Teleology and Spinoza's Conatus

Jonathan Bennett

From: *Midwest Studies in Philosophy* 8 (1983), pp. 143–160.

1. Spinoza's challenge to teleology

Reports on Spinoza's views about goals or purposes or 'final causes' tend to focus on his rejection of cosmic or divine purpose. But that is not all he rejected: he was opposed to all 'final causes', all teleological explanation, even of human action; and that gives the *Ethics* some peculiar features that I will expound in this paper.

Spinoza has two general objections to teleology, both given in the Appendix to Part I of the *Ethics*. To get a hold on them, let us consider a small fragment of the natural world: A certain event occurs in my brain, which causes me to raise my hand, which in turn causes the deflection of a stone that has been thrown at my face. For short: Brain causes Raise, which causes Deflect. If I say that I raised my hand in order to deflect the stone, I purport to explain why Raise occurred. 'Why did you raise your hand?' 'So as to deflect the stone.' But Spinoza thinks that to explain something is to say what causes it, and so he thinks that the above explanation purports to give Deflect a role in the causing of Raise. He objects to this on two grounds: **(i)** the role of 'cause of Raise' is already filled, namely by Brain, and **(ii)** Deflect cannot enter into the causing of Raise since the causal flow runs the other way, i.e. Raise causes Deflect. I will concentrate on **(ii)** rather than **(i)**. That is, I will not emphasize Spinoza's view that teleological explanations are wrong because they put items into causal roles that have been preempted by other items; rather, I will emphasize his view that they are wrong because: 'This doctrine concerning an end turns Nature completely upside down. For what is really a cause it considers as an effect, and conversely what is an effect it considers as a cause. What naturally comes before, it puts after.' In short, Deflect cannot help to explain Raise, because Raised causes Deflect.

In an earlier treatment of this matter, I said that Spinoza's point concerned the attempt to explain an event through a *later* event, and Parkinson has criticized this on the ground that 'Spinoza does not view causation in this temporal way'.[1] He is right about that. Spinoza's extreme rationalism makes it hard for him to give theoretical weight to time differences, and, in particular, his conflation of causal with logical necessity forbids him to make such differences central in his account of the cause-effect relation. Still, Raise does occur before Deflect, and Spinoza would have regarded that as showing that Deflect does not cause Raise. Whatever

[1] G. H. R. Parkinson, 'Spinoza's Concept of the Rational Act', *Studia Leibnitiana Supplementa* 20 (1981), pp. 1–19, at pp. 6–7 (n. 11).

his theoretical difficulties about time, if he were told that 'something happened yesterday that caused a house to burn down two days ago', he would surely think 'That can't be right!' on the ground that causes cannot postdate their effects; and so he could also give that as a sufficient reason for objecting to many teleological explanations, including the Raise-Deflect one that I have been discussing. Still, I concede that to state his entire case against teleology in that temporal way is to lose generality and to mislocate the center of gravity of his thought. In what follows, I will use the general causal statement of Spinoza's objection to teleology, though keeping the temporal version in sight as well. The choice between these makes no difference to strategy and virtually none to tactics. Spinoza says that teleological explanations order things wrongly, and we have to show that they do not: We will not be able to straighten him out unless we know what ordering of events he thought to be uniquely correct, but we do not have to know why he thought so.

Spinoza complains that a teleological explanation misrelates a pair of events, such as Raise and Deflect. Sometimes, however, there is only one event: We might say, 'He raised his hand so as to deflect the stone', in a case where the hand went up too slowly and the stone got through. In this case, the charge of 'misrelating a pair of events' does not get off the ground, because the only relevant event is Raise—there is no second event we could name Deflect.

Spinoza would say that there is trouble here, too, I think. He would object to our pretending to explain an item with the aid of the concept of something further down the causal or temporal stream, i.e. something—actual or possible—that does not lie in the causal ancestry of the item being explained. Idle mentions are harmless, of course, as in 'The vase was caused to fall by a push from the man who would become President'. But no tolerable explanation can *need* to mention actual or possible effects of the thing being explained—or so Spinoza thought.

