision.

But the given cannot discharge both requirements. It captures the idea of impact by virtue of its causal character. The receptivity of the given is the receipt of something prior to the operation of concepts, for on the binary model that comes later. That is why receptivity is construed causally. Thus construed, it readily captures the idea that receptivity is concerned with things without the mind, things that impinge upon us. And it is this that gives us a clear model of the boundary between world and self, for the world is outside the mind because it is beyond the causal boundaries of the self. It gets to be outside because of the way it causally impacts on us.

However, if receptivity is characterised causally it cannot provide the rational force for belief revision. When experience gives us reason to revise our beliefs, the impact is rational, not causal. Having an experience that calls for the revision of belief is not simply a matter of being hit! The given as causal receipt prior to the operation of concepts is a mere bully; it is not the sort of thing that can stand in relations of rational support. But that means that, on the binary model, the rational force of belief revision cannot occur until after the given has been taken up by concepts. But that means that the rational force of belief revision is not an impact at all, for it arises from within the operation of concepts – it arises from within the second component. If that is the source of belief revision, then it is, as we have seen, no real source of belief revision at all.

What looks attractive in the binary model is its central flaw. The attraction is a simple account of error, for when experience is mistaken, this is explained as a deviant application of concepts to the given. Veridical and illusory experiences have a common factor - the common way we take sense-data whether or not the data come from the world or from the Matrix. Error consists in taking the wrong interpretation of the given. But, in reality, this means that there is no real notion of error, for error should amount to the idea that one ought to revise belief in the face of recalcitrance. As we have seen, the binary model does not, in truth, have a credible concept of recalcitrance, for the given as such has no right interpretation. The given is just causal given. Rightness or wrongness of interpretation is wholly to do with what happens after concepts have been applied.

IV

Can the unitary model do any better? What is required is a model of experience in which what is in experience is both something that impinges on us and carries the rational force needed for belief revision. With the world in view, the unitary model perhaps captures the idea of rational force required for belief revision, but it can appear to do so in a way that makes it inadequate for the same reason as the binary model.

The problem with the binary model was that it had a clear model of receptivity and an inadequate model of belief revision. The unitary model, in having the world in view, claims to have the very thing that impinges on us within the space of reasons. It claims to have that which is outside the mind fully engaged with concepts in belief formation and revision. That being so, it should have a ready and simple model of belief revision. But, of course, what is problematic here is the idea that it really is the world that comes into view. The worry must be that the unitary model collapses to just one half of the binary model, the half represented in the second component. That is to say, the obvious worry with the unitary model is that it accommodates the idea of the rational force of belief revision by treating beliefs as measured only against other beliefs. If so, it fails at the same hurdle that stops the binary model. If that were right, it would turn out that there was no adequate account of a self/world boundary in the unitary model, for there would be no adequate account of the world.

The issue about the nature of the self/world boundary is the issue about being clear that it really is the world that comes into view on the unitary model and not merely a facsimile of the world composed out of our concepts. What is needed then is a concept of the impact characteristic of experience that both captures the intuitive idea of impact and provides something that rationally engages with belief. The unitary model claims to have such an account, for it claims that the world is in view. The world is what impacts on us and, on the unitary model, concepts apply directly to the world. The bit that I am insisting is not quite clear in this model concerns the concept of 'impact'. Without a causal model of impact, with what right can the unitary model of experience genuinely capture the receptivity of experience? That is the problem.

There is an answer to this and it is simple to state, but what it means offers a shift in the way we have been used to think about experience. The concept of the world that comes into view cannot be characterised causally nor can its impact be characterised causally. If it were characterised causally, it would not be suited to rationally engage with concepts. The problem is to find an alternative to a causal characterisation. The alternative that I suggest is to conceive of the world as that which is independent of will. The concept of receptivity is handled in terms of things independent of will. My experience impacts in so far as it has a content that thwarts my will. Of course, often enough, the normal way in which things thwart my will is due to their causal powers and capacities – tables and chairs get in my way as I move around. But the idea of receptivity is not understood in terms of the causal bumps and bruises such things provide; it is understood in terms of the more primitive idea that these are things independent of will. It is not that such things causally impinge on me that makes them part of the world, it is that their causal impinging is contrary to my will.

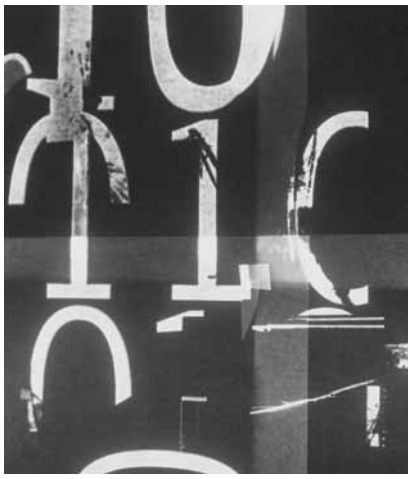
With this more abstract account of the impact in experience there is nothing to stand in the way of the thought that what is independent of will can rationally engage with concepts. The problem with the binary model is that receptivity is characterised in a way that makes it unavailable to concepts. The more abstract characterisation does not have this consequence. One way of making this plain is simply to remark that the concept of the world as that which is independent of will is a concept of the world as understood from within the conceptually structured point of view of the subject.

The concept of the world under the binary model is of a place characterised with a privileged set of concepts - those that characterise its causal properties that account for the physics of its interactions with our perceptual systems. A richer characterisation, for example in terms of properties that we find rationally salient, is, at best, a characterisation of the world as conceptualised by us. For the binary model, that richer characterisation is, potentially, a characterisation only of the world as we take it to be after processing through our concepts, rather than a characterisation of how it is.

For the unitary model, there is no principled requirement for a gap between a characterisation of the world in terms of basic concepts that characterise it as it is, and those concepts that characterise it only as we happen to take it to be.[5] We might want to make a distinction between those concepts that reflect how the world really is and others that merely reflect something local about our perspective on it, but there is no requirement to draw such a distinction that flows from the basic account of what is going on in experience.

That is one respect in which a unitary model of experience challenges fundamental ideas about experience, ideas that have their roots in traditional empiricism. The challenge, however, runs much deeper. If we think of the world as, in the first instance, that which is independent of will, that can only make sense if we acknowledge that the subject of experience is a self-as-will.[6] The concept of the world is not given independently of the concept of the subject of experience. Our concepts of both items are mutually dependent. This has a number of consequences. I shall remark here on just one.

[The World]



Traditional empiricism conceived of experience fundamentally as a given. The role of the subject is passive - the subject receives inputs. It is because of this passivity that it looks easy for the binary model to account for the boundary between self and world in causal terms. If the account of the unitary model that I have suggested is correct, we have to consider a much more dynamic model of experience. Experience is not a receipt of items from the world, it is a tussle, an ongoing tension between will and non-will. Experience is a dynamic engagement in which the will is variously satisfied and thwarted. The traditional empiricist model is often conceived in terms of a display in the Cartesian inner theatre. It is like watching a private slide show, or movie film. That, of course, is the image suggested by the Matrix. On the unitary model, that cannot be right. Experience is not fundamentally an inner display, it is an ongoing tussle between will and that which can thwart it as we strive to bring our will into a stable relation with the world. In experience, we do not receive 'things' from the world at all; we adjust the tensions in the balance between will and non-will. In experience we are in the world.

But what about the Matrix? Surely that film illustrates clearly the empiricist binary model in which experience is the receipt of an inner movie? Although that thought is tempting, I think it is wrong. Indeed, the Matrix endorses the unitary model of experience that I have been describing. On the unitary model, the boundary between self and world is marked by the infringements to will. The distinction between reality and appearance is the distinction between that which is independent of will and that which is not. And that is precisely the account of the boundary between self and world that we get in the Matrix. The account of experience illustrated by that film is the unitary model. Neo learns the difference between reality and the Matrix because the latter is subject to will. When he is in the Matrix he learns that how things go can be determined by will. It is a difficult lesson, but he learns how to dodge bullets. He comes to know that, in karate, as Morpheus tells him 'you are faster than that', even though he is already moving at impossible speeds. In the end, in the Matrix Neo learns that a man can fly. In reality, he cannot fly. It is only when plugged into the Matrix that how things are becomes subject to will so that eventually, even when hit by a hail of bullets, he does not die. The message is all about the self-as-will. It is a message that requires that the self/world boundary in experience is not the fine line where receptivity ends and spontaneity begins. The self/world boundary in experience is itself a negotiation between receptivity and spontaneity, the negotiation between will and non-will. And this is the unitary model of experience in which we have the world in view.

Notes

1 Cf W. Sellars, *Empiricism and the Philosophy of Mind* reprinted with an introduction by R. Rorty and a studyguide by R. Brandom (Cambridge, Mass.: Harvard University Press, 1997). See also McDowell's Woodbridge Lectures 1997: Having the world in View: Sellars, Kant and Intentionality', *Journal of Philosophy* vol **95**, No. 9: 431-91.

2 G. Evans, *The Varieties of Reference* (Oxford: Clarendon Press 1982), cf M. Luntley, *Contemporary Philosophy of Thought: truth, world, content* (Oxford: Blackwell, 1999), especially Chapters 11, 12.

3 The contentious bit concerns the scope for false thoughts. How is error possible if the thought is object-dependent? This turns out not to be a problem, cf. Luntley op. cit Ch.12. Rather than explain that issue, let me instead approach the same point from a different angle.

4 This phrase is due to Sellars op. cit.

5 The privileged account is often thought of as an absolute conception of the world. See Bernard Williams, *Descartes* (London: Penguin, 1978), Chapter 2 for a detailed study of the idea of the absolute conception of the world primary/secondary qualities distinction.

6 This move is central to my reading of Wittgenstein. See my *Wittgenstein: the conditions for the possibility of judgement* (Oxford: Blackwell, *in press, forthcoming* 2003).

Stephen Grant

[On the Ontological Argument]

Introduction

The ontological argument is one of the three classic philosophical arguments which aims to prove the existence of God. Where it differs from the cosmological and teleological arguments is that it runs along the lines of pure reason, and demands no empirical evidence to support the premises. The thrust of the argument is that if we properly understand our concept of God and we are rational, then we must accept He exists. Although this argument enjoyed a revival in the twentieth century, my focus in this article will be to set out the two earliest and best-known versions of the argument, which are found in the work of Anselm and Descartes, as well as considering what are widely regard as the most challenging responses, those of Immanuel Kant and Gottlob Frege.

Anselm

St Anselm of Canterbury (1033-1109) offered the first version of the ontological argument in the second and third chapters of *Proslogion*.[1] The argument is presented as a response to the sceptical concerns of the Fool of *Psalms* 13 and 52 who doubts God's existence. Anselm begins by characterising God as 'something-than-which-nothing-greater-can-be-thought', and asks us to consider which is greater, an object which exists only as an idea or one which exists both as an idea and in reality. Anselm claims that even the Fool must accept that the second option is the correct one, because if we choose the first, we can always think of something greater - an idea which also has an instance in the real world. This is taken to be the classic statement of the argument, but Anselm amplified this initial account in *Proslogion* 3 where he states that not only must we think that God exists, but that we cannot think of Him not existing. In other words, not only does He exist, but He necessarily exists, and it is this additional claim which provided the starting point from which the twentieth century versions of the argument proceeded.

The first response to the argument came from Anselm's contemporary, the twelfth century monk Gaunilo of Marmoutiers.[2] Gaunilo offers a parallel argument which generates an absurd conclusion. He asks us to imagine the fabled 'Lost Island', which was said to be 'superior everywhere in abundance of riches to all those other lands that men now inhabit'.[3] Gaunilo argues that as an island of such unparalleled beauty is clearly better if it exists than if it doesn't, and given that as soon as we come to understand that the idea of the best possible island must include its existing (for otherwise it wouldn't be as good), then we must conclude the Lost Island really does exist. He takes it that any argument which can be used to demonstrate the existence of mythical islands is absurd and asks us on the basis of this to reject Anselm's proof.

Anselm's defence against Gaunilo provides the ontological argument with a further refinement which we find more prominently in Descartes' version, and which gives the argument much of the specific content with which it is usually associated. He accepts that if the argument could be applied to an island then the island must exist, but flatly denies that the argument can be applied in this way. The reason for this is that the meaning of the concept 'island' does not entail that it must exist. As such, it is obvious that we can think of the Lost Island or any other such object as not existing in reality. But this just isn't the case with God. It is part of the meaning of our concept of God that He is eternal and that He did not come into existence at any point. In this respect, He differs from islands and everything else, in that He is the only being for whom it impossible for us to think of Him as not existing once we have understood what He is. In thinking of God, we just have to think He exists, and this distinguishes Him from any other being or object.

Descartes

At this stage it is worth setting out the version of the argument found in 'Meditation 5' of the *Meditations on First Philosophy*[4] by the French philosopher Rene Descartes (1596-1650). Descartes offers a more economical and rather more readable account of the argument, but one which proceeds from the same premises as those set out by Anselm. After characterising God as 'a supremely perfect being', the key points of his argument are captured in the following passage;

> But when I concentrate more carefully, it is quite evident that existence can no more be separated from the essence of God than the fact that its three angles equal two right angles can be separated from the essence of a triangle, or that the idea of a mountain can be separated from the idea of a valley.

For Descartes, our idea of God is 'clear and distinct' - an idea which we cannot doubt, and which is held with the same certainty as our ideas in the fields of geometry or mathematics. The idea of God includes existence, just as the idea of a triangle includes having three angles, and just as we must think that any triangle has three angles as soon as we understand what the concept means, we must understand that God exists once we have understood the concept of God.

