

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Frontiers of Philosophy in China**.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/75698/>

Published paper:

Steward, HC (2011) *Agency, Causality and Properties*. *Frontiers of Philosophy in China*, 6 (3). 390 – 401.

<http://dx.doi.org/10.1007/s11466-011-0146-2>

Agency, Causality and Properties

In this paper, I want to try first to highlight, and then to question, an assumption which I think is often made by philosophers writing on topics which concern causation – of which of course, agency is one of the most important. The assumption is that in so far as a substance may be said to be the cause of any phenomenon, a better, fuller, and metaphysically more accurate view of the situation can always be had by looking to the *properties* in virtue of which the substance was able to cause the effect in question. I shall proceed by taking a look at a couple of different examples of places in the literature where it seems to me that this assumption has led philosophers astray. The first example concerns a case of causation in the inanimate world; the second, a case where the type of causation concerned is human agency. In both cases, I shall allege, an important mistake has been committed, and I shall try to diagnose its source. But in the first case, there are reasons for thinking that comparatively little damage may be done when the mistake is made. In the second, though, I shall argue, where the causation in question is agency, the mistake is disastrous – and I shall suggest that it leads both to a radical misunderstanding of the nature of agency, and ultimately to an unfair assessment of the prospects for a coherent version of libertarianism.

(i) *Humphreys*

In his (1989), Paul Humphreys denies that physical objects can really be causes. He accepts that such objects are attributed causal efficacy in ordinary talk, and offers as an example of such an ordinary implication the sentence ‘The car demolished the wall’. Here, it seems, if we take the English sentence at face value, the car itself is attributed causal efficacy vis-à-vis the effect which was the wall’s falling down. But Humphreys denies that what is said here can literally be correct. According to Humphreys, when cases such as this are examined more closely, it always turns out really to be some *aspect* of the object that was the real cause of the effect in question – for example, in this case, the high momentum of the car. Humphreys alleges that it is clear that it was not the car *per se* that demolished the wall, since a car parked touching the wall has no effect; and moreover any object of a similar size moving at similar velocity would have done as well as the car. We should infer, then, according to Humphreys, that it is really the car’s high momentum which was the cause of the wall’s falling down – and that the car itself ought not really to be said to be the cause.

Before going on to look more closely at Humphreys’ reasoning here, I want to pause to point out, for a moment, how bizarre the conclusion of that reasoning is. A very simple reason for finding it odd is that verbs like ‘demolish’ – and many other sorts of verbs by means of which we appear to attribute effects to physical objects – are often given a causal analysis by linguists. ‘Demolish’, for instance, one might think, simply *means* something like ‘cause to fall down’ – and this is indeed how it, and many other similar verbs (e.g. ‘uproot’, ‘unravel’, ‘raise’, ‘dissolve’, etc.) *are* generally analysed in linguistics. And given this premise, if the car demolished the wall it appears simply to *follow* that the car caused the wall to fall down. It seems that if we are to concur with Humphreys, our only options are (i) to deny the literal truth of the sentence ‘the car demolished the wall’; or (ii) to reject the causal analysis of ‘demolish’. But surely neither of these alternatives is very happy. The first involves finding a perfectly ordinary sentence of English to be false in circumstances in which we would all

ordinarily agree it was true; and the second involves denying a well-established account of the workings of a certain large class of English verbs. It seems to me that it would be preferable, then, before turning to adopt either of these solutions, to take a closer look at Humphreys' reasoning.

Humphreys claims, recall, that we are prevented from acceding to the claim that it was the car itself that demolished the wall both by the fact that if the car had merely been parked touching the wall, nothing would have happened to the wall at all, and likewise, by the fact that any object with a similar amount of momentum would have demolished the wall just as well as the car. But – to take the first of these inferences first - why should it be thought to follow from the fact that the car wouldn't have demolished the wall if it hadn't been driven into it at high speed, that the car wasn't really the cause of the wall's being demolished? We cannot *usually* infer from the fact that A would not have V-ed if such and such had not been the case, that A did not V. For example, we can't infer from the fact that I wouldn't have read this book if it hadn't looked interesting that I didn't read this book. And the second inference looks no better, on the face of it. We cannot usually infer from the fact that another thing, B could have V-ed just as well as A did, that A did not really V. Why here, in the case of these specifically causal contexts, should it be thought to follow?