Why not? It will not do merely to say that what explains Raise must cause it and Deflect cannot cause it since it is caused by it; for now we are looking at a case in which there is no such event as Deflect. Spinoza might say that that is worse than ever, for now it is being pretended that Raise is caused by a subsequent event that does not even happen! But that sounds wild and unconvincing: someone who says it, we are apt to think, must have gone astray somewhere in his or her thinking about teleology. Still, it is one thing to say that Spinoza no longer has a clear account of what is wrong with saying 'I raised my hand so as to deflect the stone'; it is another to show what is right with it. *Can* we really explain Raise in a way that essentially involves mentioning a possible event which, if it is actual, is caused by Raise?

Yes, we can, and in section 6 I will show how. First, though, a simpler and more inviting way of dealing with Spinoza's problem ought to be discussed.

2. Braithwaite's partial response

It seems natural to suggest that at least some teleological explanations are all right, namely, those that explain an action by reference to *thoughts about* its possible effects. Thus, according to Braithwaite:

> There is one type of teleological explanation in which the reference to the future presents no difficulty, namely explanations of intentional human action in terms of a goal to the attainment of which the action is a means. [Such] explanations are always understood as reducible to causal explanations with intentions as

[1] R. B. Braithwaite, *Scientific Explanation* (Cambridge University Press, 1953), pp. 324f, quoted with omissions.

> causes; to use the Aristotelian terms, the idea of the 'final cause' functions as the 'efficient cause'.[1]

This seems so obviously available as an answer to the question 'How can an event be explained with the aid of a mention of a possible effect of it?' that one wonders why Spinoza did not avail himself of it and drop much of his opposition to teleology.

There is an easy, obvious reason why Spinoza would not accept Braithwaite's proposal just as it stands. Since Spinoza holds that there can be no causal commerce between the mental and physical realms, he could not allow that any thought or 'idea' could cause Raise. But we can modify Braithwaite's proposal, consistently with Spinoza's own views, so that it clears this hurdle and yet still presents Spinoza with a teleological challenge.

The key to the modification is Spinoza's doctrine of parallelism between the mental and physical realms (2p7), which implies that physical causal chains are always matched by mental ones, and vice versa. This implies that when Raise seems to be caused by a thought of mine, it is caused by some physical (perhaps cerebral) partner of that thought; and the thought itself does cause (not Raise itself, but) a mental partner of Raise. So there are two causal chains here, according to Spinoza. One is physical:

(1) physical correlate of thought of deflection → Raise.

The other is mental:

(2) thought of deflection → mental correlate of Raise.

Neither involves a causal crossing of the boundary between physical and mental. In thus avoiding interaction between mind and body, far from destroying Braithwaite's proposed form of teleological explanation, we have turned it into two!

If **(1)** and **(2)** above are genuine causal transactions, why can they not support teleological explanation? In each of them, an item x *is caused by*—and is thus explainable through—a previous item that somehow involves the concept of a possible *effect of* x. In **(1)** Raise is caused by a brain event that is, in a sense, 'of' the stone's being deflected; in **(2)** the mental correlate of Raise is caused by a thought of the stone's being deflected. But neither of them has the faintest appearance of implying that Deflect causes Raise; so Spinoza is not entitled to say outright that they reverse the order of nature by treating effects as causes. What, then, can he say about them?

3. How could Spinoza have replied?

I do not know whether Spinoza explicitly thought about this way of explaining an item with help from the concept of a possible effect of the item. If he did, he must have rejected it, and reasons for doing so can be found in his thought.

They turn on the fact that a cause of x has features that do not contribute to its causing x, including some that do not contribute to its causing anything. The fall of a vase may be caused by a push that **(i)** occurs across the middle of the table in a northerly direction, **(ii)** extends through ten inches, **(iii)** is accompanied by a snapping of fingers, **(iv)** is just like a certain movement that Olivier makes in his film version of Hamlet, and **(v)** is performed by someone who will become President ten years later. The first two of these are relevant to the push's causing the fall, the third is relevant to some of its causal powers though not to that one, and the last two are arguably irrelevant to all its causal powers.