The parallel with Anselm is already clear. There is little or no practical distinction to be drawn between Anselm's definition of God as a being 'greater-than-which-nothing-can-be-thought', and Descartes's reference to the idea of 'a supremely perfect being'. Each believes that a clear understanding of the meaning of the concept 'God' will be sufficient for the rational agent to understand that God must be thought to exist, and each is also forced into the same sort of refinement to the argument which will create the opportunity for the devastating arguments which Kant was later to press against them. For like Anselm, Descartes was forced to defend himself against the *reductio* argument we saw from Gaunilo, and his response was also to argue that God's existence was necessary, and that He is the only being who has the property of necessary existence. Only by taking this path can Descartes avoid committing himself to the absurd claim that we can show how any object, if defined as being perfect or necessary, must be thought to exist.

Kant

There are two common reactions when one comes across the ontological argument. The first is that it is ingenious, and the second is that for all its ingenuity there must be something wrong with it. This second intuition is perhaps best captured in the slogan that one cannot simply define something into existence, and the most famous critique of the argument came from the man who gave it its name, the great German philosopher Immanuel Kant (1724-1804). Kant's criticism is part of one of the most influential texts in western philosophy, The *Critique of Pure Reason*, a work almost as famous for its technical difficulty as for its philosophical brilliance.

Kant offers a series of arguments which attack certain key premises. Descartes insisted that to say God necessarily exists was equivalent to saying that God has the property of necessary existence, which he took to be no different logically to saying God has the property of being omnipotent or omniscient. Descartes argued that 'we can take the word "property" to stand for any attribute, or for whatever can be predicated of a thing'.[5] (To predicate simply means to say about a subject that it has a particular property, so that in the sentence 'the apple is red', 'the apple' is the subject, and 'is red' is the predicate which picks out the property of redness.) This means that just as the statement 'a triangle has three sides' must be true, it is claimed that 'God exists' must also be true. In each case, it is argued that this is because the predicate is contained in the subject - having three sides is part of being a triangle, and existing is part of being God.

It is here where Kant disagrees. He accepts that we can certainly have a *definition* which includes the notion that the object must be thought of as existing, but claims it is a separate question as to whether or not there really is an *object* whose non-existence is unthinkable. His first move is to question Descartes' analogy between the statement 'a triangle has three sides' and 'God exists', both of which would be true a *priori* if we accept the arguments of the fifth meditation. Kant asks us to distinguish between what he calls 'the unconditioned necessity of judgements' on the one hand and 'the absolute necessity of things' on the other. To put it more simply, it is one thing to say that we just have to think that a triangle has three sides - here we are talking solely in terms of the *concept* of a triangle - but quite another to say that we just have to think that there are triangles in the world, where we are talking about the existence of *real objects* called triangles. It is indeed a logical contradiction to say 'a triangle doesn't have three sides', but there is no *logical* contradiction in saying 'there are no triangles'.

Needless to say we cannot fall back on the strategy of pointing to actual triangles, in that we cannot have recourse to *a posteriori* evidence (based on experience) to support an argument which is said to be true *a priori* (based on reason alone).

Kant then goes on to offer two explanations of why there couldn't be a logical contradiction in denying the existence of any object - even of a supreme being. The first is that it is impossible for any logical contradiction to arise in such statements. If we take the example at the end of the previous paragraph, we say in the statement 'a triangle doesn't have three sides' that the triangle lacks a certain property - having three sides. We generate the contradiction because we have denied that the subject has a particular property which it must have in virtue of its meaning - part of the meaning of the concept triangle just is that it has three sides. But when we deny something exists, we don't deny the subject has a particular property, we get rid of the subject and all its properties, so there are no subject and predicate left between which any contradiction can arise. In Kant's own words, 'We have thus seen that if the predicate of a judgement is rejected together with the subject, no internal contradiction can result, and that this holds no matter what the predicate may be'.[6]

Kant's second argument is one which has provoked considerable dispute, and remains a live issue in modern philosophical logic. He argues that existence is not a '*determining* predicate', by which he means that when we say something exists, we do nothing to enlarge on our understanding of the *concept* of that thing. In order to explain this, consider the following case. Imagine someone new to religious thought who thus far knows only that God is omniscient, and who then reads these three sentences: 'God is omnipotent.' 'God is omnibenevolent.' 'God is.' Kant claims that the first two statements would add something to the student's concept of what God is, but the third wouldn't. This is because the third sentence contains no predicate, and therefore cannot expand our understanding of the concept of God.

One response has simply been to dispute this conclusion. If we say something exists, does it not add to our understanding of that concept to know that there are objects in the real world to which the concept corresponds? But this does not really answer Kant's point. He does not argue that such statements provide us with no information about the concept, but rather that it does not add to the *meaning* of the concept. If we say something exists or something is, we are not expanding the list of predicates which we now know belong to the subject, we are saying that this subject with all its predicates corresponds to an object in the real world.

Frege

For Kant, the ontological argument was a futile attempt to think something into existence. He saw the argument as one which confused what is possible with what is actual, and one which simply could not deliver the intended result - 'we can no more extend our stock of (theoretical) insight by mere ideas, than a merchant can better his position by adding a few noughts to his cash account'.[7] The strength of his arguments was such that for over two centuries there was little or no interest in defending the ontological argument, but this did not prevent the Austrian philosopher Gottlob Frege (1848-1925) from providing what he believed to be a superior demolition of it to that of his German predecessor.[8]

Frege argued that the problem with the argument comes to light more clearly when we compare the way in which we use the concept of existence with the way in which we use numbers. To begin with, let us consider the confusion he felt arose when we use numbers. When we say things such as 'I have red apples' and 'I have two apples', it looks as if 'red' and 'two' are playing the same role in their respective sentences - each looks as if it 'qualifies' the apples (gives us some information about them). That they do *not* play that same role can be seen from the fact that in the first sentence each apple is red, but in the second sentence each apple is not two. Frege argues that a standard predicate such as 'red' can tell us something about an *object* in the real world, but a number such as two can only tell us about a *concept*. This becomes clearer if we compare two further sentences, 'Venus has red moons' and 'Venus has 0 moons'. The redness is a property of

the physical objects circling Venus, but this obviously cannot be the case with the zero in the second sentence as there aren't any objects of which it could be a property. According to Frege, the zero is a property not of any object, but of the concept 'moon of Venus'. To say 'there are 0 moons' is to say that there are no objects which fall under this concept, just as to say 'I have two apples' is to say that there are two objects which fall under the concept 'my apples'.

Now, what has this talk of numbers to do with God's existence? The answer is that when we say something exists, we are effectively assigning a number to it - the number one. When we say God exists, we are claiming that there is one real object which falls under the concept God, and the problem for the ontological argument now becomes clear. Frege's response to anyone claiming that existence is part of the meaning of God would simply be that this proves absolutely nothing with regard to whether or not there is a real God. As existence is only ever a property of a concept then we still need to address the further question as to whether there is a real object which falls under the concept of a supreme being with necessary existence. And it is here where we find perhaps the most eloquent statement of why we can't define God into existence. Regardless of what the definition is, there will always be a further question as to whether our definition picks out a real object, and no matter how much necessity or existence we build into the definition, this further question will always remain.

Frege's central claim is that just as the language of numbers confuses us by giving the impression that 'two' and 'red' are used in a similar way, the language of existence produces the same mistake. We tend to assume that just as redness can be a property of a real object, existence can be too,

and it is here where the mistake lies. Once the confusion is cleared up then one key premise of the ontological argument falls. According to Frege, we cannot know from the definition of God that there is a real God which has the property of existence, because no real object ever has this property.

Conclusion

It would be wrong to think that Kant and Frege had consigned the ontological argument to history. The question arises as to whether or not their views are as devastating to Anselm's version as they are to Descartes', and both Norman Malcolm and Alvin Plantinga have recently put forward more technical versions of the argument pitched in terms of modal logic. But the thought remains that it is unlikely that anyone who approaches the literature with a genuinely open mind is likely to feel convinced of God's existence based purely on these arguments. Regardless of the elegance of the logical proofs on offer, the intuition that the real existence of an object cannot be conclusively proved purely by considering a definition remains stubbornly in place.[9]

Notes

1 The key extracts from each of the first three primary texts to which I shall be referring here can be found in Brian Davies' *Philosophy of Religion: A Guide and Anthology* (Oxford: Oxford University Press 2000).

2 *ibid*, 313-317.

3 *ibid*, 316.

4 *ibid*, 327.

5 *ibid*, 331.

6 *ibid*, 338

7 *ibid*, 341.

8 See *The Frege Reader*, 102-103, ed Michael Beaney (Oxford: Basil Blackwell 1997).

9 Plantinga's account of the argument can be found in Brian Davies' *Philosophy of Religion: A Guide and Anthology*, 342-353, (Oxford: Oxford University Press 2000). For Norman Malcolm's version, see his 'Anselm's Ontological Argument', in *Philosophical Review*, 69, (1960). An excellent, brief overview of all the material covered in this article as well as these last two arguments can be found in Chapter 4 of Brian Davies' *An Introduction to the Philosophy of Religion* (Oxford: Oxford University Press 1993).

James Hill

On Billiard Balls

David Hume against the Mechanists

I would like to approach David Hume's theory of causation from an historical angle. That does not mean of course that I take the question of whether Hume was right or not to have an historical answer, true only in a certain place at a certain time. What it means is that to understand Hume on causation I think it can help to appreciate the context in which he was writing - against whom he was primarily arguing and what the significance of certain examples and terms might have been at the time. This kind of contextual understanding is needed to help us judge what exactly Hume was saying, and that must be a first step towards deciding whether or not he is right.[1]

Let us begin with a point that may puzzle modern readers new to the First *Enquiry*.[2] Hume seems curiously concerned with goings-on on the billiard table. Of the innumerable cases of causal links that he might have chosen, Hume seems to find conclusions about the actions of billiard-balls - or, more exactly, of one billiard ball hitting another causing the second to be set in motion - sufficiently representative to stand in for causal relations *per se*. Someone picking up the *Enquiry* when it was first published, however, would find the significance of Hume's choice of example unmistakable. Why? Because at the time, 1748, the billiard-ball model had become standard in explaining the nature of the universe. Atomism ruled.

This doctrine was essentially a revival of the physics of Democritus and Epicurus which thought of nature as reducible to minute particles of impenetrable matter (atoms) and empty space (the void). It gathered support throughout the seventeenth century and by the time of John Locke's death in 1704, it had taken the British intellectual world by storm. Locke himself was a cautious proponent of atomism, or 'corpuscularianism' as it was now more often called. Locke might just have had time to read Isaac Newton's *Opticks*, which came out in that same year and pronounced that 'God in the Beginning form'd Matter in solid, massy, hard, impenetrable, moveable Particles'.[3] Newton's public endorsement of atomism was highly influential in the following decades because he was a thinker of such towering authority. It is likely that Hume introduced the billiard balls with one eye on the ultimate particles advocated by Sir Isaac and others.

The billiard-ball example had another significance, however, in addition to being a model of atoms in the void, and it is this that I wish to concentrate on. It stood for a kind of causal action that was thought to be transparent. It was an example of 'impulse', that is of one body causing changes in another body by means of contact - by pushing it or striking it. 'Impulse', Locke wrote in his *Essay concerning Human Understanding*, is 'the only way which we can conceive Bodies operate in.'[4] The peculiar intelligibility of impulse was thought to consist in our being able to perceive how the first body acts on the second to bring about the change. Nothing is mysterious or hidden. If we know the perceivable qualities of both objects we can foresee exactly how they will interact.

That impulse was the only intelligible physical causality was a central principle in 'the mechanical philosophy', which was endorsed in differing forms by almost all of Hume's immediate predecessors and which Hume himself was implicitly challenging. The Mechanists contrasted the transparency of impulse with our experience of other more opaque causal relations. For example, when we see a magnet attract iron-filings we see a sequence of events - the magnet is placed in the vicinity of the filings, the filings move towards the magnet, sticking to it or clumping round it - but we do not perceive how the magnet manages to make the filings move towards it. The magnet's *modus operandi,* or way of operation, remains obscure. Or to take another example, if we light a firework, the flame ignites gunpowder

within the rocket (this is a seventeenth century firework!) leading to its take off. We are not really able to discern by what means the take off is brought about. We know that it *follows* the burning of the fuse and various characteristic sounds and flashes, but not *how* it is produced.[5] In the case of impulse, however, the mechanists were convinced that the 'how' was known. To return to Hume's favoured example, the mechanist would say it was because one solid object of the same size as another was pushing into the same space that the second ball was forced to move off.

So Hume homed in on the interaction of billiard balls because it represented for others a *fundamental* and *transparent* case of the operation of causal power. In the light of this context, Hume's strategy can be expressed as follows. He set out to show that the transparency that mechanists claimed to find in impulse was really a kind of illusion produced by habit or 'custom'. So much of our experience of causality is of contact-action, Hume argues, that our very familiarity with these cases makes us feel that they have a transparency that they do not really possess. In fact, Adam, on first perceiving the impulse of one billiard-ball on another, would find the resulting motion of the second ball just as surprising and mysterious as, say, the action of the magnet on the iron-filings. It is custom that blinds us to this truth.

> We fancy, that were we brought on a sudden into this world, we could at first have inferred that one Billiard-ball would communicate motion to another upon impulse; and that we needed not to have waited for the event, in order to pronounce with certainty concerning it. Such is the influence of custom, that, where it is strongest, it not only covers our natural ignorance, but even conceals itself, and seems not to take place, merely because it is found in the highest degree.[6]

On causation, Hume is the great leveller. He denies any sort of hierarchy between more and less transparent cases of causal relation. All causal relations, in his view, are equally untransparent. All of them amount to constant conjunctions and our perceptions of them never give us insight into the *modus operandi* of the connexion. The only distinction between the different constant conjunctions is that some are perceived more often and thus we have more developed expectations about how the chain of events will continue.

