42
THE DIRECTION OF TIME
D. H. Mellor

Formalities

The direction of time is the difference between being *earlier* than something and being *later* than it. The difference is not formal, since *earlier* and *later* are formally similar, each being the other's converse (any *x* is earlier than any *y* if and only if that *y* is later than that *x*) and both being transitive (if *x* is earlier/later than *y*, and *y* than *z*, then *x* is earlier/later than *z*). And if time is linear, i.e. if the passage of time returns nothing to its origin, *earlier* and *later* will also be irreflexive and asymmetrical: nothing will be earlier or later than itself, and nothing will be both earlier and later than anything else.

If anything does return to its origin, its world line (its path through space and time) will be a "closed time-like loop": global if time is circular and returns the whole universe to the Big Bang, local if the loop is the world line of a thing within it, like Dr Who's time machine TARDIS (Time and Relative Dimensions in Space), should backward time travel return that to its origin. In both cases *earlier* and *later* still share many formal properties, being now reflexive and symmetrical, since everything in a time-like loop is both earlier and later than itself and than everything else in the loop. And similarly in spatial loops, like the London underground's Circle line, where both clockwise and anticlockwise trains link every station to itself and to every other station.

But in both cases there are now formal differences too. For example, clockwise Circle line trains from Aldgate to Paddington go via Victoria, which anticlockwise trains do not. Similarly, going from earlier to later round circular time puts Cleopatra between the Big Bang and the Beatles, whereas going round from later to earlier puts the Big Bang between Cleopatra and the Beatles.

Extrinsic and intrinsic differences

Everyone agrees that time, whether linear or looped, is directed in a way that space is not. If so, *earlier* and *later* must differ in some substantive, non-formal way if the parallels noted in the first section, "Formalities," are not to give time's direction spatial counterparts. The difference between *earlier* and *later* must also be *intrinsic*, to distinguish it from substantive spatial differences, like that between *clockwise* and *anticlockwise*, which are merely extrinsic. For the Circle line direction from Paddington to Aldgate via

Victoria is only clockwise seen from above: seen from below it is anticlockwise; and in itself it is neither. No intrinsic non-formal feature distinguishes the two ways round a closed spatial loop, or the two opposed directions along an open spatial line: in both cases each is just the other's converse along a one-dimensional spatial path, with no substantive feature of the path itself telling us which is which.

I infer that, for time's direction to differ from that of any spatial dimension, it must follow from a substantive and intrinsic distinction between time and space, a distinction which, since it applies at all space–time points, must be local as well as global. These conditions set the tests that I think any adequate account of time's direction must pass: it must make that direction intrinsic, local and devoid of spatial counterparts.

The flow of time

The obvious account derives time's direction from its *flow*, i.e. from everything moving from the future to the past via the present. Theories that give this account I call *A-theories*, after McTaggart's (1908) distinction between an A-series of events ordered from past to future and a B-series ordered from earlier to later. And while different A-theorists credit A-series locations with different entailments – e.g. that only the present exists (Bourne 2006: Pt 1), or that only the past and present do (Tooley 1997: Ch. 2) – all that matters here is whether time flows at all (Oaklander and Smith 1994), a controversy about which I need only make two points. First, since the flow of time, if it exists, distinguishes time from space everywhere, it is, as required, both intrinsic and local. But second, it will still only give time its direction, by distinguishing *earlier* from *later*, if this distinction is derivable from that between *past* and *future*, not *vice versa*. That is why A-theorists think that "the intrinsic sense of a series of events in Time is essentially bound up with the distinction between past, present and future. A precedes B because A is past when B is present" (Broad 1923: 58). They are wrong: A-series locations are distinguishable only by how much earlier or later they are than the present: *yesterday* as one day earlier than *today*, for example, and *tomorrow* as one day later than it (Mellor 1998: Ch. 1.3). And so in general, *pace* Broad, the fact is that *A is past when B is present* because *A precedes B*, and not the other way round. An A-series is simply a B-series plus a present moment that, by mere definition, moves from earlier to later times rather than from later to earlier ones. But then even A-theorists need a B-theory account of time's direction to tell them what it is for a present to move from earlier to later times and not the other way. What might that theory be?

