

# The Philosophical Quarterly

Volume 57 • Number 228 • July 2007

## CONTENTS

### ARTICLES

---

The Objective Attitude	Tamler Sommers	321
Propositional Clothing and Belief	Neil Sinclair	342
Unified Semantics of Singular Terms	John Justice	363
The Wrong of Rape	David Archard	374
Gettier and Factivity in Indo-Tibetan Epistemology	Jonathan Stoltz	394
Democracy and Equality	Steven Wall	416
The Impossibility of Backwards Causation	Hanoch Ben-Yami	439

### DISCUSSIONS

---

Two Kinds of Holism about Values	Campbell Brown	456
The Importance of Frankfurt-Style Argument	John Martin Fischer	464

### CRITICAL STUDY

---

Insensitive Semantics	Mat Carmody	472
-----------------------	-------------	-----

---

BOOK REVIEWS		479
--------------	--	-----

## THE IMPOSSIBILITY OF BACKWARDS CAUSATION

BY HANOCH BEN-YAMI

*Dummett and others have failed to show that an effect can precede its cause. Dummett claimed that 'backwards causation' is unproblematic in agentless worlds, and tried to show under what conditions it is rational to believe that even backwards agent-causation occurs. Relying on considerations originating in discussions of special relativity, I show that the latter conditions actually support the view that backwards agent-causation is impossible. I next show that in Dummett's agentless worlds explanation does not necessitate backwards causation. I then show why even relative backwards causation is impossible in his and Tooley's scenarios of parallel processes in which causes apparently act in opposite temporal directions. We thus have good reasons for thinking that backwards causation is impossible.*

In 1954, Michael Dummett published a paper in which he argued that some sort of cause – ‘quasi-cause’ was his term – may occur later than its effect.<sup>1</sup> This claim, which most philosophers find, at least initially, quite puzzling, was criticized in the same volume of the Aristotelian Society's *Proceedings* by Antony Flew.<sup>2</sup> One of Flew's criticisms, subsequently dubbed ‘the bilking argument’ after the term and metaphor he used (p. 57), was elaborated two years later in a paper by Max Black.<sup>3</sup> This bilking argument (to be presented below) seemed to many to supply a refutation of Dummett's claim.

Yet Dummett, undeterred, defended his original position in his celebrated 1964 paper ‘Bringing About the Past’.<sup>4</sup> In that paper he replied to the bilking argument (without mentioning either Flew or Black), by elaborating the conditions on an event's being a cause of a preceding event – this time ascribing full-blown causality, *backwards* causality, and not merely some sort of *quasi*-causality, to the later event.

Dummett's later paper again drew some criticisms, probably the best known being Hugh Mellor's. But Mellor's criticism was not widely accepted,

<sup>1</sup> M. Dummett, ‘Can an Effect Precede its Cause?’, *Proceedings of the Aristotelian Society*, Supp. Vol. 28 (1954), pp. 27–44, repr. in his *Truth and Other Enigmas* (London: Duckworth, 1978), pp. 319–32.

<sup>2</sup> A.N. Flew, ‘Can an Effect Precede its Cause?’, *Proceedings of the Aristotelian Society*, Supp. Vol. 28 (1954), pp. 45–62.

<sup>3</sup> M. Black, ‘Why Cannot an Effect Precede its Cause?’, *Analysis*, 16 (1956), pp. 49–58.

<sup>4</sup> Dummett, ‘Bringing About the Past’, *Philosophical Review*, 73 (1964), pp. 338–59.

and was, I think, satisfactorily rebutted by Dummett himself and by Peter Riggs.<sup>5</sup> At present, some philosophers not only think that Dummett has shown backwards causation to be possible, but also attempt to make use of it in explaining several puzzling phenomena in quantum physics.<sup>6</sup>

I suspect these attempts are misguided. Backwards causation still seems to me (as it does to many other philosophers) incoherent. In fact, Dummett's claim, that backwards causation is possible, sounds like one of those 'philosophical discoveries' Wittgenstein warned us against, discoveries which are actually castles in the air (*Luftgebäude*, *PI* §118; 'houses of cards' in Anscombe's translation). But calling names is not enough; one should also clear the ground, which might be a difficult task. This is the task of this paper.

Dummett repeated the claims and arguments of his 1964 paper in a few later publications.<sup>7</sup> He himself, however, continued to regard 'Bringing About the Past' as an authoritative presentation of his position. I shall do the same here, and take this paper as my main target. All page references to Dummett's work below are to this paper.

Dummett thinks that if we limit ourselves to worlds containing no 'creature to whom can be ascribed intention and purpose' (p. 339), then 'we can conceive of a world in which a notion of causality associated with the [later-to-earlier temporal] direction would have been more appropriate and, so long as we consider ourselves as mere observers of such a world, there is no particular conceptual difficulty about the conception of such a backwards causation' (p. 340). In support of his claim he describes an agentless world in which backwards causation seems to him the more appropriate kind of cause. I think the world Dummett describes does not support his claim, as I shall try to show later in this paper. Yet Dummett, being convinced of its conceivability, proceeds to deal with something which does appear to 'generate absurdities' (*merely* appears to do so, as he will try to show), namely, backwards causation that governs events in which we, as agents, can intervene (p. 340).