Now, to be a mechanist or a Democritean atomist today would be foolish. It would be comparable to believing that the earth is flat or that the sun literally rises in the morning. Science has moved on and though in modern physics there is still talk of 'atoms', they have very little in common with the hard 'massy' little particles of matter that Democritus, Epicurus, Newton *et al* imagined. But there may still be insight in the mechanists' analysis of natural causation: indeed I think there is one important truth that Hume's critique does not do proper justice to.

To appreciate this insight, let us take another example of impulse, this time from the railways. At the end of the line we can watch a train slowing down before hitting the buffers, which bring it to an abrupt halt. The buffers - although they have a bit of give - obstruct the train. How? By being made of hard, inflexible matter of the right shape and size to get in the way. This uncontroversial reflection suffices to show that there is something wrong with at least one of Hume's negative claims:

> In reality, there is no part of matter, that does ever, by its sensible qualities, discover any power or energy, or give us ground to imagine, that it could produce any thing, or be followed by any other object, which we could denominate its effect. Solidity, extension, motion; these qualities are all complete in themselves, and never point out any other event which may result from them.[7]

Pace [with due respect to - ed] Hume, the solidity, extension and motion of the buffers and the train combine to make the stopping of the train a necessary effect of the buffers being in its way.

A Humean might object as follows: 'we can easily imagine that the train continues in the same direction through the buffers and the end of the line; our expectation that it will not is based on custom, not on a perceived necessity.' But what exactly are we imagining when we imagine the train continuing regardless of the buffers? We can certainly imagine this happening if the buffers and their supports suddenly become as soft as butter, or if they were already cracked and so break up on impact, or if the buffers themselves moved with the train in the same direction indefinitely, or any number of other possibilities. But notice, in order to imagine the train not stopping we have to change the *qualities* of the buffers. We have to imagine that that they were not solid after all (but like butter, or cracked), or that they are not at rest (but in motion along with the train). But can we imagine that the buffers remain solid, at rest and undiminished in size *and* that the train simply continues? If the train is imagined as a hard, material object - and not a phantom - I think this is impossible.

In our example certain qualities of objects determine the kind of effects those objects - trains and buffers - can have. This is true, I think, of almost all the kinds of causal interaction involving impulse that the mechanists thought were transparent and fundamental. To take the case of the billiard balls again, when we see one ball moving towards another, and we know that the balls are solid and that the second ball stands in the way, then certain effects are quite excluded. The moving ball cannot, for example, just pass through the second ball and come out the other side continuing at the same speed; nor can the first ball stop at *exactly* the same place as the second ball (at best they must be adjacent); nor can one of the balls suddenly vanish, and so on and so forth. The qualities of the balls determine the kind of effect that the impulse of the first ball will have on the second. We can only imagine radically different effects by imagining the balls as having - or suddenly adopting - radically different qualities.

This is not true in the case of the magnet. Here the effect does seem to be genuinely opaque. One could study the magnet - a small lump of metal - for as long as one liked and one would discern no quality by means of which the attractive power could be predicted. In the case of magnetism it seems fair to say, that the effect is known, at least in the first place, by a constant conjunction of events: if we put this lump of metal near the iron filings they will be drawn to it in a characteristic way, but the nature of the power is hidden. For this case Hume's theory works very well.

Of course, we might be able to set up the *appearance* of a necessary connection by saying that the magnet must draw the iron-filings because it has the quality, or power, of attraction. But this would be to play a game with words. It would be like the medical student in Molière's play *Le Malade Imaginaire* responding to the question 'how does opium make a person sleep?' by saying 'because it has a soporific power *(virtus dormitiva)*'. This may have satisfied the scholastic examiners in Molière's satire, but it would not satisfy someone who really wanted to know how opium acts on us. Likewise, saying that a magnet draws iron-filings because it has an attractive power does not make the cause in the least bit transparent.

A distinction can, then, be drawn between relatively transparent causal connections which arise from the observable qualities of the objects involved and causal connections which do not arise from any perceivable qualities and are known by the constant conjunction of events.

I now wish to ask why Hume was led to ignore this distinction in intuitive transparency and to proceed with his levelling definition of all causal connections in terms of constant conjunction. Part of the answer, I think, lies in his definition of a cause. For Hume it is a 'necessary connection'. When Hume searches his experience of the inner and outer worlds for an impression of the causal nexus he has a very demanding conception of what he is looking for. So demanding that it is not really surprising that he ends up with nothing. A cause, on Hume's view, must produce its effect with an absolute necessity comparable with geometry or logic. If we really had a grasp of what is a cause in the strong, traditional sense, Hume argues, we would know why there cannot arise anything else other than the pre-determined effect.

So far I have been careful to describe the impulse of billiard-balls as 'relatively' transparent. It is not completely so and thus fails to meet Hume's rigid requirement. Although we can exclude certain effects (such as the first billiard ball passing through the second or the train passing through the buffers) we cannot exclude *all* possibilities. Perception of the qualities of the two balls gives us only a framework within which we can narrow down the kind of effects that are conceivable. Hume is certainly right to say that there is still indeterminacy for someone relying on a knowledge of the perceivable qualities of the balls. For example, there is no reason why the first ball should not stop just before hitting the second, or why the motion of the second ball should not take any number of directions. The necessary connection Hume is seeking is 'all or nothing': either we can say exactly and incontrovertibly what will happen when we have knowledge of a cause, or we must accept there is no connection, but only a conjunction of quite independent events. It is because we cannot specify exactly what the effect will be of the collision of two billiard balls (but only a framework to which the effect must conform) that, in Hume's view, we are dealing with a customary sequence not a necessary relation.

Hume would have whole-heartedly agreed with the early Wittgenstein, who in his *Tractatus Logico-Philosophicus* wrote, 'A necessity for one thing to happen because another has happened does not exist. There is only *logical* necessity.'[8] But of course if one looks for logical necessity outside logic one should not be surprised when one cannot find it.

But there is another deeper reason for Hume's dismissal of the relative transparency of mechanical causation. When Hume is thinking of causal relations he is not thinking of them as primarily being between objects. It is true he still talks of 'objects'. But he cannot mean objects in the sense of three-dimensional bits of matter. He has in mind 'objects of perception', or more precisely the content of 'ideas'. This can be seen in his two definitions of causation which both speak of relations between 'objects', and when Hume supplies examples it becomes clear that these 'objects' are actually perceived events. For example his first definition of causal connection runs as follows:

> An object followed by another, and where all the objects similar to the first are followed by objects similar to the second.[9]

The example that Hume then gives of this is:

> this vibration [of a string] is followed by this sound, and ... all similar vibrations have been followed by similar sounds.[10]

Vibrations and sounds are events. It is true that they may be, at least on one interpretation, reducible to objects in the traditional sense: the vibration is presumably a quality of the string in question and the sound might be thought of as constituted by waves of movement in the particles of air. But this is not what Hume has in mind. He is thinking of events as contents of perception. In fact he is quite explicit about this elsewhere in chapter VII of the first *Enquiry* 'Of the idea of necessary connexion'. In one passage, for example, he moves from talking about 'events' to talking about 'objects' in the sense I have outlined:

> It appears, then, that this idea of a necessary connexion among *events* arises from a number of similar instances which occur of the constant conjunction of these *events*; nor can that idea ever be suggested by any one of these instances, surveyed in all possible lights and positions. But there is nothing in a number of instances, different from every single instance, which is supposed to be exactly similar; except only, that after a repetition of similar instances, the mind is carried by habit, upon the appearance of one *event*, to expect its usual attendant, and to believe that it will exist.

> This connexion, therefore, which we feel in the mind, this customary transition of the imagination from one *object* to its usual attendant, is the sentiment or impression from which we form the idea of power or necessary connexion.[11]

What influence does the primacy of events in causality have? One thing is clear, if we think primarily of perceived events rather than enduring physical objects, then the mechanical cause of impulse loses its transparency. This is because events are not the kind of things that bring certain qualities to different situations; they are rather 'moments of happening'. This makes it very easy to pass from one kind of happening to another quite different one. Even an enduring object can be understood as a succession of events. For example a banana might be loosely described as yellowness of a characteristic curved shape and tactile feel perceived here at times t, t_1, t_2 and so on.[12] A moving train would be a large, streamlined, appearance of a characteristic shape and feel, perceived in a progression of places at times t, t_1, t_2 and so on. There is no reason why a train understood thus should not suddenly appear beyond the buffers. An event-train is really a ghost-train.

Roy Holland, in a valuable, neglected article, 'The Link between Cause and Effect',[13] points out how the analysis of causation in terms of discrete events rather than in terms of the qualities of objects, or 'stuff', makes constant conjunction the only viable account of causal connection. He writes that 'the wonderfully diaphanous relationship that [causality] turns out to be is a consequence of the style of analysis and is really only what was to be expected of a relationship whose terms are...events. Since events lack all such properties as hardness, springiness, liquidity, porosity, causticity and tensile strength...' For someone, like myself, naïve enough to start from an ontology of physical objects, Hume's reduction of all causal connections to constant conjunction is not as compelling as it might otherwise seem. In fact the mechanical philosophy, despite its many shortcomings, still has at least one thing to be said for it: it recognised that some causal interactions of external objects are more transparent than others.

Notes

1 For someone interested in the historical significance of the *Enquiry* two recent and very useful sources are: Peter Millican (ed.), *Reading Hume on Human Understanding* (Oxford: OUP, 2002), essays 1 and 2 (27-96) and Stephen Buckle, *Hume's Enlightenment Tract* (Oxford: OUP, 2001).

2 I will refer here only to the *Enquiry concerning Human Understanding*, Third Edition, ed. Peter Nidditch (Oxford: OUP, 1975). This article is meant only as an interpretation and critique of the account of causation presented in the *Enquiry*. What the *Treatise* has to say is not, I think, always in harmony with the *Enquiry*. Since Hume himself wrote, towards the end of his life, that the *Enquiry* 'may alone be regarded as containing his philosophical sentiments and principles' *(Enquiry, 2)*, I find exclusive concentration on the account there justifiable.

3 See Newton, *Opticks*, Query 31, reprinted in Cohen, I.B. and Westfall, R.S. (ed.), *Newton: Texts, Backgrounds, Commentaries* (Norton, 1995), 40-55.

4 *An Essay concerning Human Understanding*, ed. Peter Nidditch, (Oxford: OUP, 1975), II.viii.11.

5 Of course, there were many hypotheses about how the firework and the magnet worked. The point being made here is just that these cases must be explained using some kind of imperceptible mechanism since what is detected by the senses does not make the causal relation transparent.

6 *Enquiry*, 28-29

7 *Enquiry*, 63.

8 *Tractatus Logico-Philosophicus*, trans. C.K. Ogden (London: Routledge, 1922), 6.37.

9 *Enquiry*, 76. (The original is in italics.)

10 *Enquiry*, 77.(The original is in italics.)

11 To highlight the terminological point I am making, I have removed Hume's emphases and added my own.

12 This kind of analysis is offered in the *Treatise* (I.iv.2) where Hume argues that an enduring object is constructed by the imagination by running together distinct but resembling, perceptions.

13 This is contained in his collection *Against Empiricism: On Education, Epistemology and Value* (Oxford: Blackwell, 1980), 210-228

Pierre Cruse

[Scientific Realism]

Should we believe in Unobservable Entities?

One of the most notable features of modern science is that it explains phenomena we observe by postulating entities that we do not. Many of the explanations appeal to entities that are too small to be seen – the TV works because electrons are being fired at the screen and causing it to illuminate in certain patterns; you have the eye colour you do because you inherited DNA that codes for eye colour when you were conceived; you have a fever because a virus is attacking your immune system. Some appeal to entities that are too big, or too far away – radiation from distant stars has a slightly lower frequency than we expect because the universe (too big to see) is expanding; it isn't expanding as fast as we think it should be because the universe contains 'dark matter' – objects (e.g. planets) in space that do not emit light, so cannot be detected by normal astronomical observation. And some entities are unobservable not because of their size, shape, or distance, but simply because of their nature – metal objects get attracted to a magnet because of the magnetic field it causes, the mass of the object causes it to resist acceleration, but neither the field nor the mass itself can directly be observed.

Entities falling into these categories pervade modern science. But the fact we don't actually directly observe them raises the question: do we know they are really there, and if so how? These questions divide philosophers of science into *scientific realists* – who believe in the reality of theoretical entities – and *anti-realists* or *instrumentalists* (after the view that theoretical postulates are just instruments for generating predictions) who do not. In this article I will look at some of the reasons which have led philosophers to take up these positions, and put forward my own view on how some of these disputes might be resolved.

Scientific Realism and Explanation

Let us begin by looking at how one might justify realism about theoretical entities. Realism is probably the most intuitive position, since most would probably assume that what our best scientific theories say about the world is true. After all, do scientists not have *evidence* for the claims that they make about the world?

Well, let us ask what sort of evidence there is for holding that, say, electrons exist. Although the name 'electron' was used for various hypothetical posits beforehand, electrons are said to have been 'discovered' by physicist J. J. Thomson in 1897. A potted history leading up to Thomson's discovery is this. If one places an 'anode' and a 'cathode' – two metal terminals connected to a wire carrying an electric current – and places them inside a vacuum tube, one can create a visible 'ray' that travels between the anode and the cathode.

[Realism]

During the latter part of the nineteenth century there was some controversy about what these rays are – do they consist of waves or particles? J. J. Thomson was the first to succeed in showing that the rays were deflected by an electric field, showing that they must consist of negatively charged particles; and he was able to measure the mass/charge ratio of these particles. The evidence for the existence of the electron, then, seems in Thomson's time to have been that the existence of tiny charged particles could explain the various results available about the behaviour of cathode rays – their deflection, their power of penetration, the fact that they can do 'work' (they can, for example, turn a 'windmill' placed in their path). In short, electrons were thought to exist because they were the *best explanation* of results relating to cathode rays.