The expansion of the universe

B-theories of time's direction use an "arrow of time," a "process or phenomenon that has … a definite direction in time" (Savitt 1995: 1), such as the expansion of the universe (Zeh 2007: Ch. 5.3). But even if the universe expands for ever, this "cosmological" arrow cannot be intrinsic to time. For if it was, the universe could never stop expanding, since that would make the end of its expansion the end of time itself. But that is neither credible nor consistent with theories that take the universe's continued expansion to depend on a contingent balance of gravitational and other forces (Dainton 2001: Ch. 4.6), a contingency which requires time to have a logically independent direction that need not be reversed or destroyed if the expansion stops.

The cosmological arrow fails our other test, too, because it cannot give time a local as well as a global direction. For just as a child's growth is a fact about its whole body, not about any one of its cells, so the universe's expansion is only a fact about all of it, not about any point or thing within it. The expansion of the universe could only give time a local direction if events, like my typing of these words, could only occur successively, rather than simultaneously, while, and because, the universe continues to expand; and no one believes that.

Increasing entropy

Another arrow of time is provided by the fact that while all other forms of energy are completely convertible into heat, the converse is not true (Atkins 1986: 86). More precisely, let *S* be the entropy of a thing *a*, where *S*'s changes are, when well-defined, given by

$$dS = dq/T,$$

where *T* is *a*'s absolute temperature; and d*q* is *a*'s net gain in heat from internal and external sources. Then while *a* is thermally isolated, i.e. while no heat is transferred into or out of it,

$$dS/dt \geq 0 \text{ (Denbigh 1955: 26)},$$

where *t* is time. In short, the entropy of a thermally isolated thing never decreases, so that if it is lower at t_1 than at t_2, then t_1 is earlier than t_2. This is the "thermodynamic" arrow of time.

The trouble with this arrow is that nothing is ever wholly isolated thermally, which is why the entropy of many things, including ourselves, can and often does decrease. The fact that these decreases have to be matched or exceeded by increases elsewhere only lets this arrow, like the cosmological one, give time a global direction, and then only because, by definition, there is nothing outside the whole universe for it to interact with. Nor does it help that if a thing *was* thermally isolated, its entropy would never decrease. For this can only be true, as a matter not of definition but of thermodynamic fact, if all world lines already have a direction for entropy to increase in, a direction that must therefore have a different basis. That is why, "when we … write $dS/dt \geq 0$, this is to take $+t$ as being towards 'the future' and $-t$ as being towards 'the past'" (Denbigh and Denbigh 1985: 15), a conclusion that is only reinforced in the statistical and quantum theories that have succeeded classical thermodynamics. For in these theories, the entropy of a thermally isolated thing has a non-zero chance of decreasing for a while, a chance it could not have if the thermodynamic arrow gave time a local direction.

Irreversibility

Neither these nor any of the other arrows listed by Zeh (2007: 5) gives time a direction that is both local and intrinsic. The real question about these arrows is not what links them to the direction of time but what links them to each other: why, for example, is the direction in which the universe expands that in which entropy increases? That is a good question, but it is not a question about the direction of time itself. To take a spatial analogue: "No one has seriously maintained that space is "handed" … because of the *de facto* asymmetries between left- and right-handed objects. What then is supposed to make time different from space in this respect?" (Earman 1974: 32). What indeed? What tempts us to identify the direction of time with that of one of its arrows?

The best way to answer this question is to ask what proves that a movie is being played backward. Not how it shows things moving, for things can move in any direction; nor how people in it speak, since nothing rules out a language that sounds like a real language spoken backward. The real giveaway is its showing time's arrows pointing the wrong way: the universe contracting, or entropy spontaneously decreasing (as in the separation, with no energy input, of brine into fresh water and solid salt). It is because a movie that shows this *must* be running backward that we are tempted to identify time's direction with that of more or less irreversible processes.

To see why we should resist the temptation, consider the theory that positrons are electrons travelling backward in time, i.e. that the direction of time is reversed along the world lines of positrons (Reichenbach 1956: Ch. 30). The theory exploits the fact that positrons differ from electrons only in their charge being positive instead of negative, thus making positive charges, which attract electrons, repel them, and negative charges, which repel electrons, attract them. This makes a movie of electrons played backward look exactly like one of positrons played forward, just as a movie of positrons played backward looks like one of electrons played forward. But if the first similarity shows positrons to be electrons travelling backward in time, must not the second show electrons to be positrons doing so? But then which is it: which are the real backward time travellers, positrons or electrons?