The reason it has been thought to do so, I think – and not only by Humphreys – the line of reasoning he sketches is quite commonly found in the philosophical literature on causation - stems ultimately from the attempt to apply certain methodological principles to the example under consideration which are sometimes offered as guides to the discernment of true causal relations. In particular, it seems as though Humphreys is here attempting to make use of Mill's 'Method of Difference' and 'Method of Agreement' in order to identify what was the true cause of the wall's falling down in the case he imagines. Mill's 'Method of Difference', as applied to causes, states that 'If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the ... cause, or an indispensable part of the cause, of the phenomenon'. (Mill 1970: 256). This might perhaps seem to license the claim made by Humphreys, that since a parked car has no effect on the wall, it must be the car's momentum rather than the car per se, which was the cause of the wall's being demolished. And Mill's 'Method of Agreement', as applied to causes, states that 'If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause ... of the given phenomenon' (Mill 1970: 255). This might seem to license the claim made by Humphreys that since anything with similarly high momentum – but not e.g. anything of a similar colour, or made from a similar material or having the same shape – would have demolished the wall, the high momentum must be the cause of the wall's falling down. But can these principles really be applied in the way in which Humphreys hopes to apply them to justify a preference for a property over a particular entity, when it comes to discerning what is the real cause of some phenomenon or other?

Mill's discussion of the 'Method of Difference' and the 'Method of Agreement' is part of his discussion of laws of nature. The question he is considering is how to single out 'from among the circumstances which precede ... a phenomenon those with which it is really connected by an invariable law' (Mill 1970: 253). He gives as a rather gruesome example of use of the Method of Difference taking a bird from a cage and plunging it into carbonic acid gas (Mill 1970: 257). The bird was alive before the immersion in the gas; after immersion, it suffocates. It must, then, be the immersion in the gas, Mill suggests, which has caused the suffocation – since it is the one circumstance that has changed in the condition of the bird. All

the other things, we may suppose – temperature, pressure, etc., are just as they were before – or at any rate, we will try to ensure that this is so in performing our experiment. It is none of these things, therefore, but the thing which is different, the sudden presence of carbonic acid in the bird's environment, which was the cause of the bird's death. And as an example of the use of the Method of Agreement, he offers the discovery by Baron Liebig of the common factor present in deaths produced by metallic poisoning by means of compounds as various as arsenious acid and the salts of lead, bismuth, copper and mercury; all were discovered by Liebig to combine with animal products such as milk, muscle fibre and animal membranes in such a way as to inhibit further decomposition of these materials – decomposition on which the continuation of life in any organism turns out partly to depend (Mill 1970: 267). Thus, despite the apparent difference in the poisons involved, all were discovered by Liebig actually to have a common mode of operation which involves the chemical's preventing a type of functioning crucial to the well-being of organic tissues from occurring as it ought. It seems obvious that Mill is here highlighting principles of scientific investigation which are widely recognised to be amongst the most important marks of the experimental method, and about the general validity of which, various minor qualifications and points of detail apart, there can be little doubt. Is Humphreys doing no more really, then, than applying what look to be sensible principles of empirical enquiry which have proved their worth many times over, to the particular case of the car and its properties?