Now, with reference to causal chain **(1)**, Spinoza would say that the physical event that causes Raise is not helped to do so by having the feature that links it with the deflection of the stone, and in **(2)** he would say that the thought that causes the mental correlate of Raise is not helped to do so by being a thought of the deflection of the stone. In each case, the cause has a deflection-involving feature, but its

other features are sufficient for it to cause Raise **(1)** or the mental correlate of Raise **(2)**. Indeed, the position is even stronger than that. Spinoza, as I understand him, would say that those deflection-involving features are entirely causally impotent: They do not contribute to any of the causal powers of the items that have them. He would regard them as representative rather than intrinsic features of the items that have them, and he thinks that a thing's causal powers depend only upon its intrinsic nature. To get the feel of this position, consider the thesis that the causal powers of a bit of paper with ink marks on it may depend on size, shape, chemical composition, etc., but will never depend upon whether it is a map of Sussex. Though the example needs refining, even in its rough form it may help to give the general idea.

Having attributed this very strong view to Spinoza, I should defend it in more detail, taking the two cases separately.

(1) Spinoza allows that a state of one's body may be 'of' something else: He calls such physical states 'images'. The only ones he explains are caused by what they are images of, so that my image of' you is a state that my body is caused to be in by your body. I do not see how to extend that to cover a brain state that is 'of' a nonexistent state of affairs, e.g. a possible future deflection of a stone, but I need not wrestle with that problem. What matters here is that Spinoza seems to have assumed, firmly and deeply, that the causal powers of a physical item depend wholly upon its intrinsic properties, such as the shapes, sizes, positions, and velocities of particles, and never on any representative or 'of'-ish feature it might have. The physical theory expressed in the lemmas inserted between 2p13 and 2p14 leaves no room for doubt about this. It follows that, although the cerebral item that caused Raise is correlated with my thought of the stone's being deflected, that feature of it cannot have contributed to its causing of Raise. A fortiori, this causal transaction does not enable us to put the concept of an effect of x to work in explaining x, and so it does not threaten us with teleology.

(2) I contend that Spinoza would also hold that the representative features of mental items contribute nothing to their causal powers. Thus, when I have the thought that P, this thought is a psychological particular that has various features that enable it to have various mental effects, but its representative feature—its having the content *that P*—is not causally efficacious. Spinoza is forced to accept this by pressure from his doctrine of parallelism between the mental and physical realms: 'The order and connection of ideas [mental items] is the same as the order and connection of things [basically physical items]' (2p7). The very wording of this and still more Spinoza's handling of it throughout the *Ethics* imply a strict isomorphism between the mental and physical realms, with similarities in one mapping onto similarities in the other and with causal chains in one mapping onto causal chains on the other. Now, if I was right in my previous claim that Spinoza holds that a physical item's causal powers depend solely on features of it having to do with positions, velocities, shapes, and sizes of particles, then he must make the causal powers of mental items depend upon features of them that are systematically correlated with those physical features.

That seems to foreclose the possibility that the causal powers of mental items should depend on their content, their representative features, what they are 'of' or 'about' or what they 'say'. Spinoza's parallelism doctrine is a lot to swallow, but it would be even harder to choke down the thesis that features having to do with positions and velocities, etc. can be systematically mapped onto such

mental features as *being about Vienna* or *being of the form '. . . that God is good'* or *being of the form '. . . in order to deflect the stone'*. It seems reasonable to suppose—and these days *is* widely supposed[1]—that many different kinds of cerebral event might serve in the brains of different people, or even at different times in the brain of one person, as the physical correlate of a thought about Vienna; and, if that is right, then a psychological theory that was isomorphic with some physical theory such as cerebral neurology could not have *being about Vienna* as a causally significant feature of some thoughts. The same is true for all the other representative features of thoughts.