Neither: the two particles simply have different properties, which is why movies of them differ when played the same way, whether forward (i.e. as shot) or backward. The fact that this difference happens to make either movie played backward look like the other one played forward is irrelevant. And as in this case, so in general: irreversible processes whose directions are not intrinsic to time – as its flow is – can tell us nothing about a direction of time that they must all presuppose. This is not to decry the questions they raise: the one noted above, of how independent their directions are, and the prior one of why they *have* directions, i.e. why they always or mostly go one way when no basic law of physics stops them going the other way. But answering those questions will not tell us what gives time itself its direction.

Seeing the direction of time

Perhaps nothing does. Why, after all, must time get its direction from something else, when the "directions" of increasing mass, size and other quantities do not? Why seek *The Physical Basis of the Direction of Time* (Zeh 2007), when no one seeks a "physical basis," i.e. a basis in something else, for differences between increasing and decreasing values of other quantities? It cannot be because telling which of two events is the earlier is harder than telling which of two things is the lighter or smaller: it is not. Seeing that a clock hand is moving clockwise, for example, includes seeing that its end passes nearby points in a certain order, e.g. that it passes the figure "1" *earlier* than it passes "2." If we could not see the time order of events, we could not see which way clock hands are moving, which we can.

Yet the direction of time that we perceive, and then use to distinguish reversible from irreversible processes, and to say which way the latter go, still needs a non-temporal basis. The reason lies in the fact, noted by Kant (1968 [1781]: Second Analogy) and others, that

> a succession of feelings, in and of itself, is not a feeling of succession. And since, to our successive feelings, a feeling of their own succession is added, that must be treated as an additional fact requiring its own special elucidation (James 1890: Vol. 1, 628–9).

For example, no single "feeling," i.e. experience, can tell me that my clock hand passes "1" (event *e*) earlier and not later than it passes "2" (event *f*). I need two: first the experience of seeing *e*, and then that of seeing *f*. Yet these two experiences are, as James says, not enough in themselves to tell me that *e* precedes *f*, since they will not tell me this if, when I see *f*, I have quite forgotten seeing *e*. This is why, if my seeing *f* does tell me that *e* preceded *f*, that is "an additional fact requiring its own special elucidation."

I think the elucidation is causal (Mellor 1998: Ch. 10.5): my seeing *f* will only make me see that *f* is later than *e* if it is *affected* by my having seen *e*. What the effect is, what produces it, and whether I am aware of it (i.e. whether, as James assumes, it is itself an experience) are immaterial; but an effect there must be. Of course the effect, whatever it is, will not ensure that what it makes my seeing *f* tell me, namely that *e* precedes *f*, is *true*, since it only links my perceptions of *e* and *f*, not *e* and *f* themselves. That is how, since light travels faster than sound, when I hear thunder seconds after I see lightning, my senses can tell me, falsely, that the thunder is that much later than the lightning. But all this shows is that here, as elsewhere, we should not believe everything our senses tell us.

How much can this causal account of how we perceive the time order of events tell us about their actual order? Well, since most if not all causes do in fact precede their effects – time *does* have a causal arrow – it does at least explain why the time order of our perceptions of events is the time order which these perceptions tell us, truly or falsely, that those events have. That is the causal basis of our perception, and hence of our conception, of the direction of time. But this is not enough to give the direction itself a causal basis; that requires another link between causation and time.

Causal and temporal order

The obvious candidate for the extra link we need between causal and temporal order is our inability to affect the past (some of which we can perceive) and to perceive the future (some of which we can affect). More precisely, and in B-theory terms, at any time *t*, we cannot affect anything earlier than *t* but may perceive it, and cannot perceive anything later than *t* but may affect it. And these are not two differences –between what we can and cannot affect, and can and cannot perceive – but one, since our perceptions are themselves effects of what we thereby perceive. What stops our senses showing us the future is the very fact that stops us affecting the past, namely, the fact that causes precede their effects, a fact that identifying time order with causal order immediately explains.