Dummett supplies (pp. 344–8) an elaborate argument against a fatalistic-like argument for the impossibility of backwards causation involving agents.

<sup>5</sup> See D.H. Mellor, *Real Time* (Cambridge UP, 1981), and 'Fixed Past, Unfixed Future', in B.M. Taylor (ed.), *Michael Dummett: Contributions to Philosophy* (Dordrecht: Nijhoff, 1987), pp. 166–86. See also S. Gorovitz, 'Leaving the Past Alone', *Philosophical Review*, 73 (1964), pp. 360–71; Dummett, 'Reply to D.H. Mellor', in Taylor (ed.), *Michael Dummett: Contributions to Philosophy*, pp. 287–98; P.J. Riggs, 'A Critique of Mellor's Argument against "Backwards" Causation', *British Journal for the Philosophy of Science*, 42 (1991), pp. 75–86.

<sup>6</sup> E.g., P. Dowe, 'Backwards Causation and the Direction of Causal Processes', *Mind*, 105 (1996), pp. 227–48; H. Price, *Time's Arrow and Archimedes' Point* (Oxford UP, 1996).

<sup>7</sup> Dummett, 'Causal Loops', in R. Flood and M. Lockwood (eds), *The Nature of Time* (Oxford: Blackwell, 1986), pp.135–69, 'Reply to D.H. Mellor', and very briefly in his *Truth and the Past* (Columbia UP, 2004).

I too think that this fatalistic-like argument is unsound, so I shall not discuss it here.

Dummett then gradually builds up a description of a scenario in which he thinks it would be rational for the agent in the story to believe that his action brings about an earlier event (Dummett does not mean that the action *changes* the past, but that an event is caused by an action occurring later than itself). Dummett wants to show not only that there is no contradiction in the idea of backwards agent-causation in the appropriate circumstances, but also that it would be rational in some circumstances to believe that it has actually occurred. His scenario is therefore based on metaphysical and epistemological considerations, as well as considerations pertaining to rationality, all with regard to the agent in his story.

My purpose here is different: I shall try to show that the idea of backwards agent-causation is incoherent. Consequently, as it is never rational to believe in what you know involves a contradiction, I need to consider only the metaphysical aspects of Dummett's scenario. Since Dummett thinks that backwards causation is problematic when the cause is an *agent's action*, I shall introduce into the scenario characteristics peculiar to agents, even to human or rational agents. But I can ignore questions like whether the acting agent could know that one event caused another, how the agent learns about causal relations, or whether it would be rational for the agent to believe that backwards causation was instantiated in a specific case.

On many occasions, agents are justified in believing that a specific event *a* of type A caused a specific event *b* of type B only if events of type B are connected by some rule to events of type A. (So as not to beg the question, I include no reference to temporal order in the rule.) The agents should also take into consideration the frequency in which Bs come together with As, the reliability of the information available to them, and probably other factors as well. But these are irrelevant if we want to show that the very idea of backwards agent-causation involves a contradiction.

For instance, *pace* Hume, it is doubtful whether the concept of *cause* is reducible to regularity plus several additional conceptual ingredients.<sup>8</sup> Moreover, even if causation presupposes regularity, this is in the sense that if the same causal conditions recur, then the same effect will recur. But the same conditions need not recur. So even if causality presupposes regularity, Dummett's scenario needs to involve the same thing happening over and over again only if we include epistemic questions, but not if we ask the metaphysical one alone. In addition, we also do not need to estimate the reliability of the information available to the agent in the story in order to

<sup>8</sup> Cf. P.M.S. Hacker, 'Causality', in his *Human Nature: the Categorical Framework* (Oxford: Blackwell, forthcoming).

determine whether the idea of backwards agent-causation involves a contradiction: we, who are telling the story, know exactly which events have taken place.

Consequently, unlike Dummett, I shall not be asking whether the agent in the story is justified in believing that a case of backwards causation has occurred in which he is the agent. Rather, I shall be asking whether it is coherent to assume that the agent in the story has caused an event earlier than his action. This will also simplify the intricate scenario Dummett constructs (pp. 348–9), to which I now turn.

Suppose we come across a tribe who have the following custom. Every second year the young men of the tribe are sent, as part of their initiation ritual, on a lion hunt: they have to prove their manhood. They travel for two days, hunt lions for two days, and spend two days on the return journey; observers go with them, and report to the chief upon their return whether the young men acquitted themselves with bravery or not. The people of the tribe believe that various ceremonies, carried out by the chief, influence the weather, the crops, and so forth.... While the young men are away from the village the chief performs ceremonies – dances, let us say – intended to cause the young men to act bravely. We notice that he continues to perform these dances for the whole six days that the party is away, that is to say, for two days during which the events that the dancing is supposed to influence have already taken place.

To simplify things, I shall suppose that the chief dances only once, on the fifth day, in order to cause the young men to act bravely. I shall also concentrate not on the custom in general (as I have said, regularity need not be assumed), but on one specific pair of lion hunt and dance. Suppose the young men acquitted themselves with bravery (event *b*) and the chief did dance (*d*). Is it possible that his dance caused their success?