When considering the question how generally the Methods of Difference and Agreement may be applied, I think it ought to be borne in mind that for Mill, both methods are for utilisation in connection with the discernment of *laws of nature*. His interest is in the means of discovering which phenomena are 'invariably connected' with others – which of the 'circumstances' which precede a phenomenon those which are connected with that (type of) phenomenon by a true law. But this context, it seems to me, is important to the understanding of how those principles are supposed to be applied. What sorts of things will these preceding 'circumstances' be? I suggest that it will be best to think of them propositionally - essentially, the candidate causes should be thought of as *facts* to the effect that a certain general property or feature was present on the particular occasion in question. The element of generality is essential – for otherwise the idea of invariability makes no sense – nothing which is not at least potentially multiply instantiable can be 'invariably' connected with something else. It is true that it will not be the general thing itself, but rather the fact that it was instantiated on some particular occasion, which will be singled out as the cause of any given effect – e.g. 'the fact that the wall was impacted by an object having such-and-such a degree of momentum' will be the cause of its falling down. Or again: it may be a law that birds in general (and perhaps even animal life forms in general) die on exposure to a sufficient quantity of carbonic acid gas; and it may likewise be a law that no animal life can survive the type of interference with the natural processes of putrefaction that contact with a sufficient quantity of metallic salts tends to produce – and knowing such things as this may help one divine on a given individual occasion that *this* bird has died because it was plunged into a particular mass of carbonic acid gas, and *this* person has died because their tissues have responded with the usual reaction to the ingestion of some particular quantity of metallic salt. But the crucial thing is that all potential contenders for the prize of true cause in the individual case, for Mill, will be in the same ontological category as one another – they will all alike be facts about the instantiation of certain general properties, features or aspects of situations by certain particular entities. There is no question, for Mill, of applying the Methods of Difference and Agreement so as to adjudicate the putative competing causal claims of particulars, on the one hand, and their properties, on the other, to argue for *metaphysical* conclusions about which sorts of ontological entity are best suited to be

accounted causes. Since both are methods which are concerned essentially with the discernment of laws, they are designed from the first to operate within a domain in which *all* the possible candidate causes are instantiations of general features or properties. They are not suitable for assessing the pretensions to causal status of such things as the particular entities which instantiate these features or properties themselves. The methods are, moreover, principles of empirical and not of ontological enquiry – they are for use by scientists in the attempt to discover which types of general phenomena are causally related to which others, not for use by philosophers, attempting to regiment the notion of causation. It seems to me, then, that the use that Humphreys has attempted to make of the Methods of Difference and Agreement is a misapplication of an empirical method in the service of metaphysical ends.

Confusion here is encouraged, I think, by the fact that *causation* and *explanation* are closely related concepts. This means that in our assessment of what is ‘the real cause’ in instances of causation, there are often methodological principles at work which sustain an understandable preference, in certain sorts of case, for causal explanations which mention aspects or properties of substances over those which mention only the substances themselves. It is certainly not hopeless, when confronted with the question, say, why someone who has been poisoned has died, to say that a given quantity of arsenic was the cause – that is a causal explanation of sorts. But if we can say, in addition, that it was because arsenic forms salts of a type which react with organic matter in such a way that natural processes of decomposition are disrupted, we do better – we locate the property of the arsenic in virtue of which it acts as a poison on the human body. But we must be careful here not to mix up the metaphysics of causation with the heuristics of explanation. It may be right to say that if we want an explanation of why the person died, we are generally likely to be more satisfied if we have some account in general terms of the properties in virtue of which arsenic acts as a poison. But first, it is not completely obvious that we should *always* prefer an explanation in such general terms over a mention of the particular cause. Sterling Lamprecht, a great champion of the claims of particular causes, writes as follows, in criticism of the view of Schlick that causation is nothing but ‘regularity of sequence’:

Sitting in my home at night, I may hear a knock at the door. Someone might chance to tell me that, always and invariably, according to a law of nature, sound results from the reverberation of a solid block of wood which is disturbed by blows upon it, and that this explanation is the full and entire causal account of the knocking. Would I be satisfied? Would you? Would even Moritz Schlick? The laws of nature in terms of which a particular knocking might be adequately described would probably not be of much interest to you or me or Schlick at that moment. I am sure that I should want to know who or what concretely was making blows on the door. I should want to identify the specific agent who or which did the knocking ... That there would be uniformity of result in similar cases would be of no account to me at the moment. (Lamprecht 1967: 121-2).

And second, even if it is usually true that explanations in terms of general properties are preferable to those which cite only particular, substantial causes, it simply does not follow from this, as is often assumed, that it is the general properties that ‘do the causal work’. Certainly it follows that properties, or facts involving them, can be used to do important *explanatory* work in accounting for why something occurred. But might it not be a category mistake to suppose that properties do *causal* work? Jonathan Lowe has argued recently that events cannot do causal work, because they do not have causal powers, and the same might, I think, be plausibly said of properties. Poisonousness, for example, seems to be a property, but

poisonousness is surely not naturally thought of as something which *has* powers; rather, it seems itself to be a power. What has the causal power to produce changes in other entities is surely the poisonous thing itself – the arsenic or the bismuth or the mercury.