Thus, Spinoza is pushed into denying the causal efficacy of the representative features of thoughts by his physics together with his parallelism thesis. He may also be pulled toward that denial by a certain advantage he can get out of it. I was first put onto this point by C. L. Hardin, though the details of the handling are my own. It deserves a section to itself, but readers can jump to the start of section 5 without losing the main thread.

4. Why do we know no psychology?

The 2p7 parallelism doctrine says that true psychology maps onto true physics, right across the board, and human psychology is just the mental special case corresponding to the physical special case of the human brain. Spinoza thought he knew some general physics but must have realized that he had no inkling of a corresponding psychology. This asymmetry makes itself felt in the *Ethics* through repeated reminders that for Spinoza the physical realm calls the tune, as is strikingly evident in the structure of Part 2 'On the Nature and Origin of the Mind'. The opening propositions of this say only that the realm of thought is systematically correlated with the physical realm without interacting with it, and then 2p11 through 2p13 say that a person's mind and body are instances of this correlation. Before giving detail about the human mind, however, Spinoza breaks off at 2p13 and inserts a physics and a biology; then he returns to the topic of the human mind in 2p14, which says that the more versatile your body is, the more sensitive and capable your mind will be. As I said, the body calls the tune.

Still, Spinoza does not say outright that we have more access to the physical realm than to the mental; still less does he explain why. He ought to see this as a problem. It is understandable that the thick detail of human psychology defeats us by its complexity, matching the complexity of the still unknown fine structure of the human brain; but Spinoza should be troubled by our possessing some general physics but no corresponding general psychology.

Although this asymmetry must be an embarrassment to him, there are two things he can do to reduce it a little, making our ignorance of true basic psychology look less blankly contingent than is, say, my ignorance of economic geography or yours of New Zealand poetry.

The first is to suppose that we do not yet know how to classify mental items in a manner suitable for psychological theory: We have no glimmerings of true psychology for about the same reason that someone could have no glimmerings of true chemistry if he tried to found his chemical theory on the categories 'dirt', 'rock', 'liquid', and 'greenery'. (For a fuller presentation of this general line of thought, see

[1] See, for example, Jaegwon Kim, 'Causality, Identity, and Supervenience in the Mind-Body Problem', in Peter A. French, Theodore E. Euhling, Jr., and Howard K. Wettstein, eds., *Midwest Studies in Philosophy* 4, Studies in Metaphysics (University of Minnesota Press, 1979), pp. 31–49, at pp. 38–39; and David Lewis, 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50 (1972), pp. 249–58, at p. 256.

Thomas Nagel's intensely Spinozist paper, 'Panpsychism', in his *Mortal Questions* (1979), pp. 181–95.)

What is wrong with our present taxonomy of the mental? Does it just happen not to carve up mental reality at the joints, or has it some general feature that positively disqualifies it for scientific use? Spinoza, who hates brute facts, would prefer the latter option, and there is something he could say in support of it—this being the second of the two discomfort-reducing moves I said he could make.

It consists in the observation that we know almost nothing about our thoughts except for their representative features (i.e. what they are like *objective*), whereas what determines their causal powers, and thus what matters for science, is their intrinsic nature (i.e. what they are like *formaliter*). That is my main point in this section: If Spinoza does hold that mental items owe none of their causal powers to their representative features, i.e. to what they are 'of' or 'about', that can help him to explain why, although we do have physics, we do not have psychology. Of course, the explanation is incomplete, since it leaves unexplained our ignorance of the intrinsic natures of our thoughts, but it is better than nothing.