As Black has argued,<sup>9</sup> for *d* to cause *b* the following two conditions have to hold: (i) *b* has no earlier cause; (ii) there is a sufficient cause for *d*, causally independent of *b*.

But now the bilking argument sets in. What is distinctive of agent-causation is that agents can act on the basis of the information available to them. We are assuming that *b* does not causally necessitate *d*, and that *d* is an action of a person. Accordingly, the chief can decide not to dance if he knows that the young men acquitted themselves with bravery. But then *b* would occur whether or not *d* occurs. So *d* is not *b*'s cause.

I shall first reply to an immediate objection to this bilking argument, an objection which both Black and Dummett recognized as *not* constituting an adequate response. The bilking argument claims that had *d* not occurred, *b* would still have occurred, and that therefore *d* is not *b*'s cause. The

<sup>9</sup> Black, 'Why Cannot an Event Precede its Cause?', p. 54, in general agreement with Dummett's 'Can an Effect Precede its Cause?', pp. 322–3, quoted by Black on pp. 55–6.

argument thus seems to presuppose a counterfactual theory of causation: event *a* is the cause of event *b* only if, had *a* not occurred, *b* too would not have occurred. But despite attracting support from some distinguished representatives,<sup>10</sup> this theory is known to be problematic. Even if *d* is *b*'s cause, if it had not occurred, then some other event could have occurred which would have caused *b*. So the fact that if *d* had not occurred *b* would still have occurred does not show that *d* is not *b*'s cause.

Although this is a good objection to the counterfactual theory of causation, it is not so in my case. I am describing a scenario involving agent-causation because of the peculiarities of this kind of causation. And these include the ability to act on the basis of one's knowledge. (As I mentioned above, I shall consider later in this paper cases in which such causation is not involved.) So in order to remain within this framework, we need to limit all alleged future causes of *b* to such actions as are within the powers of agents to perform or not. Specifically, we need to assume that *any* alleged cause can be prevented if agents know that *b* has occurred. Accordingly, *b* could occur without *any* alleged future cause; we cannot assume that some cause or other would always occur. So no such alleged cause is really *b*'s cause.

(Still, one should be wary of a certain fallacy that often crops up in such reasoning. From the fact that any event of kind *F* might not have occurred, it does not follow that all *F* events might not have occurred. This invalid inference is an instance of what may be called 'the centipede fallacy': from the fact that if any of its legs is cut the centipede can still walk, it does not follow that if all the centipede's legs are cut it can still do so. In the language of formal logic, this is the fallacy involved in substituting  $\diamond\forall$  for  $\forall\diamond$ . I have not committed this fallacy here. Since I assume that any agent can act on the basis of his knowledge, I can assume that the only factor relevant for his action is his knowledge that *b* has occurred, and that independently of others' actions, each agent could refrain from acting if he knew that *b* has occurred. I can therefore assume that no agent might have acted had he known that *b* has occurred, and that therefore no action is *b*'s cause.)

Since, as I have just pointed out, adding agents does not affect my argument, I can for simplicity's sake assume that only one such action takes place: the chief's dance.

To sum up, since the chief could have refrained from doing *d* if he knew that *b* had occurred, and of course in that case *b* would have occurred, *d* is not *b*'s cause. It is apparently impossible to construct a story that contains backwards agent-causation.

<sup>10</sup> E.g., D.K. Lewis, 'Causation', *Journal of Philosophy*, 70 (1973), pp. 556–67.

But why introduce knowledge into the considerations? Is it not enough that we assume the chief to be a free agent, and also assume that his dance is not causally determined by the bravery with which the young men acquitted themselves? Should we not say that whether or not he knew of the young men's bravery, the chief could have refrained from dancing, and that therefore his dancing is not the cause of the young men's bravery?

I am trying to refute backwards agent-causation by means of peculiarities of *agent*-causation. I am not trying to do this by simply assuming indeterminism, which is a possible characteristic of laws, both of those describing agents' behaviour and of those describing inanimate nature. I am, though, assuming the ability to act for reasons, which is peculiar to rational agents. I aim to see whether we can include this ability in the story without making backwards agent-causation impossible. If I relied only on the fact that the past does not determine the future, then no peculiarities of agent-causation would be involved.

In addition, it seems that indeterminism by itself is insufficient to rule out the possibility of backwards causation. Indeterminism is a logical relation between the full description of the world at a certain moment and its full description at later moments: given the laws of nature, the latter does not follow from the former. But given indeterminism, this is generally true in the opposite temporal direction as well: given the laws of nature, the full description of the world at an earlier moment does not follow from its full description at a later moment. So if indeterminism were sufficient to rule out backwards causation, then because of this logical symmetry, it would be sufficient to rule out forwards causation as well. But forwards causation seems possible even given indeterminism; in fact, this seems to be what happens in our world. So it seems that on its own, indeterminism is insufficient to rule out the possibility of backwards causation.

Of course, if free will is incompatible with determinism, then the introduction of free agents presupposes indeterminism. But the compatibilism debate is still raging, and we should not take sides here. Moreover, perhaps acting for reasons is compatible with the absence of free will. Consequently I shall not try to argue against the possibility of backwards agent-causation on the assumption that agent-causation involves indeterminism. I shall, however, assume that it involves the ability to act for reasons.