Time's causal arrow also shows how the positrons of the sixth section, "Irreversibility," differ from time-travelling electrons, by making a locally reversed time order entail a locally reversed causal order that positrons and electrons never exhibit. (For example, deflecting positrons at any time *t* only ever affects where they are after *t*, not before it.) Whereas when Dr Who's TARDIS travels back in time, its doing so automatically makes effects within it, that are later than their causes by TARDIS time, earlier than those causes by outside time.

Better still, the causal arrow explains special relativity's distinction between events that could be linked by things moving at or below the speed of light, which is what makes their space–time separation *time-like*, and events that could not, whose separation is *space-like*. This can be explained as follows (Mellor 1998: 108). If nothing causes a changeable property *F* of a thing *a* to change between a time t_1 and a later time t_2, then *a*'s being *F* at t_1 can cause *a* to be *F* at t_2. This enables whatever causes *a* to be *F* at t_1 to be an indirect cause of whatever *a*'s being *F* at t_2 causes, as when light made red by reflection from Mars causes us to see later that Mars is red. And if this is what transmits causation across space, then relativity's letting nothing accelerate to more than the speed of light makes light the fastest possible transmitter of causation. This in turn explains why the space–time separation of causally related events is always time-like, and why the time order of all events with time-like separations is the same in all reference frames. These two explanations, by making causation what distinguishes time from the spatial dimensions of space–time, and making time order coincide with causal order, are what make time's causal arrow intrinsic to time and therefore a credible source of time's direction.

Causation and time

Time's other arrows cannot begin to match this: none of them is both local and intrinsic to time; and none explains how we perceive (and conceive of) the direction of time, why we cannot affect the past or perceive the future, and what distinguishes time from space. Yet despite its long history (Robb 1914; Reichenbach 1956: Pt 2; Grünbaum 1968: Ch. 7), the causal theory is still often rejected (Lacey 1968) or ignored – it is not on Zeh's latest (2007: 5) list of time's arrows – for reasons I must therefore now outline and try to rebut.

(a) Because causal theories of time cannot define a cause as the *earlier* member of a cause–effect pair; they cannot use Hume's definition of a cause as "an object, *followed* by another, and where all the objects similar to the first are *followed* by objects similar to the second" (Hume 1975 [1748]: §60; my italics). But then this definition needs supplementing with a different theory of time order to account for *non*-temporal differences between cause and effect, for example, that causes explain their effects but not *vice versa*. One such theory is the A-theory:

> that past and present events and states of affairs are fixed …, whereas at least some future ones are still to be fixed … [and] if at any time A is fixed while *B* is still unfixed, *B* cannot be causally prior to A, because … *B* [might] not occur. (Mackie 1974: 178)

How can B-theorists, who cannot use A-theories of how causes differ from effects, account for that difference? One way follows from Hume's counterfactual rewording of his definition: "… in other words where, if the *first* object had not been, the *second* never had existed" (Hume 1975 [1748]: §60; my italics). In other words, stripped of its temporal implications: if a cause had not existed, nor would its effects. This account of causation, which may or may not include a possible-worlds analysis of its counterfactuals (Lewis 1973), beats Hume's first definition hands down by making causation local (Lewis 1973: 558) but not – since a counterfactual does not entail its converse – symmetrical: it does not make effects as necessary for their causes as it makes causes for their effects. But nor does it make causation asymmetrical: it does not stop effects causing their own causes. And while a possible-worlds analysis of its counterfactuals may stop most of them doing so, it will not rule out backward causation altogether (Lewis 1973: 567). And although I take this to be a defect, most philosophers do not.

(b) Counterfactual theories, however, have trouble with causes that are not necessary for their effects, i.e. where "C causes E" fails to entail that *E* would not occur if C did not, as it often seems to do. Fred's smoking can, for example, cause his cancer even if he might have got cancer had he quit. What *can* handle these cases is a probabilistic theory of causation, provided its probabilities are not just the frequencies – e.g. the frequencies with which smokers and non-smokers get cancer – that Reichenbach (1956: §12) and others think they are. For these frequencies, like Hume's constant conjunctions, cannot apply to single events: Fred's cancer cannot be more or less frequently conjoined with his smoking or with his not smoking. Only theories whose probabilities are single-case chances (Mellor 1995: Chs 4–5), like Fred's chances of getting cancer if he smokes and if he doesn't, will make probabilistic causation give time a local direction.