You might think that it is wrong because we really know a lot about the intrinsic natures of many mental items: the intensity of sensations, the subjective 'what-is-it-like' quality of perceptual states, and so on. There is something in that, but Spinoza would attach little weight to it. His picture of our mental life is a severely intellectualist one: it is dominated by our 'ideas', which tend to be beliefs but which, even when they fall short of that, are always propositional thoughts (2p49s). It is surely true that we know almost nothing about the intrinsic nature of that kind of mental item. I am now consciously entertaining the thought that I will be in New York City on Friday: that is a particular episode in the history of my mind, but if I were asked to describe it, what could I say? I could say when and 'where' it happened and what else I was thinking and experiencing at the time, but that would be all I could report, except for the episode's content, its being a thought that I will be in New York City on Friday. And so it is in general with mental items that have content: we know virtually nothing about them except their content.

There is just one feature of mental states that is intrinsic and that Spinoza does not snub. When my physical health improves, so correspondingly does my mental health, and the latter improvement is—Spinoza seems to hold—a feeling of pleasure. He has a general theory about pleasure and unpleasure, which he holds to be the mental sides of changes in physical health and level of vitality, and he apparently thinks there are causal laws about such kinds of feeling. That would be a start on a true basic psychology, but what a small one! It is on a par with —and indeed is really the mental counterpart of—the fragment of the science of biology that you get just from being able to tell whether a given organism is becoming more or less ill. In Spinoza's theory about feelings, all the fine detail is given in terms of the beliefs that cause the feelings, and that puts the theory out of touch with true basic psychology. But this is an aside within an aside; it is time to rejoin the main path.

5. The rejection of human teleology

Whether or not his reasons were as I have conjectured in section 3, Spinoza does try to avoid teleological notions in his initial account of the human condition. He allows himself the term 'desire', defined as 'appetite with consciousness thereof'; but according to him my appetite for P's being the case amounts merely to those intrinsic features of me that cause me to act in ways that make P more likely to become the case. (That, anyway, is the best I can make of what he

says in 3p9s, 3p56d, and the first of the affect definitions at the end of part 3.) Suppose that I am in a physical state S and a corresponding mental state S* and that S causes me to move in ways that increase the chance of my eating an apple; that fact can be expressed by calling S my 'appetite' to eat an apple and by calling S and S* together my 'desire' for an apple. But a proper (i.e. causal) explanation of my movements or their mental counterparts will refer only to S or S* itself; it will adduce only the intrinsic features of the relevant physical or mental item; the item's being an appetite or desire for an apple is not an intrinsic fact about it and so has no explanatory role. In short, Spinoza is resolutely refusing to let a possible effect of x have a working role in the explanation of x.

That is a pity because his account of the human condition is distorted and cramped by his refusal to allow anything teleological and his consequent inability to wield a sound concept of intention or purpose or goal. And there was no need for this, since one does not have to 'reverse the order of nature' or engage in any other malpractice in order to explain an item in a manner that essentially involves mentioning items that are temporally and causally subsequent to it. In section 6 I will show how this is done. Spinoza might not have liked the procedure in question, but his own basic principles give him no reason to reject it.

There is another reason for bringing a sound theory of teleology to bear upon our study of Spinoza and comparing it with his substitute theory of 'appetite'. When the text of the *Ethics* is examined in the light of that comparison, we can explain a profoundly puzzling feature of that work. Spinoza says that teleological explanations are always improper; yet he attributes to organisms a drive—he calls it conatus—that in his hands becomes a principle of self-interest. But to be self-interested is to have a certain kind of goal or purpose, which is the whole essence of teleology or 'final causes'. What on earth is going on here? In sections 9 and 10 I will answer this question.

6. Sketch of a theory of teleology

A full account of teleological explanation is a lengthy affair, which I have presented elsewhere.[1] Here I will keep it brief. The crucial notion is that of an instrumental property, that is, a property attributed to x by a proposition of the form:

> x is so situated and constructed that: *if Fx soon, then Gx thereafter*.