Dummett acknowledged the difficulty which the bilking argument creates for the possibility of backwards agent-causation. He presents the argument and tries to reply to it on pp. 352–8 of 'Bringing About the Past'. His conclusion is that for backwards agent-causation 'to make sense' (p. 357) the agents should be *incapable of knowing* whether the earlier event, which they are now trying to cause, has occurred:

My conclusion therefore is this. If anyone were to claim, of some type of action A, (i) that experience gave grounds for holding the performance of A as increasing the probability of the previous occurrence of a type of event E; and (ii) that experience gave no grounds for regarding A as an action which it was ever not in his power to perform – that is, for entertaining the possibility of his trying to perform it and failing – then we could force him either to abandon one or other of these beliefs, or else to abandon the belief (iii) that it was ever possible for him to have knowledge, independent of his intention to perform A or not, of whether an event E had occurred.

As I mentioned above, Dummett is interested not only in arguing for the possibility of backwards agent-causation, but also, unlike me, in making it rational for the agent in his story to believe that it does occur. That is why he talks about *types* of events and not of *particular* events, and about probability judgements on experiential grounds. But since I am interested only in reinstating backwards causation's impossibility, I shall both discuss particular events and ignore the probability judgements of the agent in the story.

The situation is therefore as follows: if, on the fifth day, the chief *cannot* know whether the young men acquitted themselves with bravery, then his *nature as agent* does not commit us to making it possible in the story that he does not dance if the young men acquitted themselves with bravery. The bilking argument therefore does not work in this case. I have trimmed down the situations that allow for backwards agent-causation, but I have not cancelled them: agents can act in order to bring about an earlier event only if they cannot know whether the earlier event has occurred.

I have allowed for backwards agent-causation only in cases in which the agent cannot know whether the relevant past event has occurred. But what is the nature of the impossibility implied by this 'cannot'? Suppose that the young men's lion hunt is broadcast live on TV, but that the chief's set is broken; would that count as a sufficient sort of impossibility? I think it would not. The chief could pay a visit to his next-door neighbour, for instance. But suppose the local channel decided not to broadcast the biennial lion hunt; should we then say that the chief could not know the results of the event in time? Again, it seems we should not: the local channel *could* have broadcast it, or perhaps one of the young men *could* have used his mobile (times have changed in Africa since 1964).

But what if there is no TV in the chief's land, nor any mobiles or even wireless? Well, the observers could then signal by flaming torches that the young men had acquitted themselves with bravery. And then, from one post to another, each observer uncovering his pre-lit torch only if he sees the preceding one, the message would reach the chief almost at the speed of light – a system actually used in antiquity (as in Aeschylus' *Agamemnon*). Again, even if the chief's men do not use such a system, they could; and so

the chief *could* know the hunt's result on the fifth day. Consequently his dance could not be the cause of the young men's bravery.

So it seems that we are driven to some extreme kind of impossibility. We should say that the chief could not know the lion hunt's results only if it was physically impossible for information to arrive from *b* to *d*. No signal can leave *b* and arrive at the chief before he dances. I have demonstrated the need for such impossibility above by means of light, the fastest signal in our world; but had, say, sound been the fastest signal, I would have constructed an appropriate scenario utilizing it for my purposes. An agent's action *d* can cause an event *b* only if the fastest physically possible signal cannot leave *b* and arrive at the agent before *d*.

But this introduces a new twist into consideration. Temporal relations between distant events, as Einstein taught us, are not a trivial matter.<sup>11</sup> In fact, Einstein himself maintained (pp. 893–4) that there is more than one way to determine them. And in philosophical work on special relativity, the only theory in play which is capable of establishing a temporal ordering of distant events is the causal theory of time. Some maintain that it leaves room for differing conventions,<sup>12</sup> others that only one determination of temporal order relative to an inertial observer is possible by its means.<sup>13</sup> But I am not familiar with any other theory which can put constraints on the determination of temporal relations between distant events.

The classical formulation of the conceptual relation between causality and temporal order was supplied by Reichenbach. I think his formulation should be somewhat modified,<sup>14</sup> but these adjustments do not affect my argument here. I shall therefore make use of his formulation (p. 143):

We now turn to the comparison of two time series at different points in space. For this purpose we shall again use signals... Since we use signals, our previous definition of time order offers an important result. Let *e* be the event of the departure of the signal from *p*, and *e'* the event of its arrival at *p'*; then *e* and *e'* are two events connected by a signal; consequently they are ordered; *e'* must be later than *e*. Certain events are therefore already ordered although they belong to different temporal sequences.

On the other hand, Reichenbach continues (p. 144), if it is 'impossible to connect the events *e* and *e'* in either direction by a signal ... their time order

<sup>11</sup> Einstein, 'Zur Elektrodynamik bewegter Körper', *Annalen der Physik*, 17 (1905), pp. 891–921.