Lowe himself goes further than this; in addition to claiming that substances are the only proper bearers of causal power, he claims also that it is substances which are the relata proper of the causal relation. To say this, I think, might be to overreact to the worry about causal power as Lowe himself seems to recognise in some places.¹ For even if events and properties do not have causal powers, it need not follow that they cannot be causes - for perhaps there are kinds of causes, the kind that events, properties and facts are, which cause things in ways other than by exerting causal powers. Perhaps, for instance, as I have argued elsewhere (Steward 1997) there are causes which trigger other events by *occurring*, and to occur need not be to exercise any kind of causal power. Or perhaps there are causes which *matter*, which are *relevant*, to the occurrence (or non-occurrence) of particular effects without being efficacious. But still, Lowe is surely right to question whether we can move smoothly from the observation that it is often more appropriate and more helpful to cite an event or a property as the cause of something than it is to cite a substance, to the conclusion that it is never really the substance which ‘does the causal work’, that the substance simply drops out of the causal metaphysics, giving way gracefully to the items which are offered in what are judged to be the fuller and more detailed explanations of what has happened.

My own view, in fact, is that the question what the relata of the causal relation are, though often posed, is simply not a sensible one. It is not true either that events are the relata of the relation, nor that substances are – nor indeed that any other of the multifarious candidates that have been proposed (e.g. facts, processes, properties, tropes, etc.) are its relata – because there simply is no single relation that is *the* causal relation – nor is causation always and everywhere a relation at all. Causation is best thought of as a *category* – a large and ontologically flexible umbrella concept under which we bring a wide diversity of ontologically various relations and relationships, unified only by their connections to our interest in the explanation of how things have come to be and why they have (or have not) changed. We need to ask why it is obligatory to suppose that causation *has* an ontology in the first place. After all (as is frequently observed), items in almost all the ontological categories it is possible to think of are *spoken* of as causes – objects, persons, events, facts, states, properties, and so on. Why think that any of these ways of speaking can be sensibly thought of as revelatory of *the* ontology of causation? Might not the simple truth be that we need a plurality of irreducibly distinct ontological categories to do justice to the totality of causal phenomena?

If this is indeed the simple truth, the way should be clear to embrace a pluralism about causes which will enable us easily to avoid Humphrey’s unpalatable and absurd conclusion that cars cannot demolish walls. Cars can demolish walls all right. When they do so, it is perfectly true, of course, that some of their properties are relevant to the fact that they have this demolishing ability, and some are not. But that does not mean that these properties demolish the walls – nor indeed that any horrid surrogate entities – such as tropes or property instances or exemplifications of properties at times – do so. Demolition is simply not a job

¹ “I speak of ‘causal relations’ in the plural here advisedly, because I think it is tendentious to assume that there is such a thing as ‘the’ causal relation, although this assumption is very widespread amongst contemporary analytical metaphysicians ... some causal statements undoubtedly have an event-causal formulation ... However, we also have what might be called ‘mixed’ causal statements, such as ‘The bomb caused the collapse of the bridge’ in which the grammatical subject of the verb ‘to cause’ is a noun-phrase denoting a particular persisting object or individual substance ... (Lowe, 2008: 142).

for a property. It is the sort of job that only something like a car – with mass, momentum and volume – could do.

I want to turn now to look at a second kind of argument in the literature where I think a similar kind of reluctance to credit substances with powers they really have and a preference for ceding those powers to their properties can be found. This time, though, the mistake occurs in connection with the explanation of human action and the related mistake may be, I think, both more important, and harder to see.