In shorthand, I put this by saying that F/Gx or that x has the instrumental property F/G: the animal is kills/eats, the pane of glass is dropped/shatters, and so on. Instrumental properties are not in themselves teleological. But now suppose that there is an organism x and a property G such that for *any* property f and time t,

> If f/Gx at t, then fx at t+d

—that is, whenever x is does-something/becomes-G, it does the 'something'. If Gx is 'x eats', then x is an organism that does whatever leads to its eating: when it is kills/eats, it kills; when it is climbs/eats, it climbs; and so on. This would be an animal that has becoming-G as a goal, and its G-seeking conduct could be explained in those terms. Why did it do F at time T? Because then it was F/G—which is to say that at T the animal was so constructed and situated that if it did F shortly thereafter that would lead to its becoming G a little later still. That is how an event can be explained with

[1] J. Bennett, *Linguistic Behaviour* (Cambridge University Press, 1976), ch. 2. The account presented there is developed from a basic idea—which would have sufficed for my main purposes in this present paper—in Charles Taylor's *The Explanation of Behaviour* (London: Routledge and Kegan Paul, 1964).

help from a mention of a possible event which, if it became actual, would be causally and temporally subsequent to the explained event. This explanation is not causal; i.e. it is not a matter of mechanistic, efficient causation. But neither is it a rival to mechanistic causation: There is no reason why each movement that is explained in terms of a goal should not also be mechanistically explicable in terms of the animal's intrinsic states at the time.[1]

There are no interesting cases of organisms conforming to a teleological law of the kind I have given. Any non-trivial teleological law must be restricted to values of f that in some sense belong to the repertoire of the given organism; and there must be allowance for multiple goals, for the animal's being prevented from doing F, and so on. I will pretend that all of that has been silently built into the account.

7. The cognitive complication

One complication must be treated explicitly, however. We do not expect any actual organism to do whatever *will* make it become G but only what *it thinks will* do so. We can always fix things so that an animal's doing F is the route to its becoming G (for example, walking clockwise in a circle for nine minutes is a way to get food, because we have arbitrarily chosen that reward for that performance), but we do not expect that to affect its behavior unless the relevant instrumental fact is registered upon the animal. I use 'registration' to name a genus of which belief is a vaguely demarcated species. And what I am saying is that a true teleological generalization would almost certainly have to be not of the form:

If f/Gx at t, then fx at t+d,

but rather of the form:

If at t x registers that f/Gx at t, then fx at t+d.

That, I submit, is the fundamental source of the famous interplay between belief and desire. They are well known to be intimately tied together, at least through the formulae:

His behavior shows what he wants, if you know what he thinks;

His behavior shows what he thinks, if you know what he wants.

The ultimate source of that link is the theory of teleological explanation: The concept of *desire* comes from that of *goal*, which is defined by the teleological patterns to which the animal dependably conforms; and the basic use of the concept of *belief*—I submit—is in the antecedents of teleological generalizations, with all its other uses depending on that. Properly to explain what beliefs are, you must start with their role in the pursuit of goals.

Spinoza's rejection of teleology, therefore, deprived him of a well-grounded concept of belief. His theory of cognitive content was bound to be thin and inadequate, since he did not have the teleological context within which to launch a good theory.

Still, the cognitive element in teleological explanations is not central to my present theme: Spinoza's basic objection to teleology can be stated and rebutted without reference to anything cognitive, namely, as an objection to explaining an item with help from the concept of a possible effect of that item. This is countered by showing how teleological explanations can have that feature without thereby being guilty of misconduct, e.g. 'reversing the order of nature' by treating effects as causes. The crucial idea is that of an organism's being so constructed that it can be depended upon to do whatever will make it G later.

[1] See D. C. Dennett, 'Intentional Systems', in his *Brainstorms* (Montgomery: Bradford Books, 1978); Bennett, *op. cit.*, section 21.

8. Distinguishing the conatus from teleology

From 3p4, which says that no organism can possibly destroy itself, Spinoza infers his conatus doctrine, his thesis about universal self-interest, according to which from each human's nature 'there necessarily follow those things which are conducive to his preservation' (3p9s).[1] In this section, I will try to provide a clearer view of the gap between those two.