<sup>12</sup> H. Reichenbach, *Philosophie der Raum-Zeit-Lehre* (1928), tr. as *The Philosophy of Space and Time* (New York: Dover, 1958), §22; A. Grünbaum, *Philosophical Problems of Space and Time*, 2nd edn (Dordrecht: Reidel, 1973), ch. 12, §B.

<sup>13</sup> D. Malament, 'Causal Theories of Time and the Conventionality of Simultaneity', *Noûs*, 11 (1977), pp. 293–308.

<sup>14</sup> See my 'Causality and Temporal Order in Special Relativity', *British Journal for the Philosophy of Science*, 57 (2006), pp. 459–79.

is consequently not determined. We shall call such events *indeterminate as to time order*.<sup>7</sup>

It is now easy to see the implications of this position for backwards agent-causation. The chief's dance *d* is a cause of the young men's bravery *b* only if the chief at *d* could not know whether *b* has occurred. But the chief could not know at *d* whether *b* has occurred only if no signal from *b* could reach the chief at *d*. But then, objectively, *b* did not occur before *d*. So if *d* is the cause of *b*, *b* did not occur before *d*. So this is not a case of backwards agent-causation. Backwards agent-causation is therefore impossible.

What is going on here? I think Dummett presupposed unawares that the temporal order between distant events is somehow given, independently of the causal nexus. He then tried to impose causal relations upon the given temporal order as some additional structure. But despite Dummett's illustrious predecessors (Hume above all), this is mistaken. Temporal relations between distant events, to the extent that they are objectively determined, are determined together with the causal relations between them.

This result is plausible for the following reason as well. One could have claimed that there is something suspicious in the condition Dummett imposes on the possibility of backwards agent-causation. The chief's dance is supposed to have causal powers that reach not only into the future, but into the past as well. But why should the availability of information about the hunt influence the efficacy of the chief's dance? After all, his jumps and turns could be the same whether or not he knew about the hunt's results. We now have a solution to this puzzle: the causal powers of the dance are not affected by the availability of information. Rather, the dance can influence only what cannot have influenced it.

How can Dummett answer this objection? I can see only two possible ways to do so, but neither seems very promising.

First, he could claim that temporal relations are not determined by means of causal relations, or that they are not determined by means of the specific Reichenbachian rules formulated above. I am not acquainted, however, with any other way by means of which objective temporal relations can be determined; and it is doubtful whether a substitute for Reichenbach's approach is even possible. In addition, a claim one might be tempted to make, namely, that temporal order is primitive and not related conceptually to any other relation, is also implausible. The change which special relativity has introduced into our conception of causal order has affected our conception of temporal order as well, and this indicates that we do think of causality and temporal order as conceptually related. And secondly, we detect the temporal order between distant events not directly, but by observing other facts about them (the time at which we see each event, its distance

from us, etc.). So if temporal order between distant events is to be not a transcendent fact, but a relation with some empirical reality, as it seems Dummett assumes along with everyone else, it should be conceptually related to other notions.

I am not merely asking here how the arrow of time is to be determined, i.e., what distinguishes past from future. The way *this* question is usually posed in the literature already presupposes that we have some linear temporal order of the relevant events, a foliation, and we just need to determine in which direction the future lies. The question about the arrow of time presupposes that the events are already organized along a line, and it asks which direction along the line is the one towards the future. But if we reject Reichenbach's criterion, the very order or foliation of distant events, partial or total, is also lost. And if, as I think likely, no substitute for Reichenbach's approach is possible, then rejecting it would leave us with no objective temporal relations between distant events, and the claim that backwards causation is possible would lose most of its meaning. Some conventions would then allow for backwards causation, some would not; but the question whether backwards causation is *objectively* possible would have a false presupposition.

So as long as Dummett does not come up with an appropriate substitute for Reichenbach's approach, it seems that this response – 'temporal relations are not determined in this way' – is not open to him.

The second kind of response to my objection which may seem possible is as follows. Dummett could claim that I was not justified in interpreting the modality involved in his claim, that the chief could not know the outcome of the lion hunt, as *physical* impossibility. If it is some weaker kind of impossibility – if the chief could have known the results had he had a working TV set, or had his men signalled by torches, but he cannot since he does not have one, and they do not use this signalling system – if this kind of impossibility is sufficient for Dummett's purposes, then I have not found a contradiction in the idea of backwards agent-causation. But what kind of modality could that weaker modality be? Should we say the chief was able to know the lion hunt results only if he could know them by means he usually employs? Or by means people in his tribe usually employ? Or people in his country? Or should we also allow innovations that are not too difficult to come by? All this is *ad hoc*, devoid of any justification. Moreover, since we are enquiring into the *metaphysical* possibility of backwards agent-causation, it seems that some extreme sort of modality *is* the one relevant to the question whether the agent could know about the alleged effect of his action.