(ii) *Mele's Argument*

An example of the second kind of argument is offered by Al Mele in his book, *Free Will and Luck*. Mele formulates what he calls 'a problem about luck for libertarians' as follows. He notes that the typical libertarian believes that a free decision to A, made by a given agent, at a particular time *t*, could, at that very moment, have gone the other way – the agent could have decided at *t* not to A, instead. In the actual world, this agent – following Mele, I shall call him Joe – decides at *t* to A. But in another world *with the very same laws of nature and the very same past*, the libertarian believes, Joe decides at *t* not to A. But in that case, Mele argues, the libertarian faces the following difficulty:

If there is nothing about Joe's powers, capacities, states of mind, moral character, and the like in either world that accounts for this difference, then the difference seems to be just a matter of luck. And given that neither world diverges from the other in any respect before *t*, there is no difference at all in Joe in these two worlds to account for the difference in his decisions... the difference in his decisions ... [is] just a matter of luck (Mele 2006: 9).

And if the difference in his decisions in these two possible worlds is just a matter of luck, Mele goes on to ask, how can it have been in any sense *up to Joe* which decision was made? How can it have been up to him which possible world became actual? And unless we can see that which decision was made was up to Joe, unless which possible world became actual was up to him, how on earth does the indeterminacy which has been posited by the libertarian contribute to his freedom and moral responsibility?

What I am interested in here is not the argument offered by Mele taken as a whole, but rather a particular aspect of it – that is, the link which Mele attempts to make between there being nothing *about* Joe's powers, capacities, states of mind, moral character, and the like, to account for the difference between the situation in which he decides at *t* to A and the situation in which he decides at *t* not to A, and its being a matter of luck that Joe does what he does. Clearly, more really needs to be said, in order to tackle this inference, about what it is for something or other to be a 'matter of luck', but a full discussion of the concept of luck would be a complex matter which would take me well beyond the bounds of this paper. Let me try to cut the work short, then, by observing that what Mele seems to need in order for his argument to go through, is for it to *follow* from its being a matter of luck whether Joe decides at *t* to A or decides at *t* not to A that it not be *up to Joe* what in fact gets decided. Rather than discuss the concept of luck in more detail, then, let me just home straight in on this implication of the understanding of it that Mele seems to need. What he appears to want to insist is that if there is nothing *about* Joe which could account for the difference between the situation in which he decides at *t* to A and the situation in which he decides at *t* not to A, it could not be *up to Joe* which thing happens. And that is the inference I want to question.

The inference, I take it, is from there being no *property* of Joe's which one might point to in order to explain why he decides in one way rather than another, to its not being up to Joe whether he does the one thing or the other, to Joe's not really being the source, the true agent of what happens. But why should we accept that any such inference is in order? Let me distinguish between two kinds of reason one might have for believing that such an inference might be valid. One reason might have to do specifically with thoughts about rationality – it might be said, for instance, that in order for a decision genuinely to count as the true action of a human agent, it must be made for a reason. And in that case, it might be said, the agent's appreciation of the preponderance of the weight of that reason, or its preponderance over the reasons which argue for the alternative course of action can then be cited in explanation of what she does, must be part of the explanation of why she decides as she does. Relevant properties of the agent – properties concerning relative strengths and weights of beliefs and desires, for instance - will therefore always be available to explain why she does what she does, provided we have an instance of true agency at all. I do not think the premise of this line of thought – that actions must be done for reasons - is true, nor do I think the inference it involves is valid, but I want to set it aside for the purposes of this paper, because it requires a discussion of the relation between agency and rationality which would take me much too far afield. What I want to focus on is another kind of reason I think one might have for believing that the inference from 'there is no property of the agent which could explain why one thing happened rather than another' to 'it was not up to the agent what happened' might be valid. This second reason, I think, is belief in a principle which might be stated in something like the following way: when a substantial thing does anything or acts in any way, what it does must always be something which is explicable in terms of certain properties which it antecedently possessed – and more specifically, if it does one thing rather than another, then that must be explicable in terms of its antecedently possessing certain properties rather than certain others. If this were true, Mele would, I think, be within his rights to infer that since there is, by hypothesis no such explanation in the case of Joe, that he cannot really have been the agent of the inexplicable occurrence which is therefore called his 'decision' only by courtesy. But I want now to argue that the principle presupposes determinism, and so that the libertarian should have no qualms about its rejection.