(c) Some causes and effects seem simultaneous, as when "if I view as a cause a ball which impresses a hollow as it lies on a stuffed cushion, the cause is simultaneous with the effect" (Kant 1968 [1781]: A203). This, however, conflicts with laws like Newton's third law of motion, which implies, for example, that the momentum of each of two colliding things will cause the other's momentum to change. For that, if these causes and effects were simultaneous, would require each thing to have its changed and unchanged momentum at the same time, which is impossible (Le Poidevin 1991: Ch. 6). This is one of several incentives to try and explain away apparent cases of simultaneous causation, which I do in my *Real Time II* (1998: Ch. 10.3).

Modern physics may also allow causation across space-like intervals, provided it is either unmediated or mediated by *tachyons*, entities that always travel faster than light, which relativity allows. But the only evidence for this is quantum *non-locality*, where the result of one measurement fixes the result of another made at a space-like interval (Redhead 1983); and what this threatens is not the time order of cause and effect but only "the sense of locality that requires that correlation between space-like separated events always be factorable-out by a common cause" (Skyrms 1980: 127). In particular, since non-locality only links the *results* of measurements, not their being *made*, it does not enable faster-than-light signalling, that is, it does not turn the making of a measurement into an effective means of producing a specific result at a space-like interval. Yet it is a long-recognised implication of causation that causes *are* means to their effects in this sense (Gasking 1955; Mellor 1995: Ch. 7; Price 1996: Ch. 6), an implication that stops non-locality ruling out a causal distinction between time-like and space-like separations.

(d) Identifying time order with causal order seems to rule out backward causation, which is after all conceivable. But that is no objection to an identity proposed, not as a mere analysis of our concepts of causation and time, but as a substantive claim about causation and time themselves. Conceptual analysis of our causal discourse may be "… a guide to our main topic and an introduction to it; but it is not in itself our main topic, and with regard to that topic its authority is far from absolute" (Mackie 1974: 1). And anyway, as we saw in subsection (a), above, this identity does not rule out what we *call* backward causation, like that entailed by backward time travel: it merely makes it entail a local reversal of time.

Nor does identifying temporal with causal order rule out the time-like loops of the first section: identifying the *later* direction in them with that of causation does not make them impossible. Their possibility is indeed often inferred from Gödel's (1949) proof that a global loop is consistent with general relativity (Bourne 2006: Ch. 8) or Lewis's (1976) argument for the consistency of some local loops produced by backward time travel. I reject both inferences, for reasons given in my *Real Time II* (1998: Ch. 12), but either view is compatible with a causal theory of time.

(e) How can causation give the time order of events that are *not* causally related? My answer (Mellor 1998: 113) is that each space–time point is the location of many facts – about density, temperature, the intensity of electromagnetic and other fields, etc. – that *are* causally related to similar facts at other points. And all it takes for causation to fix the time order, if any, of two space–time points t_1 and t_2 is that some fact at t_1 (say) causes some fact at t_2, thereby making each fact at t_1 precede all facts at t_2, whether or not it causes them.

But if the time order of all facts at t_1 and t_2 follows from the causal order of only some of them, might not one fact P at t_1 cause a fact Q at t_2 while another fact R at t_2 causes a fact S at t_1, thereby making t_1 both earlier and later than t_2? Indeed it might, if the causal loops are possible that the causal theory makes backward time travel entail, as when TARDIS travels back in time from t_2 to t_1, with causation outside TARDIS making t_1 earlier than t_2, and causation inside it making t_1 later than t_2. If such loops are possible, as many philosophers suppose, it will also be possible for two space–time points to be both earlier and later than each other. Yet even if this is metaphysically possible, because "the openness of the causal chains [is] an empirical fact" (Reichenbach 1956: 37), and not the contradiction I think it is (Mellor 1998: Ch. 12.4), that openness may still, and I believe will, stop it ever happening in fact.

Conclusion

Inferring from subsections (a)–(e) that time *can* get its direction from its causal arrow, and from earlier sections that it *cannot* get it any other way, I conclude that time is indeed the causal dimension of space–time, with an intrinsic direction that it gets from that of causation.

References

Atkins, P. W. (1986) "Time and Dispersal: The Second Law," in R. Flood and M. Lockwood (eds), *The Nature of Time*, Oxford: Blackwell, pp. 80–98.