The same considerations show that an attempt to interpret the modality involved as context dependent, as Lewis did in his defence of the time-travel

variety of backwards causation,<sup>15</sup> is not promising either. Lewis maintained (pp. 142–6) that ‘can’, ‘is able’, ‘has the power’ and so on are equivocal: ‘To say that something can happen means that its happening is compossible with certain facts’, but which facts these are is determined by context. Accordingly, someone may be able to do something if we consider certain facts, but incapable of doing it relative to other facts. Lewis’ example was of a man who tries to hit a target but fails for some commonplace reason, such as a distracting noise. Given the man’s skill and training, and similar facts, he *could* hit his target; but given the fact that he did not hit his target, he *could not* hit it. Yet even if we accept Lewis’ analysis of modal concepts, it seems inapplicable to my case. In order to apply it, we need to claim that the chief could indeed know the hunt’s results in the sense that there were physical means of supplying him with the necessary information (and therefore the hunt preceded his dance), but that this is not the sense relevant to his action as an agent. But then what would the latter sense be? Again arbitrariness and considerations inappropriate for the question of the *metaphysical* possibility of backwards agent-causation seem inescapable.

Accordingly, this way of responding to my objection by exploiting a ‘weaker sense of *could*’, although perhaps not entirely blocked, does not seem promising either. And since I cannot think of any other way of trying to respond to it apart from the two I mentioned and rejected, I think Dummett failed to detect a route that could make backwards agent-causation possible.

To move on, my argument against the possibility of backwards *agent*-causation can be generalized to an argument against the possibility of backwards causation *simpliciter*. According to Reichenbach’s formulation, if *c* could cause *e*, then *c* occurred before *e*; therefore a cause cannot occur later than its effect. Or the other way round: if a cause can objectively occur later than its effect, then Reichenbach’s formulation should be rejected; but then we are left with no way of generally determining objective temporal relations between events; so we cannot then say that a cause can objectively occur later than its effect; consequently no cause can objectively occur later than its effect.

This argument against backwards causation *simpliciter* is fairly obvious (although I do not recall seeing it put this way in the literature). Still, if it is to be rejected, one should justify some method of determining objective temporal relations between events, either independently of their causal relations or at least in a way sufficiently different from Reichenbach’s.

This need to supply an alternative way of determining causal order often goes unnoticed. Dowe, for instance, defended ‘a version of Reichenbach’s

<sup>15</sup> Lewis, ‘The Paradoxes of Time Travel’ (1976), repr. in R. Le Poidevin (ed.), *The Philosophy of Time* (Oxford UP, 1993), pp. 134–46.

own theory about the direction of causation, the fork asymmetry account' (p. 227). He assumed that backwards causation is possible, and examined which conceptions of causality are consistent with it. But in doing this, he assumed that temporal relations are determined independently of causal ones, as if the causal network is superimposed on an already existing temporal one. Moreover, he assumed not only that the arrow of time is independent of causal relations (a claim often made, and usually followed by the claim that the entropy gradient determines time's arrow), but also that a temporal foliation of the relevant events is given as well. But without justifying the claim that this foliation, irrespective of the past–future direction of its leaves, is independent of the causal order, Dowe's causal theory cannot be taken as a sufficient justification of the coherence of backwards causation. And I cannot see how this can be objectively determined independently of the causal relations between events.

Although I think the prospects for this determination of temporal relations irrespective of causal relations seem rather poor, Dummett, as I mentioned above, thought that 'there is no particular conceptual difficulty about the conception of ... backwards causation' in a world that contains no 'creature to whom can be ascribed intention and purpose' (pp. 339–40). In support of this claim, he described an agentless world in which backwards causation seemed to him the more appropriate kind of cause. I shall therefore now turn to this description (p. 339), to see whether it supports Dummett's claim:

... imagine ourselves observing events in a world just like the actual one, except that the order of events is reversed.... The sapling grows gradually smaller, finally reducing itself to an apple pip; then an apple is gradually constituted around the pip from ingredients found in the soil; at a certain moment the apple rolls along the ground, gradually gaining momentum, bounces a few times, and then suddenly takes off vertically and attaches itself with a snap to the bough of an apple tree.

Similar poetic excursions can also be found in Tooley's work.<sup>16</sup> Unlike Dummett's, Tooley's fictional worlds do contain agents, who walk backwards, unfry an egg from an omelette, or perform similar feats.

Dummett next explains why, in the world he described, 'we should have to substitute a system of explanations of events in terms of the processes that led back to them from some subsequent moment'. I shall first show why such a substitution is probably unjustified for Dummett's world even from an explanatory point of view. I shall then supply a more general argument against the *possibility* of backwards causation in both Dummett's and Tooley's worlds.

<sup>16</sup> M. Tooley, *Causation: a Realist Approach* (Oxford: Clarendon Press, 1987), pp. 213–14, and *Time, Tense, and Causation* (Oxford: Clarendon Press, 1997), pp. 64–6.

Dummett's justification of the alleged substitution's necessity consists in demonstrating (pp. 339–40) how explanations by means of causes that precede their effects would leave us with 'inexplicable coincidences':

If through some extraordinary chance we, in this world, could consider events from the standpoint of the microscopic, the unpredictability would disappear theoretically ('in principle') although not in practice; but we should be left – so long as we continued to try to give causal explanations on the basis of what leads up to an event – with inexplicable coincidences. 'In principle' we could, by observing the movements of the molecules of the soil, predict that at a certain moment they were going to move in such a way as to combine to give a slight impetus to the apple, and that this impetus would be progressively reinforced by other molecules along a certain path, so as to cause the apple to accelerate in such a way that it would end up attached to the apple tree. But not only could we not make such predictions in practice: the fact that the 'random' movements of the molecules should happen to work out in such a way that all along the path the molecules always happened to be moving in the same direction at just the moment that the apple reached that point, and, above all, that these movements always worked in such a way as to leave the apple attached to an *apple* tree and not to any other tree or any other object – these facts would cry out for explanation, and we should be unable to provide it.