The fact that the existence of electrons would best explain experimental results would seem to give us strong grounds for thinking that electrons are real. But let us ask why we think this is true. To draw this conclusion we need to use the following principle, known as the principle of 'inference to the best explanation' or IBE:

We observe evidence E

Theory T is the best explanation of evidence E,

Therefore,

Theory T is true

But we might well ask what justification there is for thinking that this principle will give us true conclusions when the premises are true. Note that it doesn't seem to do this as a matter of *logic*. Compare the principle of IBE with the principle of *modus ponens* (the principle that if it is true that *P*, and it is true that *if P then Q*, then it is true that *Q*). In the case of modus ponens, it looks as though it is *impossible* for *P* to be true and *if P then Q* to be true, but *Q* false. But it doesn't seem impossible that the best explanation of some evidence is false. For example, in the case of electrons, the evidence certainly seems to suggest that electrons are particles – since waves don't get deflected by electric fields – but it is surely *possible* that they are not particles at all, but a special sort of wave that violate the usual laws about the way that waves behave, and *do* get deflected by electric fields. The reason we think they are not seems not to be that we think this impossible, but simply that it seems highly *unlikely* to be the case, since it involves all sorts of needless complications to our theory. But this just raises the question again – is there any reason to think that the best explanation of our evidence is *likely* to be true?

The No Miracles Argument

There is another argument that scientific realists often appeal to in order to justify the view that the best explanations of our evidence are likely to be true. We can see how it works by looking at an analogy. Suppose that a friend of yours comes up with an implausible conspiracy theory, for example, that the government has rigged the result of the national lottery, so that a prominent member will win it at the next draw. Then suppose that a prominent government member does in fact win the next lottery draw. In that case it is *possible* that your friend's theory is false, and that the government member won just by chance. But for this to be the case your friend's prediction would have to have been incredibly lucky. Given her successful prediction, a more reasonable conclusion seems to be that her theory was actually true.

Something similar might be said in the case of electrons. Suppose I make a theoretical claim about unobservable entities (e.g. that cathode rays consist of electrons), as an explanation for certain phenomena (e.g. the fact that they get deflected in electric fields). Suppose I then predict on this basis that certain further things will happen (e.g. the cathode rays will turn a 'windmill') and that then these things *do* in fact happen. Now if my theory was false, it seems that the truth of my prediction would have to be put down to mere good luck – if cathode rays *don't* consist of electrons, then there is no reason to expect that they should behave as my theory says. But if my theory is true – and cathode rays really do consist of electrons – then we should expect them to behave just as I said they would. As with the conspiracy theory, if the events that the theory predicts genuinely occur – especially if the events are surprising or novel events – it seems much more reasonable to conclude that the theory was true than that it was false.

This observation is the basis of what has become known as the 'no-miracles' argument for scientific realism. Philosopher Hilary Putnam puts it succinctly as follows: 'the positive argument for realism is that it is the only philosophy that does not make the success of science a miracle'.[1] His idea is basically the one we have looked at. Many scientific theories throughout history – which were put forward because they were thought to be the best explanations for known evidence – led to further surprising and novel predictions, which in fact came true. Now according to the scientific realist this

is just what we would expect, since those theories were really true. But if anti-realism or instrumentalism is true – and theories do *not* make true claims about the unobservable – then the ability of theories to make true predictions would be inexplicable; nothing short of an incredibly lucky fluke or a 'miracle'. Thus, the scientific realist says that it is much more reasonable to suppose that successful theories really are true.

Scientific Realism and the History of Science

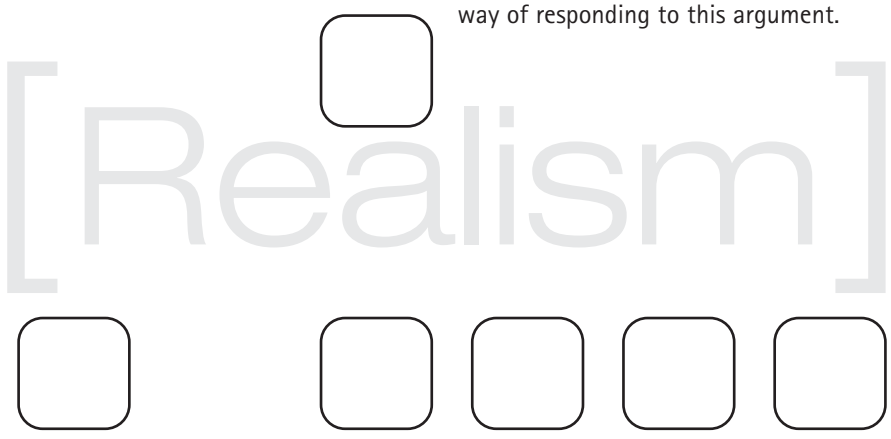
Many scientific realists hold that the no-miracles argument provides some of the strongest support available for realism. However, by no means all philosophers of science have been convinced by the argument. For the remainder of this article I will look at one of the major reasons why not.

The no-miracles argument asks us to conclude from the fact that certain theories in science have been very successful in predicting phenomena that they are true, and that the entities they postulate exist. But when we reflect on what has happened in the *history* of science, it becomes obvious that this is a very problematic claim.

We can begin to see why it is problematic by looking at our current attitude to theories of the past. It is undeniable that the history of science contains a large number of theories which we now regard as false. Many we now regard as *completely* false. For example, Aristotle believed that the planets moved around the earth embedded in crystalline spheres; Galen held that blood was produced in the liver, pumped by the heart around the body and then simply consumed; Descartes thought that gravity worked by the contact of swirling 'vortices' of solid particles that permeated the whole of space. These theories, though they were put forward with what at the time were good reasons, seem from the perspective of modern science to be fundamentally mistaken. Other theories we regard as false, but for more subtle reasons. For example, Newtonian mechanics is almost exactly true when we consider medium-sized objects moving at low speeds, but needs to be corrected to deal with phenomena at the very small or very large levels, or objects moving at very high speeds. Although the corrections are crucial, the discrepancies are so subtle they would probably have been undetectable in Newton's time.

A second feature of the history of science is that many of these theories despite being false were to a greater or lesser degree successful in explaining and predicting observed facts. For example, Aristotle's theory of crystalline spheres, though false, was able to predict the motions of the planets to quite a high degree of accuracy. And of course Newton's theory was able throughout the period of over 200 years when it was the dominant theory in physics to generate a huge variety of successful predictions. For example, the observation that the orbit of Uranus diverged from the predictions of Newton's theory led to the prediction that a planet existed outside it, pulling it slightly out of its orbit – this led to the discovery of the planet Neptune in 1846.

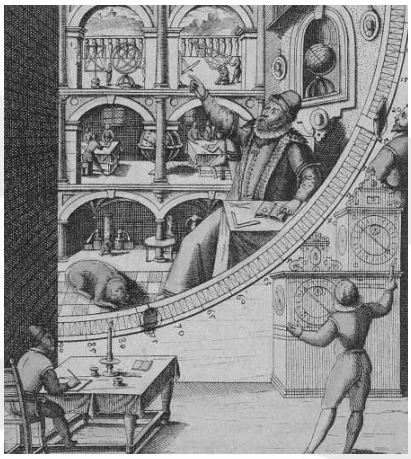
Note, then, what consequence these features of the history of science have for scientific realism. According to the realist we should believe that successful theories are true because otherwise their success in making predictions would be completely inexplicable. But in the history of science we find a number of theories which are undeniably both successful, *and false*. If this is right then the claim that successful theories are always true looks downright untenable. Moreover, since it looks as though the only justification we have for thinking that *any* theories are true is that they are successful in explaining and predicting experimental results, the fact that many theories have been successful without being true even undermines our justification for thinking that the theories we have *now* are true. This problem has become known as the 'pessimistic induction', since it suggests that generalising from the history of failures in science, we should be pessimistic about the chances of our own theories being true. The realist clearly needs some way of responding to this argument.



A More Careful Realism

In fact, given the way we have described the problem, the realist has quite a convincing response. Consider again the list of theories that we have said were false. Some of these theories are false, a realist might say, but not disastrously so. For example, although Newton's theory is false, it is also thought to be a 'limiting case' of subsequent theories such as special relativity, that is to say, Newton's theory is true if we assume that the speeds at which objects travel are vanishingly small in comparison with the speed of light (an assumption that is almost exactly true of the 'middle-sized' bodies we find on earth). One could therefore claim that although Newton's theory is strictly false, it is still *approximately* true, or *close to the truth*. The realist's claim can therefore be stated slightly more carefully to incorporate cases like this. Instead of saying that successful theories are *completely true*, the realist can claim that they are generally *approximately* true, and that this, rather than their actual truth, explains their success.

The realist is also in a position to deal with the *completely* false theories that we mentioned. Contrast the sorts of predictions that Newton's theory was able to make with those of Aristotle's theory of crystalline spheres. Newton's theory, as we saw, was able to generate predictions of phenomena that were unknown before, such as the existence of the planet Neptune. Aristotle's theory of crystalline spheres, on the other hand, while being able to explain the known motions of the planets to a certain degree of accuracy, was unable to come up with anything genuinely *novel*, that wasn't known to occur before the theory was proposed. The realist can thus point out that it isn't really a miracle if a false theory explains results that are already known to occur: these could just be 'built in' to the theory, and give us no reason to think it is true. The real 'miracle' would be if a false theory made a number of correct, *novel* predictions, about things that weren't known to occur before the theory was formulated. Thus, the realist can claim that we should only believe that theories that are successful in making novel predictions are approximately true, pointing out that many of the theories we now think are completely false did not make genuinely novel predictions.



Deeper Problems

Unfortunately, there are some remaining historical examples which seem to pose a problem for even this more cautious version of realism. One of the most notorious examples is the case of the optical aether. Before the nineteenth century, most scientists believed, after Newton, that light consisted of small, solid particles. However, around the beginning of the 19th century, it began to be realised that many optical phenomena – for example the pattern of light and dark patches created caused by interference when light is shone through two slits – could better be explained by supposing that light was constituted of waves. At the time, though, all known cases of waves occurred in a *medium*, for example in water, or on a string; and it was assumed that light waves must too. Thus scientists postulated an *aether*, which was a jelly-like, though very rigid, substance that permeated the whole of space: the immensely fast transverse (i.e. side to side) vibrations of this substance were thought to be light waves.

Aether theories were certainly successful in predicting results. In a famous case, French scientist Augustin Fresnel had submitted an entry for the Grand Prix of the Académie des Sciences in Paris in 1819, and a member of the judging panel, Siméon-Denis Poisson, worked out that Fresnel's theory – based on the waves-in-aether theory of light – would imply that if light were shone at a small opaque disc the shadow would have a small bright spot in the centre. This rather surprising prediction was later demonstrated to be true, and Fresnel was awarded the Grand Prix, despite the opposition of most of the panel members to the wave theory he defended.

The problem for scientific realism is that we now believe that there is no aether. Since Einstein's work around the beginning of the 20th century, light waves have been explained in terms of oscillations in electromagnetic fields: but fields can exist in empty space, and do not require the solid aether that nineteenth-century physicists postulated. This makes the aether a rather more problematic case for the realist than theories such as Newtonian mechanics or Aristotle's

theory of crystalline spheres. In the case of Newtonian mechanics the realist could admit the theory was successful, but call it approximately (though not completely) true, and in the case of Aristotle's theory the realist could deny that the theory was really successful. But in the case of aether theories it looks as though neither move is going to work. Aether theories were certainly successful. But given that the central entity they talked about just doesn't exist, it is very difficult to argue that they are in any meaningful sense 'approximately true'.

Although we have focussed on the aether, critics of realism point out there are a number of further theories which were genuinely successful, but which also seem crucially to involve non-existent entities. These include the phlogiston theory of chemistry, the caloric theory of heat, and others.[2] If each of these theories is indeed successful, but completely mistaken, then the scientific realist's claim that successful theories are generally close to being true looks in dire trouble.

A Realist Reply

The problem of apparently successful theories that are now known to be based on fundamentally wrong assumptions has been known in recent years to be one of the major difficulties with scientific realism, and a number of philosophers of science have attempted to address it.[3] In this final section I will describe part of what I consider to be the best response to this problem.

To begin to see how we might reply, let us compare the case of the optical aether with another entity which was being investigated at around the same time, the chemical atom. Although the idea that all matter consists of small indivisible particles goes back to the ancient Greeks, an important stage in the development of the modern concept of the atom was the proposal of English chemist John Dalton's atomic theory in 1808. Dalton proposed that each chemical element was comprised of identical atoms of a certain mass (different for each element) which combined in fixed small whole-number ratios to form compounds. This theory allowed him to use available data about the relative weights of different elements that go into forming compounds to determine the relative weights of atoms of different elements, something that had not been possible before him. However, Dalton's conception of an atom was that atoms of each element are merely solid, indivisible spheres. This model is clearly quite different to the modern view of atoms as comprised of protons and neutrons in a nucleus surrounded by orbiting electrons.

Suppose we compare, then, the relative merits of Fresnel's aether theory and Dalton's atomic theory from the modern point of view. The aether theory was correct in saying that light was essentially a wave-like phenomenon, and led Fresnel to what are from our point of view mathematically very accurate descriptions of interference and polarisation and other optical effects. However, Fresnel was wrong in thinking that the carrier of light waves was a solid aether. Dalton, on the other hand was right about the fact that atoms of different elements could combine in whole-number ratios, and right in thinking that this fact could be exploited to discover the relative atomic weights of different elements. However, he was wrong in some of his particular claims about relative weights (e.g. those of hydrogen and oxygen, since he thought water was formed from hydrogen and oxygen in a 1:1 ratio, rather than 2:1 as we think nowadays), and wrong in his view about the nature of atoms.