The easiest way to see that the principle presupposes determinism, I think, is to begin by thinking about the inanimate entities which might be involved in indeterministic occurrences. For instance, suppose a radioactive atom indeterministically ejects a particle at t . Then the radioactive atom might be said to have brought about certain consequences in the universe (that a particle was ejected at t rather than not, for example, together with any possible further consequences of this ejection) without any of the prior properties of the radioactive atom having been relevant to the fact that the question whether or not a particle would be ejected at t was settled with an ejection rather than a non-ejection. It might be retorted that the atom might at least possess properties which made it probable that it would eject a particle at t , rather than not, and that these could be cited in the explanation. But of course this need not be so. We can, indeed, increase the plausibility of the claim that the atom's antecedently existing properties were irrelevant to the explanation of the fact that it ejected a particle at t rather than not, by imagining that any objective probabilities existing and relating to the ejection were such that they made an ejection at t exceedingly *unlikely*. For in that case, it would not even be plausible that we could appeal to the thought that the radioactive atom (and its environment) somehow embodied these objective probabilities to ground the supposition that certain of its properties must have been explanatorily relevant to what happened. There seems, then, in a case like this, to be nothing we can say *about* the radioactive atom which gives us any help with explaining why the particle was ejected at t

rather than not. And yet the atom ejected a particle all right. *It* brought about consequences in the world by ejecting a particle, even though none of its properties was explanatorily relevant to the question why the particle was ejected at *t* rather than not.

We have here, then, a simple case in which an object does something – in this case, ejects something – in such a way that no property of that object can be cited in explanation of why it did that thing rather than not. Now, of course, I am not suggesting that human actions are events which should be thought of as random in the same way as radioactive emissions might be supposed to be – precisely not. A radioactive atom is certainly not the kind of thing that something could be ‘up to’ because it has no control over its own parts; it is not a possessor or controller of a body in the way that an animal or human being can be. The point of the example is not to suggest that there is no significant difference between random radioactive emissions and full-blown actions, but rather merely to cast doubt on the principle that an enduring entity’s role in the unfolding of the world must always be exhausted by the relevance of its properties to explanations of why the world in fact unfolded in the way that it did (rather than in some other way) – to cast doubt on the inference I am imagining we might be being invited to make by Mele from ‘there is no difference at all in Joe in these two worlds to account for the difference’ to ‘there is no role *for Joe* here in the unfolding of reality’. If there are to be true actions, I should like to maintain, we need for it sometimes not to be the case that the way things unfold depends only on the way they already are – even if some of those ways things are, are ways things are *with us*. Action, so it seems to me, introduces into the world another kind of dependence entirely from the kind which is exploited by explanations of how things turn out in terms of how things antecedently were – dependence on an agent, as opposed merely to dependence on the way the agent is. Showing that this is so is of course a difficult endeavour with many parts and I do not intend to attempt to show it here. But a necessary part of the project is showing that it does not follow from the fact that none of an entity’s *properties* is relevant to the explanation of why a particular sort of event happened at a given time rather than not, that the entity itself did not produce that event. The libertarian should insist that it is not because something *about us* makes us act, or explains why we act, but simply because *we act*, that it is up to us what happens to our bodies. None of this implies, of course, that our properties are never causally relevant to the question why we do what we do – that would be an absurd conclusion. My claim is only the much milder one that when it comes to the sort of contrastive explanations of why an agent has done a particular thing at a time rather than doing some other thing, or rather than merely not doing that thing, there need be nothing at all to say of a sort that relates to properties the agent antecedently possessed in the one case but not in the other. And if there is indeed nothing at all to say, that need not prevent its being the case that what occurred was up to the agent.

References

- Humphreys, P. (1989), *The Chances of Explanation* (Princeton: Princeton University Press).
- Lamprecht, Sterling P. (1967), *The Metaphysics of Naturalism* (New York: Appleton-Century-Crofts).
- Lowe, E.J. (2008), *Personal Agency* (Oxford: Oxford University Press).
- Mele, Alfred (2006), *Free Will and Luck* (Oxford: Oxford University Press).
- Mill, J.S. (1970) [1872], *A System of Logic*, 8th edition (London: Longman).