Indeed, explaining events in such a world by means of preceding causes alone would leave many inexplicable coincidences. Such explanations are therefore insufficient. But it does not follow that they should be supplemented with explanations that incorporate backwards causation. It is still possible that there is some third kind of explanation which would explain the apparent coincidences without introducing backwards causation. And such a third kind of explanation does indeed exist, and was applied to phenomena in a world quite like Dummett's imaginary one – ours.

In our world, a sapling 'is gradually constituted [from] the [apple] pip from ingredients found in the soil [and in the atmosphere]'. The sapling gradually grows, flowers, and then produces apples, 'and not ... any other [fruit] or any other object'. And we can quote Dummett further here, with appropriate adaptations: 'If through some extraordinary chance we, in [our] world, could consider events from the standpoint of the microscopic, the unpredictability would disappear theoretically ('in principle') although not in practice; but we should be left ... with inexplicable coincidences'. I could continue with these parallels, but I shall not labour the point.

These parallels demonstrate that if Dummett were right, and in the imaginary world he described, backwards causation were the more appropriate kind of causation, at least for certain processes, then it would be the more appropriate kind of causation in our world as well, at least for the parallel processes. But it is not so in our world; therefore neither is it so in Dummett's imaginary world. Dummett has failed to supply a description of

a world of which it is justifiably claimed that in it some causes act backwards.

These days we can supply a causal explanation, satisfactory ‘in principle’, for the goal-directed organic processes in our world: we refer to the omnipotent theory of evolution by natural selection. But this theory is a recent addition to our explanatory arsenal. In ancient times it was not available; yet nobody thought that organic processes involved backwards causation.

The most elaborate system of causal explanations the ancient world bequeathed to us is the Aristotelian one. Aristotle acknowledged that explanation by means of efficient causes alone would leave us with ‘inexplicable coincidences’. Some of these coincidences were explained, in his system, by final causes. Nature behaves in a goal-directed way, things directing themselves towards a final state. It is not, however, as if the final state exercises some power on preceding states, as the dancing chief was supposed to on the lion hunt: teleology is not backwards efficient causation. Rather, this directed behaviour is part of the nature of natural things. Because of some natural principle, they direct themselves towards a state of full actuality, the pip to the apple tree, the foetus to the adult, etc.

So Dummett’s imaginary world, like ours, can be explained by means of final causes, without leaving any residual inexplicable coincidences.

My argument above was intended to show that backwards causation is *unnecessary* for explaining scenarios of the kind Dummett brings. I shall next try to supply a stronger argument, intended to show that backwards causation is *impossible* even in Dummett and Tooley’s worlds.

But first, a note on my concept of causation may be in place, as I have not supplied any account of how I understand that relation. I did reject the Humean idea that causation is reducible to temporal order plus some kind of regularity; and from my discussion of final causes it should also be clear that I do not think the direction from cause to effect need be the direction of increasing entropy. But I prefer not to commit myself to any positive account of what this relation consists in. In fact, I suspect that if any concepts are bedrock, then the concept of causality, or at least of some forms of causality, is among them. We can of course characterize it in various ways (e.g., it is not symmetrical), and relate it to other concepts (e.g., temporal ones, as is done in this paper). But an explanation by means of more fundamental concepts is, I think, likely to be impossible. Yet if one does hold some reductive account of causation, then if it is consistent with the claims made in this paper about causation (for instance, its relation to temporal order), then my arguments against the possibility of backwards causation still apply, given that account.

I return, then, to Dummett and Tooley. Both include in their worlds agents equipped with human intelligence and with ordinary forwards causality between their mental events. The important thing for both is that these agents should not be able to *act* on the processes that are supposed to exhibit backwards causation. In Dummett's world 'we imagine ourselves as intelligent trees observing [the] world and communicating with one another, but unable to intervene in the course of events' (p. 339). Tooley's world is a little more elaborate; since it would be more convenient to use it as the target of my criticism below, I shall quote his scene-setting in some detail:

Consider ... the following world. It consists of two spatial regions segregated from each other by a wall with some remarkable properties. First, the wall has always existed. Secondly, it is indestructible, and so it will exist in any future time. Thirdly, there is no way of getting around the wall, so that something can travel from the one region of space to the other only by going through the wall. Finally, the wall's properties differ depending upon the direction through the wall: in one direction, no causal process at all can be transmitted from the one region to the other, while, in the opposite direction, light waves, but nothing else, can pass through the wall.<sup>17</sup>

The reader is now supposed to live 'on the side of the wall that can receive light waves from the other side', a side very similar to our actual world, and to observe the processes on the other side, processes that occur in an order opposite to that in which they occur in our world. This is supposed to supply the reader with a good reason for thinking that causation works backwards behind the wall.