There is therefore surely a case to be made that Fresnel's theory is at least as accurate or 'close to the truth' as Dalton's atomic theory. Maybe it is even *closer* to the truth, since it was more accurate in describing optical phenomena like polarisation and interference than Dalton's was in describing relative atomic weights. But despite this fact most people would still think that when Dalton talked about atoms he was talking about the same things as we now call 'atoms' – Dalton's 'atom' existed, in other words – whereas when Fresnel talked about the 'aether', he was talking about something that didn't exist.

With this in mind let us go back to the main claim of the 'pessimistic induction'. According to this argument, Fresnel's theory cannot be said to be approximately true because the central entity that it postulates – the aether – does not really exist. But comparing the aether theory with Dalton's theory it seems the aether theory is arguably *better* – i.e. closer to being true – than Dalton's theory even though Dalton is talking about an entity that *does* exist. This suggests that whether the entities a theory postulates 'really exist' is not, after all, so crucial a question in deciding whether the theory is approximately true.

I claim we can explain this as follows. Part of the reason we think that the aether doesn't exist is that some of the crucial features that Fresnel attributed to the aether – e.g. the fact that it is a solid medium – don't in reality apply to anything. But part of the reason is just the fact that at a certain point in science the community simply decided as a matter of convention not to use the term 'aether' any more. In fact it was Einstein who was first responsible for dropping the concept of the 'aether', when he claimed it was 'superfluous' according to his special theory of relativity. However, we could easily imagine that Einstein had instead decided to continue to use the word 'aether' to refer to electromagnetic fields; if he had we would probably now say that the aether really existed. The fact that this didn't happen, and that we continue to say there is no aether, does not seem to be relevant to whether theories that went before him – e.g. Fresnel's – were approximately true. My claim, then, is that it is partly just a matter of convention that we say the aether doesn't exist, and it doesn't mean that aether theories could not have been approximately true.

Can the realist also give some positive grounds for thinking that aether theories *were* approximately true? I think this is easily done. Whether or not we want to say that the aether exists, we know what Fresnel said *about* the aether. For example, he claimed that:

a) It vibrates transversely (i.e. from side to side): these vibrations constitute light waves, and explain interference, polarisation, etc.

b) It is a universal, solid, very rigid, jelly-like substance.

Now we know that b) is not true of anything. However, there *is* something that has property a) – the electromagnetic field. So we can certainly say that there *exists something* with many of the important features that Fresnel thought the aether had. Moreover, that there really does exist something with feature a) seems like a good explanation of why Fresnel was able to correctly predict the results of interference and polarisation experiments. It seems to me that this gives us sufficient grounds to say that Fresnel's theory was approximately true whether or not we say that the aether 'really existed'.

Conclusion

We have seen that a strong argument for scientific realism is that if our most successful theories were not at least approximately true then their success in predicting novel, previously unknown, phenomena would be very difficult to explain – it would look like a lucky coincidence. However, a challenge for the realist comes from the fact that many theories in history were successful, whereas modern science tells us that they were fundamentally mistaken in the way they describe the world.

My claim is that this problem is not as bad for the realist as it seems at first. One of the major reasons we think certain scientific theories of the past were mistaken is that we think the central entities they postulated – such as the aether – don't exist. But the fact we say this is just a matter of convention, and doesn't necessarily mean that the theory in question isn't approximately true. By looking at exactly what the theories said *about* the entities they postulated, we find that even theories that postulated entities that we think didn't exist say a lot of things that we would regard as true. The realist still has a lot of work to do to persuade us that successful theories are almost *always* approximately true, but I don't think things are quite as bad as the 'pessimistic induction' makes out.

Notes

1 Hilary Putnam, *Mathematics, Matter and Method: Philosophical Papers vol. 1* (Cambridge: Cambridge University Press, 1975), 73

2 The original problem was raised in an article by Larry Laudan, 'A Confutation of Convergent Realism', *Philosophy of Science* 48 (1981), 19-49, reprinted in D. Papineau ed. *The Philosophy of Science* (Oxford: OUP, 1999)

3 Two recent examples are Philip Kitcher, *The Advancement of Science* (New York: Oxford University Press, 1993) ch 5, and Stathis Psillos, *Scientific Realism: How Science Tracks Truth*, chs 4-7, (London: Routledge, 1999)

Christopher Norris

Realism, [Reference] & Possible Worlds: the Approach via Modal Logic

Introduction

This article offers a critical review of various current debates between realism and anti-realism in epistemology and philosophy of science. In particular, it focuses on the claim of some philosophers – chief among them Saul Kripke and the early Hilary Putnam – that modal logic (i.e., the branch of logic concerned with matters of necessity and possibility) provides strong support for a realist approach to these issues.

I begin by discussing Kripke's arguments for the existence of *a posteriori* necessary truths or those that have to be discovered through some process of scientific enquiry but which none the less hold as a matter of necessity in any world physically congruent with our own. This claim is backed up by Hilary Putnam's famous series of 'Twin-Earth' thought experiments designed to make the case for modal realism, that is, the idea that certain names (prototypically natural-kind terms like *gold, water, acid, lemon or tiger*) have their reference fixed across all 'possible worlds' by just what it is to be an entity of just that kind. Thus the reference-fixing may be in virtue of its molecular constitution (*water* = H_2O) or its subatomic structure (*gold* = 'the metallic element with atomic number 79'). Likewise *acids* have the property 'proton-donor' which defines their reference more precisely than earlier descriptions like 'corrosive' or 'apt to turn litmus-paper red'. In the same way *tigers* and *lemons* are distinguished by reason of their possessing certain distinctive genetic or chromosomal features, rather than through descriptive attributes such as 'striped, carnivorous, and fleet-footed' or 'yellow of skin, with a white rind, and bitter in taste'. Those features belong to them essentially and did so even at a time when nobody possessed the relevant scientific knowledge. Their usage was therefore 'truth-tracking' or 'sensitive to future discovery', rather than failing to refer altogether or – as the rival (descriptivist) account would entail – involving so disparate a range of imputed properties that we cannot think of early users as referring to the same kind of thing. This approach also claims to resolve the problem with anomalous items such as unripe (green) and sugar-saturated lemons or fleet-footed, striped, and carnivorous creatures that just happen *not* to be tigers.

What the Kripke/Putnam approach thus provides is a means of conserving fixity of reference across large (even radical) episodes of scientific theory-change. However – I argue – this benefit is lost if one follows a thinker like David Lewis in asserting the *reality* of all those 'possible worlds' where things are conceived to differ with respect to certain distinctive (e.g., microstructural) features that define what it is to be a thing of that kind in the world that we actually inhabit. According to Lewis, this claim is on a par with the realist view of mathematical statements (about numbers, sets, classes, etc.) as referring to a realm of abstract, mind-independent objects which are none the less real for our having no means of perceptual or epistemic access to them.

[Possible Worlds]

Yet the result of adopting a Lewisian modal-realist ontology is to blur certain crucial distinctions, as between the order of *a priori* or trans-world necessary (e.g., logical and mathematical) truths and those which obtain as a matter of *a posteriori* necessity in that particular world – our own – where water is H_2O, gold has the atomic number 79, tigers and lemons have a certain genetic-chromosomal structure, and so forth. That is to say, modal 'realism' of this type entails the acceptance of a massively expanded ontology whereby – as W.V. Quine puts it – there is just no distinguishing, in point of truth, between statements which quantify over such diverse items as centaurs, Homer's gods, mathematical sets and classes, and brick houses on Elm Street. This despite Quine's outlook of radical empiricism and his well-known objections to modal logic as a source of needless philosophical confusion.

My article thus puts the case for modal realism as a domain-specific approach that involves different orders of truth-claim as regards mathematics, the physical sciences, and other areas of discourse. It also points out some of the shared problems with a descriptivist approach that opens the way to a doctrine of wholesale ontological relativity across shifting scientific paradigms and a Lewis-type 'many-worlds' modal conception that likewise tends to blur such basic distinctions. I conclude that modal logic is a powerful resource in the realist's philosophical armoury but one that requires exceptionally careful handling if its application is not to produce unwanted ontological commitments.

I

Kripke is known chiefly for the arguments advanced in his book *Naming and Necessity* where he proposes a causal theory of reference as against the once prevalent descriptivist theory descending from Frege and Russell (Kripke 1980; also Frege 1952; Russell 1905). According to the latter we pick out referents (objects or persons) through a cluster of descriptive attributes which serve to specify and hence to individuate just those uniquely designated objects or persons. Thus, to repeat, people once referred to *gold* under some such description as 'yellow, ductile metal that dissolves in nitric acid', whereas now it is defined (for scientific purposes) as 'metallic element with atomic number 79'. Or again: when we refer to a historical individual such as *Aristotle* we do so by applying certain salient descriptions such as 'pupil of Plato', 'tutor of Alexander', 'author of *The Poetics*, the *Prior Analytics*, etc.'

Hence Frege's cardinal dictum that 'sense determines reference', i.e., that in so far as such proper names refer it must be in virtue of our grasping the relevant descriptive criteria. On the contrary, Kripke maintains: the reference of *gold* was fixed by an inaugural 'baptism' or act of naming, and has since held firm despite and across all subsequent changes in our knowledge concerning its nature, identifying features, physical properties, microstructural constitution, or whatever. Otherwise – on the descriptivist theory – every time that we made a new discovery about gold we should have to say (absurdly) that 'gold is not gold', since our previous beliefs had turned out false or inadequate, and it was just those beliefs that had fixed its reference. Or again: if we discovered that Aristotle had *not* in fact been a student of Plato, tutored Alexander, authored the *Poetics*, etc., then we should have to say 'Aristotle wasn't Aristotle'. Rather, what allows us to avoid this absurd consequence is the Kripkean causal theory of naming and necessity whereby 'Aristotle' refers to just that historical individual who was so named and whose identity was fixed – necessarily so – at his moment of conception. (More exactly: at the moment when his father's sperm fertilised his mother's egg.) So likewise: had George W. Bush not become US President as a result of the controversial 2001 election – had there been a recount of the Florida vote, let us say, and the Supreme Court not decided against it by upholding the official outcome – then he would still have been the selfsame George W. Bush despite this significant change of descriptive attribute. In the case of natural-kind terms like *gold* the argument works in a similar way: what the term picks out is just that substance (i.e., the element with atomic number 79) which received its name through an initial act of baptism and that has always since then been designated 'gold'. Thus the term referred to the identical stuff even when nobody knew about atomic numbers and when people had to make do with rough-and ready descriptive attributes. From which it follows that they were always wrong in mistaking fool's gold (iron pyrites) for the genuine item despite its superficial resemblance. What made them wrong was (1) the necessity that 'gold' should refer to *gold* in any conceivable world where the substance thus named possessed just that kind of uniquely distinctive microstructure, and (2) the linguistic 'chain' of transmission whereby its reference had been preserved through

every shift in its associated range of descriptive criteria (Kripke 1980).

On the strength of this argument Kripke advances some far-reaching proposals with regard to modal logic, that is, the branch of logic having to do with matters of possibility and necessity. In brief, he makes a case for the existence of *a posteriori* necessary truths – like those about the atomic constitution of *gold* or the genetic-chromosomal identity of *Aristotle* - which are neither analytic, i.e., true-by-definition, nor *a priori*, that is to say, self-evident to reason, but which none the less hold necessarily in any world where their referents exist or once existed. (See also Linsky [ed.] 1977; Schwartz [ed.] 1977; Wiggins 1980.) Thus gold *cannot but* be that kind of stuff in all worlds physically compatible with ours in respect of their constituent natural kinds while Aristotle *cannot but* have been just that individual in all worlds where his identity was fixed by the self-same act of conception. And of course one could multiply similar examples, such as *water* having the molecular structure H_2O just in virtue of its being water, or acids being proton-donors just in virtue of their being acids, or tigers possessing a certain chromosomal make-up since that is what constitutes the membership-condition for any creature that belongs to the species 'tiger' (Putnam 1975a, 1975b, 1975c). Of course these criteria haven't always applied since 'water' was once defined vaguely as the kind of stuff that fell as rain, filled up lakes, was liquid under normal ambient conditions, boiled or froze at certain temperatures, possessed certain useful cleansing properties, etc. In the same way our knowledge of acids advanced from 'acid = corrosive to certain metals, sour-tasting in dilution', etc., to 'acid = having the property of turning litmus-paper red', to 'acid = proton-donor'. Nevertheless the term 'acid' may be held to have referred to the same natural kind despite and across all these changes of descriptive paradigm, just as 'tiger' has continued to pick out the same animal species whether vaguely defined as a 'large, fast-running, cat-like creature with stripes' or with reference to its chromosome structure.

In this respect – so the argument goes – such names are 'truth-tracking' or 'sensitive to future discovery' (McCulloch 1995). That is to say, their usage at any given time might always turn out (now as in the past) to be based on a limited or partial knowledge of just what it is – scientifically speaking – that constitutes the kind in question. Very often it is a matter of superficial appearances, as in the case of 'gold = yellow, ductile metal' (which would also encompass iron pyrites) or – perhaps the most famous example – 'whale = large, water-spouting fish'.

However this gives no reason to conclude (with the strong descriptivists or paradigm-relativists like Thomas Kuhn) that shifts in the range of identifying criteria from one theory or classificatory system to the next can at times be so drastic as to break the referential chain of transmission and leave us at a loss to compare theories in point of their descriptive accuracy or causal-explanatory power (Kuhn 1970; also Quine 1961a). What leads philosophers to adopt this surely desperate position – i.e., the thesis of radical 'incommensurability' between paradigms – is their acceptance of Frege's cardinal precept that 'sense determines reference' along with the idea (which Frege sharply rejected) that the sense of any given term can only be specified in relation to the entire language, discourse, or received body of knowledge within which it plays a role (Frege 1952; also Dummett 1978). Thus scientists working before and after some major episode of theory-change must be thought of quite literally as inhabiting 'different worlds', or worlds that contain a whole different range of objects, properties, causal powers, microstructural features, and so forth.