Bourne, C. (2006) *A Future for Presentism*, Oxford: Clarendon Press.

Broad, C. D. (1923) *Scientific Thought*, London: Kegan Paul.

Dainton, B. (2001) *Time and Space*, Chesham, UK: Acumen.

Denbigh, K. G. (1955) *The Principles of Chemical Equilibrium*, Cambridge: Cambridge University Press.

Denbigh, K. G. and Denbigh, J. S. (1985) *Entropy in Relation to Incomplete Knowledge*, Cambridge: Cambridge University Press.

Earman, J. (1974) "An Attempt to Add a Little Direction to 'the Problem of the Direction of Time'," *Philosophy of Science* 41: 15–47.

Gasking, D. (1955) "Causation and Recipes," *Mind* 64: 479–87.

Gödel, K. (1949) "A Remark about the Relationship between Relativity Theory and Idealistic Philosophy," in P. A. Schilpp (ed.), *Albert Einstein: Philosopher–Scientist*, La Salle, IL: Open Court, pp. 557–62.

Grünbaum, A. (1968) *Modern Science and Zeno's Paradoxes*, London: Allen & Unwin.

Hume, D. (1975 [1748]) *An Enquiry Concerning Human Understanding*, in *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, 3rd edn, edited by L. A. Selby-Bigge and P. H. Nidditch, Oxford: Clarendon Press, pp. 5–165.

James, W. (1890) *The Principles of Psychology*, London: Macmillan.

Kant, I. (1968 [1781]) *Critique of Pure Reason*, London: Macmillan.

Lacey, H. M. (1968) "The Causal Theory of Time: a Critique of Grünbaum's Version," *Philosophy of Science* 35: 332–54.

Le Poidevin, R. (1991) *Change, Cause and Contradiction: A Defence of the Tenseless Theory of Time*, London: Macmillan.

Le Poidevin, R. and MacBeath, M. (eds) (1993) *The Philosophy of Time*, Oxford: Oxford University Press.

Lewis, D. K. (1973) "Causation," *Journal of Philosophy* 70: 556–67.

—— (1976) "The Paradoxes of Time Travel," *American Philosophical Quarterly* 13: 145–52.

Mackie, J. L. (1974) *The Cement of the Universe*, Oxford: Clarendon Press.

McTaggart, J. M. E. (1908) "The Unreality of Time," *Mind* 18: 457–84.

Mellor, D. H. (1995) *The Facts of Causation*, London: Routledge.

—— (1998) *Real Time II*, London: Routledge.

Oaklander, L. N. and Smith, Q. (eds) (1994) *The New Theory of Time*, New Haven, CT: Yale University Press.

Price, H. (1996) *Time's Arrow and Archimedes' Point*, New York: Oxford University Press.

Redhead, M. L. G. (1983) "Nonlocality and Peaceful Coexistence," in R. G. Swinburne (ed.), *Space, Time and Causality*, Dordrecht: Reidel, pp. 151–89.

Reichenbach, H. (1956) *The Direction of Time*, edited by M. Reichenbach, Berkeley: University of California Press.

Robb, A. A. (1914) *A Theory of Time and Space*, Cambridge: Cambridge University Press.

Savitt, S. F. (ed.) (1995) *Time's Arrows Today*, Cambridge: Cambridge University Press.

Skyrms, B. (1980) *Causal Necessity*, New Haven, CT: Yale University Press.

Tooley, M. (1997) *Time, Tense, and Causation*, Oxford: Clarendon Press.

Zeh, H. D. (2007) *The Physical Basis of the Direction of Time*, 5th edn, Berlin: Springer.

Further reading

B. Dainton, *Time and Space* (Chesham, UK: Acumen, 2001), chapters 1–8, provides a good introduction to the topics of this paper. For the flow of time, see R. Le Poidevin and M. MacBeath (eds), *The Philosophy of Time* (Oxford: Oxford University Press, 1993), pts 1 and 3; for time's non-causal arrows, S. F. Savitt (ed.), *Time's Arrows Today* (Cambridge: Cambridge University Press, 1995); and for causal theories of time H. Reichenbach, *The Direction of Time*, edited by M. Reichenbach (Berkeley: University of California Press, 1956), pt 2; and D. H. Mellor, *Real Time II* (London: Routledge, 1998), chapters 10–12.