I think both writers include these observers in their worlds for the following reason. If they had described a world like ours and then simply added that in that world temporal order is reversed, we might have been justified in maintaining that temporal order cannot be added merely by such a stipulation. If one held a causal theory of time, say, then one would claim that in these imaginary worlds temporal order is already determined as the causal order, and that consequently no possible case of backwards causation had been described. On the other hand, in these scenarios, which look imaginable and contain 'us' as well, the occurrence of backwards causation in the processes we cannot influence seems to have been *inferred*, relying on our very presence. Although this line of reasoning is at most implicit in Dummett's work, something close enough to it is explicit in Tooley's.

Of course, these scenarios, even if they worked, would not demonstrate the possibility of backwards causation in the world 'we' observed. One would be equally justified in maintaining that the processes on the observed side of the wall incorporated forwards causation, while the observers ('us')

<sup>17</sup> Tooley, *Time, Tense, and Causation*, p. 64.

actually had causality temporally backwards. Placing ‘us’ inside the imaginary worlds is a little misleading, for we are led to take only one of the two available viewpoints. What these scenarios would in fact demonstrate, if they worked, is that causal processes can occur in opposite directions (Tooley notes this). But I shall try to show that they fail even to do that.

The fact that in Tooley’s world we are mere observers of the events on the other side of the wall is of course essential. If we could influence what is going on there, the bilking argument would have been reinstated: just wait for the golf ball to take ‘spontaneously’ to the air, and then go and prevent the golfer from reverse-hitting it a little later. So I shall examine somewhat more carefully how the remarkable wall works. From the point of view of the people on the observed side of the wall, the wall just absorbs light waves that hit it (to simplify things, I assume it absorbs all of them). If, instead of Tooley’s remarkable wall, there were at the same place an ordinary wall, absorbing the light waves, the processes on their side would not be affected. Each thing would still have the same causal powers as it has with the remarkable wall in place. The wall does not participate in or influence the causal transactions on the observed side; it just prevents us from interfering with them.

So suppose the remarkable wall is replaced by some appropriate breakable one-way mirror or glass. The situation would now be as follows: if ‘we’, on the observers’ side, chose not to break through the wall and interfere with the processes on the observed side, the latter processes would have occurred in the same way as they did on Tooley’s original scenario. But if we chose to break through the wall, we could have prevented an effect’s later alleged cause from occurring, and in that way we could establish that it is not that effect’s cause (the bilking argument). But the nature of causality on the wall’s observed side is the same, whether or not we interfere in the processes occurring there – in his scenario, Tooley did not ascribe to the remarkable properties of his wall any *positive* causal role. So it follows that there is no backwards causation (backwards relative to the observers) in the processes on the observed side of the wall, whether or not there is any interference in these processes by agents from the wall’s observing side. So Tooley has failed to supply a scenario in which causal processes may occur in opposite directions.

More generally, the fault in Tooley’s scenario, as well as in Dummett’s intelligent trees one, can be characterized as follows. Since the direction of time cannot be introduced by stipulation, one has to include in one’s scenario two processes, apparently involving causation in opposite directions. Moreover, there should be some kind of interaction between these processes, in order to make them processes occurring in the same world – this

too cannot be done by stipulation alone. On the other hand, some constraints on the interaction between the two processes must be inserted, so that processes belonging to one series would not be able to prevent from occurring a putative later cause of an earlier event of the other series. And all this should be arranged in such a way that the constraints are not contingent relative to the nature of the causal processes involved – otherwise they could be removed, and in that way it would be demonstrated that one has failed to describe causal processes occurring in opposite directions. But Dummett and Tooley failed to supply such an arrangement.<sup>18</sup>

I have thus shown that one of the conditions Dummett finds necessary for an agent's action to produce an effect, namely, that the agent cannot know, at the time of the action, whether the effect has occurred, makes it impossible for the case to be one of backwards causation. Moreover, the same theory of objective temporal order as makes backwards *agent*-causation impossible makes backwards causation *simpliciter* impossible as well. Unless one supplies an alternative theory of objective temporal order, a theory which determines the foliation of the relevant events, one has not shown backwards causation to be possible. Next, the apparently backwards agentless world, in which Dummett thought backwards causation is the more appropriate kind of causation, can be explained, and in a sense was explained, without recourse to this kind of cause. Last, the attempt to construct a scenario that contains opposite causal processes, of the kind found in Tooley's work as well, failed, because the inability to intervene in an apparently reversed causal order was merely contingent. So there are good reasons for holding backwards causation to be an incoherent concept, and Dummett and others have failed to supply any scenario that should make us think otherwise. The ground, I hope, has been cleared.

*Central European University, Budapest*

<sup>18</sup> For lack of space I do not discuss in this paper another interesting attempt to defend the possibility of backwards causation, the one found in Huw Price's *Time's Arrow and Archimedes' Point*. Price tried to show the possibility of backwards causation in quantum phenomena that involve entanglement, relying on characteristics peculiar to these processes. Setting the stage for such a discussion would take too much space. Moreover, such a discussion, belonging to the philosophy of physics, would be different in character and audience from the one of this paper. I hope to show on some other occasion why I think Price has failed to describe a scenario in which backwards causation occurs.