Moreover, we cannot talk of scientific 'progress' in this regard since the very criteria for what *counts* as an advance in knowledge are themselves relative to this or that paradigm and hence incapable of adjudication from some standpoint of objective (paradigm-transcendent) truth. (For further discussion see Laudan 1977; Lipton 1993; Rescher 1979.) Besides, as Quine famously argued, observations are always to some extent 'theory-laden' and theories always 'underdetermined' by the best empirical evidence to hand (Harding [ed.] 1976). In which case scientists can always save a cherished theory by pleading observational error, perceptual distortion, the limits of precise measurement, etc., or alternatively save some striking empirical observation – where it comes into a conflict with a well-established theory – by making suitable adjustments elsewhere in the overall 'web of belief'. At the limit (as with certain well-known problems in the field of quantum mechanics) this might even entail some revision to the ground-rules of classical logic such as bivalence or excluded middle. (See Quine 1961a; also – for a range of views on this topic – Gibbins 1987; Haack 1974; Norris 2000; Putnam 1983.) However there is something decidedly suspect about an argument that leaves no room for such basic normative conceptions as those of good observational warrant or accordance with our best theoretical beliefs as judged by the standards of valid logical (e.g., hypothetico-deductive) inference.

Hence a main attraction, for some, of the Kripkean 'new' theory of reference: that it offers a means to avoid this unpalatable upshot of wholesale paradigm-relativism or 'incommensurability' across different theories, languages, or conceptual schemes. For if reference is fixed independently of any descriptive criteria that happen to apply from one to another paradigm then we can perfectly well explain how a term like 'electron', once introduced through the inaugural act of naming, continued to specify the same referent despite some otherwise radical revisions to its range of defining properties or imputed characteristics. (See especially Putnam 1978.) Thus pioneer usages – like that of J.J. Thomson – were descriptively and theoretically wide of the mark when assessed against our present (quantum-based) understanding, as indeed was Niels Bohr's 'planetary' model of electrons orbiting the nucleus before he abandoned that model in favour of a quantum-theoretical approach (Bohr 1934, 1958). Yet it is still the case that we can speak of Thomson, Bohr and others as referring to a certain kind of subatomic entity – the *electron* – and also as having come up with different, more-or-less adequate theories and descriptions concerning it. For what the Kripkean account of reference-fixing entitles us to claim is that Thomson set this process in train through an inaugural act of naming ('let us call "electron" the kind of thing that would explain these otherwise mysterious phenomena') and that the name then stuck – referentially speaking – despite its radical redefinition with the advent of quantum mechanics.

Thus philosophy of science can be saved from its own sceptical devices by acknowledging (1) that descriptive attributes don't go all the way down, (2) that early usages are 'sensitive to future discovery', and (3) that in the case of genuine (as opposed to empty or fallacious) object-terms their reference is preserved across even the most revolutionary episodes of theory-change. That is, the term 'phlogiston' now survives as nothing more than the name for a non-existent stuff that once figured (along with 'dephlogistated air') in a false theory of combustion while the term 'oxygen' has retained its referential good standing since we have adequate grounds to suppose that oxygen really exists and provides the best explanation of just what occurs in that process. This despite the fact that Priestley and Lavoisier – proponents of the two rival theories – conducted experiments that proved each correct by his own theoretical lights and which could arguably serve (on descriptivist grounds) to support Kuhn's case for the paradigm-relative nature of scientific truth-claims. However such ideas will appear less plausible – in fact decidedly *outré* – if one adopts the alternative Kripkean approach and takes it that the

reference of genuine (as distinct from factitious or illusory) natural-kind terms is truth-tracking and fixed by their referring to entities of just that sort.

Other instances are more problematic since they offer some leeway for reconstruing the object-terms or ontological commitments of an earlier theory in keeping with subsequent advances in scientific knowledge. Thus pre-Einsteinian talk about the 'ether' – the pervasive, intangible substance that was thought to explain the passage of light and other forms of electro-magnetic radiation throughout the universe – can be taken as co-referential with post-Maxwellian talk about the 'electro-magnetic field'. (See especially Psillos 1999; also Aronson, Harré and Way 1994; Leplin [ed.] 1984). Although the ether was shown not to exist as a result of the Michelson-Morley experiments still there is a case (so the realist might argue) for applying this retroactive principle of charity or for treating such talk as descriptively void but referentially on the right track. And again, while Black's 'caloric' hypothesis turned out to involve a false supposition – i.e., the existence of a likewise intangible fluid medium whereby to explain thermal conductivity and related phenomena – still it can be shown to have played a crucial part in developments that led to the theory of specific heat (Psillos 1999). In these cases – the latter especially – any Kripkean approach would need to be qualified so as to incorporate at least some elements of the rival (descriptivist) account. Otherwise, of course, there could be no explaining how two distinct terms with different senses and with a role in radically different physical theories might none the less be construed as referring to 'the same' (or at least to strongly analogous) kinds of physical phenomena. Indeed some philosophers have put the case for viewing the Kripkean approach not so much as an ultimate solution to problems thrown up by the Frege-Russell approach but rather as a theory which allows – and requires – some additional descriptivist component (Evans 1982; also Schwartz [ed.] 1977). Still they would mostly argue that Kripke's account is one that goes far toward resolving those problems and that it offers the best way forward not only for debates in philosophical semantics but also for epistemology and philosophy of science.

II

These claims are widely contested – not least by adherents to the 'old' descriptivist paradigm – but have all the same exerted a powerful influence on recent philosophical debate. In particular they have led to a revival of causal realism (i.e., the claim that certain kinds of object necessarily and of their very nature possess certain properties, dispositions, or causal powers) for which a main source is the Kripkean treatment of issues in modal logic.

Other thinkers – Putnam chief among them – have shown more willingness than Kripke himself to press the argument in this direction. Thus Putnam has proposed a number of ingenious thought-experiments designed to bring home the realist point that meanings 'just ain't in the head' (Putnam 1975a, 1975b, 1975c; also Norris 2002a). That is to say, what fixes the truth-conditions for our various statements concerning the physical world is not the range of descriptive criteria by which we pick out objects of this or that kind but rather the existence of just such objects with just such uniquely identifying structures and properties.

The best-known case has to do with a space-traveller from Earth to Twin-Earth who finds, on arrival, that everything looks the same as back home, including the existence of large quantities of water which fills up the lakes, falls as rain, boils and freezes at identical temperatures, etc. The only difference is that – unbeknownst to him – Twin-Earth 'water' (as referred to by the natives) has the molecular constitution XYZ, rather than H_2O.

So when the traveller exclaims with evident delight 'Lots of water around here!' he must surely be thought to have got it wrong – to have been misled by superficial or phenomenal appearances – since the stuff in question is *not* the kind of stuff that he and other Earthians standardly (correctly) refer to as 'water'. And of course the scenario can be turned around by supposing a traveller from Twin-Earth to visit Earth and likewise misidentify Earthian 'water' as just the same stuff that exists in such abundance back home, unaware as he is – not having performed the requisite chemical analysis – that this stuff is in fact H_2O and not XYZ.

There are many variations on a kindred theme in the recent literature, some (like Putnam's) designed to refine, extend, and reinforce the basic realist point while others – as I have said – adopt a more qualified approach by attempting to accommodate certain arguments from the descriptivist quarter. Then again, philosophers like Tyler Burge have argued that there is no reason in principle to restrict the Kripke-Putnam approach to natural kinds such as tigers, acids, gold, water, or electrons (Burge 1979). For the same considerations should apply just as well to artefacts or objects that don't occur naturally but which, none the less, have their reference fixed through an inaugural act of naming and thereafter passed down through a communal 'chain' of transmission that ensures a sufficient degree of continuity despite any shifts in their range of descriptive criteria. This is not to say – crucially – that the correct usage of such terms depends on the individual speaker's possessing an expert or scientific grasp of what it is that uniquely identifies the object concerned. Thus the traveller to Twin-Earth is deceived by appearances *whether or not* he happens to know that Earthian water has the molecular structure H_2O. What makes him wrong about its Twin-Earthian counterpart is the fact that there are some experts back home – physicists or chemists – who do possess that kind of expert knowledge and to whom the wider community defers should any question arise with regard to anomalous cases such as 'heavy water' or borderline (say, highly polluted or otherwise non-standard) samples of the kind. Putnam calls this the 'linguistic division of labour' and takes it to explain how someone – like himself – who has problems in distinguishing beech-trees from elms can none the less deploy those terms with a good degree of referential assurance (Putnam 1975a, 1975b; also 1988: 22-6). That is, any issue with regard to their correct usage could always be resolved (if need be) by appealing to the relevant specialist, i.e., arborological sources.

No doubt there is a sense in which arguments of this sort require that the basic position be modified so as to acknowledge the reference-fixing role of those various descriptive attributes or criteria that effectively decide what should count as expert opinion. All the same that position is by no means undermined since it still provides the best means of explaining how elms and beeches – or Earthian and Twin-Earth 'water' – can indeed be picked out as distinctive kinds whose salient (or kind-constitutive) features are those implicitly referred to when speakers use the terms in question. Thus Putnam's not knowing how to tell the difference between the two sorts of tree is made up for by the fact of his knowing that others know, just as – from a chronological perspective – we can claim that people were referring to such things as *gold, water, acids*, or *electrons* at a time when even the most expert sources could not have provided an adequate account of their constituent structures or properties. To this extent the 'linguistic division of labour' is the equivalent, in synchronic terms, of the idea that such early usages should properly be viewed as 'truth-tracking' or 'sensitive to future discovery' (McCulloch 1995). What is more, according to Kripke, it is a matter of *a posteriori* necessity that this should be the case, that is, a necessary truth about *gold, water, acids,* or *electrons* that they possess just those structures or properties that they do in fact possess, whatever the range of differing descriptions applied to them since way back when the terms were first introduced (Kripke 1980). This would also apply to terms such as Twin-Earth 'water' if we suppose the possible world in question to be one where certain natural kinds do in fact (necessarily) possess a whole range of quite distinct atomic, molecular, or genetic features. However it is crucial to Kripke's argument – at least from the realist standpoint – that we have to draw a line between logically possible worlds, i.e., those that we are able to conceive or postulate without contravening some trans-world necessary axiom of logical thought, and worlds wherein the range of possible departures from our own is subject to various specified physical constraints (Bradley and Swartz 1979; Kripke 1980; Schwartz [ed.] 1977; Wiggins 1980). For without this distinction there could be no warrant for the basic Kripke-Putnam claim, i.e., that *a posteriori* truths about the way things stand with respect to natural kinds (or this-world operative laws of nature) are also necessary truths in so far as they could not be otherwise in any world physically compatible with ours.

As I have said, such arguments have not gone unchallenged by philosophers within the analytic community. They are stoutly opposed by a sceptic like W.V. Quine who regards modal logic as a needless liability, rejects all talk about 'possible worlds' as a piece of sheer metaphysical indulgence, and adopts a naturalised (physicalist) epistemology that finds no room for such extravagant ideas (Quine 1971a, 1971b; also 1969). On the other hand they are taken to the limit – and beyond – by a modal logician such as David Lewis who argues for the literal reality (as distinct from the merely hypothetical or counterfactual existence) of all those logically possible worlds that fall within the limits of rational conceivability or which don't involve any pair of contradictory propositions (Lewis 1986; also 1973). Thus, for Lewis, there is an endless plurality of worlds in which every contingent this-world truth is negated, so that (for instance) Julius Caesar didn't in fact cross the Rubicon, or kangaroos weren't in fact equipped with heavy tails which prevent their unfortunate tendency to topple forward at every step. These worlds are just as 'real' as our own but non-actual (and hence, to us, epistemically inaccessible) since they just happen not to be the world that we actually inhabit.

On this view we should think of 'actual' by analogy with deictic or token-reflexive terms like 'I', 'here', 'now', or 'today', that is say, terms which necessarily involve some reference to a given speaker at a certain time or place of enunciation. So just as there are manifold times and places that lie beyond our first-person indexical grasp so likewise there are numerous alternative worlds whose reality is in no way affected by the mere fact that they have not been actualised in our own experience or that of persons who share our particular world. To suppose otherwise – so Lewis suggests – is the kind of parochial prejudice that must ultimately lead to downright solipsism or the refusal to credit any reality other than that which we are able to cognise from our own spatio-temporally restricted viewpoint. He also points out that if we want to be realists about mathematics then we shall have to accept that there exist certain abstract objects and associated truth-values of which we can indeed have knowledge even though they belong to a realm that by very definition cannot be accessed by any quasi-perceptual means of epistemic contact. (For further discussion see Alston 1996; Hale 1987; Katz 1998; Soames 1999.) And since mathematics is the best (most secure) kind of knowledge we possess there must surely be a place for Lewis's real but non-actual worlds together with numbers, sets, classes, and other such abstract entities. Thus we should not be over-impressed by any argument on common-sense (actualist) grounds that rejects the reality of all those possible worlds and, along with them, the only conception of mathematics that doesn't reduce to some form of shifty conventionalist or fictionalist doctrine.

Lewis is a brilliantly gifted exponent of what remains – as I have argued at length elsewhere – an exorbitant and hugely implausible hypothesis backed up by all manner of ingenious argumentation (Norris 2000). It is one that has its origins in Leibniz – the progenitor of possible-worlds talk as a device for spelling out the implications of modal logic – and which might be taken to find support (albeit from an equally exorbitant quarter) in the 'many-worlds' interpretation of quantum mechanics (Leibniz 1972; Deutsch 1997; DeWitt and Graham [eds.] 1973). However Lewis-style 'realism' is a far cry from the arguments advanced by Kripke and early Putnam with regard to the fixity of reference across all worlds compatible with ours in the relevant (e.g., physical or historical) respects. That is to say, it exploits a certain strategic blurring of the Kripkean distinction between trans-world necessary truths such as those of logic and mathematics and truths that hold good as a matter of *a posteriori* necessity, i.e., in virtue of the way things stand with regard to our actual world.

The former have to do with statements that *could not possibly* have been falsified no matter how the laws of nature lay or how events turned out in our particular world while the latter have to do with statements whose truth-value is determined – and their reference fixed – by just such intramundane laws and events. In short, what is distinctively *realist* about modal realism of the Kripke/Putnam type is its insistence on drawing such a line and thereby preventing the tendency of thought to stray over into worlds of counterfactual supposition which acknowledge no constraints on the capacity of reason to conjure up any range of alternative 'realities' subject only to certain basic logical axioms, e.g., that of non-contradiction. For this leads to such a downright profligate ontology – such an endless multiplicity of worlds all enjoying the same ontological status – that it tends to undermine the kinds of counterfactual-supporting argument ('had x not occurred, then neither would y; therefore x was a causal factor in y') that play a central role in scientific, historical, and other sorts of causal-explanatory reasoning (Hawthorn 1991; Mackie 1974; Salmon 1984).

Indeed there is a sense in which Lewis's extravagant hypothesis comes close to Quine's likewise extravagant doctrine of ontological relativity, that is, his idea that the objects or entities posited by different conceptual schemes are as many and various as the schemes themselves, and extend all the way from brick houses on Elm Street to numbers, sets, classes, centaurs, and Homer's gods (Quine 1961a, 1969). As I have said, Quine takes a dim view of modal logic since it seems to involve unacceptable consequences, such as that if it is a necessary truth that '9 is greater than 7' then it is also a necessary truth that 'the number of planets is greater than 7'(Quine 1971a: 20-21). Yet of course the latter is a contingent fact about the way things stand in our particular corner of the universe while the former is a truth-of-definition accordant with the rules of elementary arithmetic. In which case – he argues – we should stick to the first-order predicate calculus and eschew the kinds of misconceived modal reasoning that lead to such unfortunate (logically repugnant) results. However this objection can be turned back – on the Kripke/Putnam modal realist account – by distinguishing the order of trans-world necessity that applies to certain truths of logic and mathematics from the order of *a posteriori* necessity that applies to certain truths about the physical world that we actually inhabit. Moreover we can thereby resist Quine's conclusion that there is simply no difference, in point of 'reality', between the various sorts of object that have figured as posits in various (e.g. common-sense, mathematical, scientific, religious, or mythical) conceptual schemes (Quine 1961a). For one could argue that this pyrrhic conclusion is forced upon him – in large part – through Quine's refusal to apply just the kinds of reality-preserving modal distinction that would allow a more adequate treatment of metaphysical, ontological, and epistemological issues. And besides, his point about the number of planets – that modal locutions run into trouble when it comes to distinguishing necessary from contingent truths – is one that sits awkwardly with Quine's dependence on modal distinctions by way of enforcing just that logical point.

That is to say, there is a sense in which modal logic – contrary to received opinion – has a fair claim to be more basic to the process of rational (truth-preserving) argument than the first-order predicate calculus on which Quine supposedly builds his case. For that case cannot hold up except on the assumption that there exist necessary truths (like those of mathematics) and contingent truths (like that concerning the number of planets) which have to be distinguished on pain of falling into gross philosophical error. Thus:

> [g]iven that logic is concerned . . . with formulating principles of valid inference and determining which propositions *imply* which, and given that the concepts of validity and implication are themselves modal concepts, it is modal logic rather than truth-functional logic which deserves to be seen as central to the science of logic itself . . . . From a philosophical point of view, it is much sounder to view modal logic as the indispensable core of logic, to view truth-functional logic as one of its fragments, and to view 'other' logics - epistemic, deontic, temporal, and the like - as accretions either upon modal logic (a fairly standard view, as it happens) or upon its truth-functional component. (Bradley and Swartz 1979: 219)

All the same these advantages are thrown away if modal realism is pushed to the point, as in Lewis's theory, where it invites the Quinean charge of sheer metaphysical extravagance by maintaining the existence of all those non-actualised

but equally 'real' (since logically possible) worlds. Indeed – as I have said – this argument comes out pretty on a par with Quine's idea that what is real *just is* what is 'real' (within a given conceptual scheme) for all that we can possibly know, judge, or ascertain.

To be sure, Lewis has a strong case when he recruits mathematics in support of his modal-realist claim that there must be truths – such as those pertaining to various logically possible worlds – that go beyond anything knowable by means of perceptual acquaintance or epistemic contact. To reject this claim is to end up in the position of sceptical or anti-realist thinkers who declare that 'nothing works' in philosophy of mathematics since we can *either* have a notion of objective (recognition-transcendent) truth that places it forever beyond our epistemic reach *or else* a conception of mathematical knowledge that equates truth with our best methods of proof or verification. (See especially Benacerraf 1983; also various contributions to Benacerraf and Putnam [eds.] 1983 and Hart [ed.] 1996.) In which case we should have to conclude that there exist a great range of well-formed but as-yet unproven theorems – like Goldbach's conjecture that every number is the sum of two primes – that are neither true nor false since we lack (and might never produce) an adequate proof procedure. Or again, we should find ourselves driven to endorse the surely absurd conclusion that Fermat's Last Theorem was likewise devoid of an objective truth-value during the three centuries of intensive work before Andrew Wiles came up with his celebrated proof. More than that: we should be quite at a loss to explain just what it was that rendered previous attempts inadequate and that might yet conceivably turn out to reveal a flaw in Wiles's reasoning.

III

Of course Lewis's argument would count for nothing with those, like Michael Dummett, who take an anti-realist view of mathematics and other areas of discourse (Dummett 1978, 1991). On their account there is no making sense of the claim that statements can possess an objective truth-value quite apart from our capacity to find it out by some empirical or formal method of verification. Thus Goldbach's Conjecture – along with a great many others unproven theorems – would fall into Dummett's 'disputed class' of statements that are neither true nor false, as distinct from merely undecidable according to our best, most advanced or sophisticated proof procedures. This conclusion follows logically enough if one accepts Dummett's anti-realist case for the impossibility of recognition-transcendent truths, that is, his idea that any 'gaps in our knowledge' must entail the existence of corresponding 'gaps in reality'. Furthermore it is one that in principle applies across each and every area of discourse from mathematics, logic and the physical sciences to history and ethics. Thus it excludes any modal conception, such as Lewis's, which embraces not only a realist outlook with regard to abstract entities like those of mathematics and the objective (even if unprovable) truth-value of statements concerning them but also a belief in the reality of all those non-actual yet logically possible worlds and their various constituent features.

[Possible Worlds]

Now there is a lot to be said, so the realist about mathematics might feel, for Lewis's robust attitude in this respect and his insistence that if anything is to serve as a guide in such matters then it had better be our grasp of just what is required in order to make good sense of mathematical truth-claims. Yet she might well balk at the further liability introduced by Lewis's outlook of intransigent realism with regard to possible worlds and his suggestion that the case for mathematical realism stands or falls with that for the reality (as distinct from the logical conceivability) of any and every such world. Here again there is a sense, as emerged in the comparison with Quine, that by taking so extreme or ontologically profligate a view Lewis runs the risk of drowning the realist baby in the metaphysical bathwater. At any rate his version of realism is far removed from the Kripke-Putnam emphasis on distinguishing contingent from necessary truths and – among the latter – those that possess analytic (transworld) necessity from those that hold as a matter of *a posteriori* warrant. Only thus can the realist hope to produce the kind of argument that would challenge the case for anti-realism advanced by thinkers like Dummett, that is to say, an approach that treats every area of discourse as having no room for truth-apt statements whose objective truth-value transcends the limits of recognition or verification.

My own view – as should be evident by now – is that realism stands in need of such defence since we shall otherwise be wholly at a loss to explain a great many aspects of everyday as well as scientific knowledge and enquiry. Anti-realists often make much of the so-called 'argument from error', i.e., the claim that we can never be justified in asserting the truth of our current-best theories when we know that by far the greater proportion of scientific 'knowledge' to date has eventually turned out false, or else been shown (like Newton's theories of space-time and gravity) to possess only a restricted scope of application. (See especially Laudan 1981.) So why should we think that our own epistemic situation is in any way different from that which has prevailed up to now? However the realist can turn this argument around by remarking (1) that any talk of past errors presupposes our possession of other, more advanced or adequate truth-standards, and (2) that the recommended attitude of due humility concerning our present state of knowledge entails the supposition that we might yet be wrong according to (what else?) objective criteria of scientific truth and falsehood.

Thus the realist case is in no way compromised – and indeed much strengthened – by renouncing any claim to what Nicholas Rescher calls 'the ontological finality of science as we have it' (Rescher 1987: 61). Moreover there is the 'no miracles' argument which holds that we should always go for the least far-fetched or credibility-straining explanation, and should hence be sceptical of any approach – like anti-realism in philosophy of science – which would make it nothing less than a miracle that erroneous ideas should somehow have produced such a wealth of accurate predictive data and successfully applied scientific results (Boyd 1984; Putnam 1975c). In which case, according to Putnam, we have good reason to believe that 'terms in a mature scientific theory typically refer' and that 'laws of a mature scientific theory are typically approximately true' (Putnam 1975c: 290). This in turn goes along with the case for 'convergent realism' or the claim that even if our best theories so far have fallen short of the truth nevertheless they are demonstrably on the right track in so far as all the evidence points toward their having picked out a range of entities (such as 'molecules', 'atoms', and 'electrons') whose role is indispensable to further research. Thus science may be taken as converging on truth at the end of enquiry to the extent that its theories are increasingly borne out by the best evidence to hand (Aronson 1989; Aronson, Harré and Way 1994; Lipton 1993).

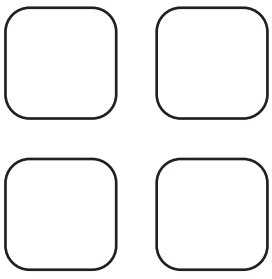
The anti-realist might readily accept all this and yet maintain – on prudential grounds – that we had much better treat atoms and suchlike as useful posits for the sake of upholding some empirically adequate theory, rather than leap to the premature conclusion that 'atoms' actually exist (van Fraassen 1980). To which the realist will once again reply that such objections miss the point since realism in philosophy of science is itself a candidate hypothesis to be judged – like scientific theories – on the strength of its explanatory virtues or its capacity to offer a plausible account of our knowledge of the growth of knowledge. That it does so better than rival hypotheses is a claim borne out by the above-cited range of arguments plus those various considerations from modal or possible-worlds logic which, as I have suggested, provide strong support for a causal-realist approach. That is, they explain how the reference of terms (including theoretical terms or names for 'unobservables' like atoms or electrons) is preserved across sometimes quite drastic episodes of scientific paradigm-change; how knowledge accrues through the

discovery of ever more detailed microstructural or depth-explanatory attributes; how theories can turn out wrong (or only partially valid) with the advent of later, more advanced or better corroborated theories; and again – most crucially for the realist – how the truth-value of well-formed statements or hypotheses might always transcend our present best knowledge or means of verification. In short, they offer strong grounds for maintaining that the burden of proof falls squarely on the anti-realist despite the current trend toward regarding anti-realism as something like a default position in epistemology and philosophy of science (Norris 2002b).

It is unlikely that sceptics will be won over by any amount of argument along these lines, whether through scientific case-studies designed to vindicate the claim of convergent realism or through the kinds of evidence that Putnam provides with his thought-experimental variations on the theme of naming, necessity, and natural kinds. Anti-realism is a doctrine so deeply bound up with certain ruling metaphysical preconceptions – most of all in Dummett's work – that it tends to adopt an across-the-board (no matter how logically nuanced) verificationist approach that treats such issues as largely irrelevant in comparison to its major thesis. Nevertheless – I would argue – they are of the utmost importance if we want to get straight about basic questions like the role of mathematics in the physical sciences or how it can be that so seemingly abstract a branch of enquiry could have offered so much in the way of applied theoretical, predictive, and explanatory power. Thus, in Eugene Wigner's memorable words: '[t]he miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand or deserve' (Wigner 1960: 237). To which the anti-realist will standardly respond with some version of the sceptical dilemma, i.e., that we can either have a notion of objective mathematical truth that *ipso facto* transcends the utmost capacities of human epistemic grasp or a scaled-down conception whereby nothing counts as a truth-apt mathematical statement unless it lies within the compass of our knowledge or available proof-procedures (Benacerraf 1952). Yet this is no answer to Wigner's problem except in the scientifically and philosophically disreputable sense of treating that problem as one best shelved for want of any ready solution. What modal realism seeks to provide is an answer which respects the distinctive kinds of knowledge that pertain in the formal and the physical sciences and which also takes account of their distinctive relationship to issues of objectivity and truth. To this extent it offers a welcome alternative to the kinds of blanket anti-realist doctrine that have largely dictated the agenda of recent epistemological debate.

IV

Such arguments need to be worked out in detail with respect to those specific areas of discourse – from the formal sciences (such as logic and mathematics) to the various natural-scientific disciplines – where a realist approach will necessarily involve different kinds of ontological commitment. That is to say, it will require a good deal of specific fine-tuning as regards the existence of objective truth-values and the issue as to how this claim can be squared with the possibility of our acquiring knowledge concerning them. No doubt there are deep philosophical problems here, especially – as sceptics are quick to point out – in the paradigm case of mathematics where there might seem to be a flat choice between objective or recognition-transcendent truth and knowledge as a matter of provability by the best methods to hand (Hart [ed.] 1996).



All the same the sceptic will be hard put to argue against all the evidence to date that we should take a purely nominalist, instrumentalist, or fictionalist view of mathematical statements and treat their role in the development of physical theories as a kind of lucky fluke. More plausible – as some would hold – is the case for regarding such 'abstract' entities as numbers, sets, and classes as having to do with our acquired capacity for generalisation from the everyday experience of bringing objects under this or that system of counting or group membership. (For further discussion see Kitcher 1983; Maddy 1990.) Of course it remains for the realist to explain how a conception of this kind – classically adopted by empiricists like J.S. Mill – might be reconciled with the objectivist (e.g., Platonist or Fregean) view of mathematical truth as inherently transcending the limits of human cognitive grasp. Hence, as I have said, the pyrrhic idea that quite simply 'nothing works' in philosophy of mathematics since one can *either* have a plausible epistemic account on which 'truth' lies within the scope of human knowability *or* an alethic (objectivist) account on which – as it seems – knowledge must forever fall short of objective truth.

However it remains for the anti-realist to offer some convincing account of how one can adopt the view that 'numbers don't really exist' while assenting to the proposition that 'there are two prime numbers between 11and 19'. Or again, they will have a problem in making the case that all statements about elementary particles should be viewed as nothing more than useful (instrumentally efficacious) fictions while none the less declaring with the utmost confidence that 'the charge on every electron is negative'. What is clear – despite these philosophic qualms – is that one cannot make sense of the history of the physical sciences to date except on the assumption that mathematics has played a chief role in that history and hence that there must be some intrinsic (however elusive or conceptually recalcitrant) relation between mathematical truths and truths about the physical world. Wigner gives voice to the widespread sense of bemusement in this regard when he writes that 'the enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious and [. . .] there is no rational explanation for it' (Wigner 1960: 223). However his remark is less than helpful if taken – as sceptics would readily take it – to entail that no such explanation could ever in principle be had. After all, one need not be any kind of Pythagorean mystic or subscriber to Hegel's idealist doctrine that 'the real is the rational' in order to think that mathematics must have some explicable purchase on those various physical phenomena that it is able to describe, predict, or explain with such extraordinary power and precision. In my view the Kripke-Putnam approach via modal logic and the causal theory of reference offers a means of laying such sceptical doubts to rest by meeting them point-for-point across the range of current anti-realist challenges.

References

Aronson, J., R. Harré and E. Way (1994). *Realism Rescued: how scientific progress is possible.* London: Duckworth.

Bohr, Niels (1934). *Atomic Physics and the Description of Nature*. Cambridge: Cambridge University Press. (1958). *Atomic Physics and Human Knowledge*. New York: Wiley.

Bradley, Raymond and Norman Swartz (1979). *Possible Worlds: an introduction to modal logic and its philosophy*. Oxford: Blackwell.

Burge, Tyler (1979). 'Individualism and the Mental'. In P.A. French, T.E. Uehling and H. Wettstein (eds.), *Midwest Studies in Philosophy*, Vol. 10: 73-121.

Devitt, Michael (1986). *Realism and Truth*, 2nd edn. Oxford: Blackwell.

Dummett, Michael (1978). *Truth and Other Enigmas*. London: Duckworth.

Evans, Gareth (1982). *The Varieties of Reference*, ed. John McDowell. Oxford: Clarendon.

Frege, Gottlob (1952). 'On Sense and Reference'. In Max Black and P.T. Geach (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell: 56-78.

Gibbins, Peter (1987). *Particles and Paradoxes: the limits of quantum logic*. Cambridge: Cambridge University Press.

Haack, Susan (1974). *Deviant Logic: some philosophical issues*. Cambridge: Cambridge University Press.

Harding, Sandra (1976). *Can Theories be Refuted? Essays on the Duhem-Quine thesis*. Dordrecht: Reidel.

Kitcher, Philip (1983). *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.

Kripke, Saul A. (1981). *Naming and Necessity*. Oxford: Blackwell.

Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions*, 2nd edn. Chicago: University of Chicago Press.

Laudan, Larry (1977). *Progress and Its Problems*. Berkeley and Los Angeles: University of California Press.

Leplin, Jarrett (ed.). (1984). *Scientific Realism*. Berkeley and Los Angeles: University of California Press.

Lewis, David (1979). *Counterfactuals*. Oxford: Blackwell. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Linsky, L (ed.) (1971). *Reference and Modality*. Oxford: Oxford University Press.

Lipton, Peter (1993). *Inference to the Best Explanation*. London: Routledge.

Maddy, Penelope (1990). *Realism in Mathematics*. Oxford: Oxford University Press.

McCulloch, Gregory (1995). *The Mind and Its World*. London: Routledge.

Norris, Christopher (2000). *Quantum Theory and the Flight from Realism: philosophical responses to quantum mechanics*. London: Routledge. (2002a). *Hilary Putnam: realism, reason, and the uses of uncertainty*. Manchester: Manchester University Press. (2002b). *Truth Matters: realism, anti-realism, and response-dependence*. Edinburgh: Edinburgh University Press.

Psillos, Stathis (1999). *Scientific Realism: how science tracks truth*. London: Routledge.

Putnam, Hilary (1975). Mind, Language and Reality (Philosophical Papers, Vol. 2). Cambridge: Cambridge University Press. (1975a). 'Is Semantics Possible?', in Putnam 1975: 139-52. (1975b). 'The Meaning of "Meaning"', in Putnam 1975: 215-71. (1975c). 'Language and Reality', in Putnam 1975: 272-90. (1978). Meaning and the Moral Sciences. London: Routledge & Kegan Paul. (1983). Realism and Reason. (Philosophical Papers, Vol. 3). Cambridge: Cambridge University Press.

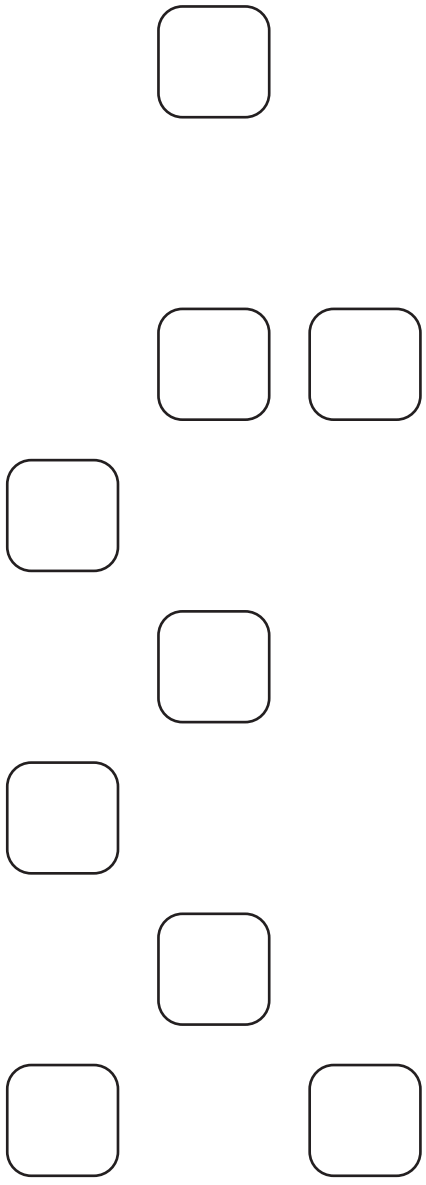
Quine, W.V. (1961a). 'Two Dogmas of Empiricism'. In *From a Logical Point of View*, 2nd edn. Cambridge, Mass.: Harvard University Press: 20-46. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press. (1971a). 'Reference and Modality'. In Linsky (ed.): 17-34. (1971b). 'Quantifiers and Propositional Attitudes'. In Linsky (ed.): 101-111.

Rescher, Nicholas (1979). *Scientific Progress*. Oxford: Blackwell.

Russell, Bertrand (1905). 'On Denoting'. *Mind*, Vol. 14: 479-93.

Schwartz, Stephen (ed.) (1977). *Naming, Necessity, and Natural Kinds*. Ithaca, N.Y.: Cornell University Press.

Wigner, Eugene (1960). 'The Unreasonable Effectiveness of Mathematics in the Physical Sciences'. In *Symmetries and Reflections*. Cambridge, MA: MIT Press: 222-37.



Notes On Contributors

Pierre Cruse

is a research fellow at the Catholic University in Louvain la Neuve. He completed his undergraduate studies in philosophy and physics at the University of Glasgow before going on to do his doctoral work at King's College, London. He has published on scientific realism, which is also the subject of his current research.

James Hill

lectures at Charles University, Prague. He studied at Oxford, Geneva and King's College, London, and has published a range of articles on the empiricists. At the moment he is preparing a book on John Locke's critique of mechanism and working on Descartes' concept of thinking. He is also collaborating on the translation of several English-language philosophers into Czech.

Michael Luntley

is professor of philosophy at the University of Warwick. His main areas of interest lie in perceptual knowledge, the metaphysics of thought and reason, the philosophy of Wittgenstein, and an increasing interest in the philosophy of education. He has published widely, including *Reason, Truth and Self: The Postmodern Reconditioned* (London: Routledge, 1995), and *Contemporary Philosophy of Thought* (Oxford: Basil Blackwell, 1999).

Stephen Grant

is joint editor of the Richmond Journal of Philosophy.

Frank Jackson

is professor of philosophy at the Australian National University. He previously taught in Adelaide, La Trobe and at Monash, before joining the ANU in 1986. His wide range of interests include perception, metaphysics and ethics, but he is best known for his seminal work in the philosophy of mind. His numerous works include *Conditionals* (Oxford: Basil Blackwell, 1987), *From Metaphysics to Ethics* (Oxford: Oxford University Press, 1997), and *Mind, Method, and Conditionals: Selected Essays* (London: Routledge, 1998).

Christopher Norris

is Distinguished Research Professor in Philosophy at the University of Cardiff. He has also taught at the University of Duisburg and the University of Wales Institute of Science and Technology. His wide range of interests includes philosophy of language and mind, metaphysics and metaethics, and issues in epistemology and science. His long list of publications includes *Resources of Realism: Prospects for "Post-Analytic" Philosophy* (London: Macmillan, 1997) and *Against Relativism: Philosophy of Science, Deconstruction and Critical Theory* (Oxford: Basil Blackwell, 1997).

Notes For Contributors

Content

We welcome articles on any area in philosophy. Papers may be broad or narrow in their focus (for instance a discussion of the mind/body problem, or an analysis of Hume's treatment of causation in the *Enquiry*). We would particularly encourage contributions which reflect original research on the following philosophical themes: epistemology, metaphysics, philosophy of religion, ethics, philosophy of mind, philosophy of science, political philosophy, religious ethics; and texts, such as: *The Republic, The Nicomachean Ethics, The Meditations, An Enquiry Concerning Human Understanding, Beyond Good and Evil, On Liberty, Existentialism and Humanism, The Problems of Philosophy, Language Truth and Logic*.

The articles should be around 3000-4000 words.

Style

The language used in the articles should be as non-technical as possible whilst preserving the richness of the arguments. Where technical terms are unavoidable they should be explained and examples offered.

Notes should be presented as endnotes. Textual references should be presented in the following format: Barry Stroud, *Hume* (London: Routledge, 1977), 77-91.

Presentation

Articles should be written in *Word* (any version).

Contributions

Articles for this journal are currently written by a panel of philosophers from a variety of universities in Britain, Australia and the United States, whose work is edited by the journal's editorial board. We invite unsolicited contributions from philosophers working in any field. The contributions should be submitted via email attachment to rjp@rutc.ac.uk

Copyright

Contributors

Richmond upon Thames College

is a large further education college located in
Twickenham, West London
offering 16-19 students one of the widest
curriculum choices in the country.

Last year, we came top of all
London further education colleges
in the Times league tables
and we are proud of our reputation
for achieving excellent results year after year.

We are well known nationally for
our high quality staff, excellent student support systems
and the inclusive education we offer to all our students.

If you would like to find out more about us -
please contact our Course Information Unit on

020 8607 8305 / 8314

or visit our website on

www.rutc.ac.uk

Richmond upon Thames College

[Subscribing]

to the *RJP*

The RJP comes out in Autumn, Spring and Summer. To subscribe you need to select the appropriate price from the table and complete the mailing information at the bottom of the form. All prices include post and packaging. The bottom half of the form should then be detached and sent with a cheque made payable to Richmond upon Thames College to the address below:

RJP Subscriptions
Philosophy Department
Richmond upon Thames College
Egerton Road
Twickenham
London TW2 7SJ
United Kingdom

Please allow one week for delivery in the UK, and two weeks for the rest of the world.

Annual Subscription : Current Rates

Institutional Subscriber in the UK	£33.00
Individual Subscriber in the UK	£18.00
Institutional Subscriber from the rest of the EU	63.00 Euros
Individual Subscriber from the rest of the EU	40.50 Euros
Institutional Subscriber outside the EU	$67.50 US
Individual Subscriber outside the EU	$45.00 US

Your name: ..... Your organisation's name (if appropriate) .....

Address .....

..... Post code/Zip code.....

Telephone number (including full international dialling code) .....

Email address .....

I enclose a cheque for the amount of .....to purchase an annual subscription

Signed ..... Date .....

[subscription]

[Philosophy]

The Richmond Journal of **Philosophy**

Richmond Journal of Philosophy
Philosophy Department
Richmond upon Thames College
Egerton Road
Twickenham
Middlesex
TW2 7SJ
United Kingdom

Tel: 020 8607 8000
Fax: 020 8744 9738
Email: rjp@rutc.ac.uk
www.rutc.ac.uk/rjp

Richmond upon Thames College

ISSN: 1477-6480

Design & Production Marketing for Education 01282 612222 [ref 24370 - 02/03]