There are two elements in it. One is the difference between '. . . does not destroy x' and '. . . is conducive to x's survival'. This of no great moment. If we take 3p4 as saying:

> If x does f, then the doing of f does not destroy x,

we could allow Spinoza to strengthen that to:

> If x does f, then the doing of f does not tend toward x's destruction,

and it would be intelligible, though wrong, to equate that with:

> If x does f, then the doing of f tends toward x's preservation.

I believe that those moves are at work early in Part 3 of the *Ethics* and that we can usefully see Spinoza as having taken his no-self-destruction thesis (3p4) to imply that whatever an organism does is helpful to it in the sense of being conducive to its survival.

I will now take the conatus doctrine in that form of it. That lets us focus on the second difference, which is more interesting. We now have the strengthened conatus doctrine saying:

> If he does it, it helps him,

and we have the remark in 3p9s, about each person doing 'those things which are conducive to his preservation', which says in effect that:

> If it would help him, he does it.

Of these, the former involves Spinoza's concept of 'appetite' or 'desire': the man is so constructed that what he does will tend to produce such-and-such results. The latter is genuinely teleological: the man is so constructed that, if something would tend to produce a certain result, he does it.

It is the difference between a conditional and its converse, and it is enormous. From the teleological statement we can infer positive predictions of behavior: The man *will do* F because that will help him. In contrast, the strengthened conatus doctrine supports only negative predictions: The man *won't do* F because that will not help him. Similarly, the teleological statement can explain why the man did such-and-such, while the other can only explain why he did not do so-and-so. The formal source of these differences is plain to see: The teleological statement has behavior in its consequent, whereas the conatus doctrine has it in the antecedent and can move it over only through contraposition:

> If it wouldn't help him, he doesn't do it.

It is because the consequent of that is negative that the conatus doctrine lets us predict and explain only negative facts about behavior.

There is no distinction between positive and negative facts in general. But I have recently established[2] a firm grounding for a distinction between positive and negative facts about the movements of a single individual, which is all I need here.

---

[1] Although in working on the *Ethics* I rely heavily on Curley's forthcoming translation, in quoting from the work I sometimes depart from Curley, and I take responsibility for all renderings. My 'are conducive to' renders the Latin *inserviunt*, which literally means 'are in the service of' or 'are devoted to' or 'are serviceable to'. Curley's 'promote' conveys the same idea.

[2] J. Bennett, 'Killing and Letting Die', in S. M. McMurrin (ed.), *The Tanner Lectures*, Vol. II (Salt Lake City: University of Utah Press, 1981).

9. How the two appear in Spinoza's text

I have been making much of the difference between 'If he does it, it helps him' and 'If it would help him, he does it'. It is undeniable that the former of these is the most that could with any semblance of validity be squeezed out of the thesis (3p4) that nothing can destroy itself; and I take that for granted. What about the other, teleological statement? I have equated it with what Spinoza says in 3p9s, namely, that there necessarily follow from each person's nature 'those things which are conducive to his preservation'. The equation would have failed if Spinoza had said: 'From a man's nature there necessarily follow things that are conducive to his preservation; and so he is determined to do such things.' That might mean no more than that the man is sometimes caused by his nature to do things that are helpful to him and could not mean more than that whatever he does is helpful to him. What Spinoza says, however, is not that but rather: 'From a man's nature there necessarily follow *those* things which are conducive to his preservation; and so he is determined to do *those same things*'; and I cannot see how to avoid taking that to mean literally that he does *all* the helpful things,[1] which makes the statement a conditional running in the teleological direction, saying that if something would help the man, then he does it.

Am I exaggerating tiny nuances in insisting that Spinoza's 'those (same)' (*ea* and *eadem*) be read as 'all those'? Well, I need not rest anything on that one sentence in 3p9s, for later in part 3 Spinoza uninhibitedly employs conditionals of the teleological sort. No fewer than eleven propositions imply that the conatus doctrine predicts what people will do in certain circumstances. In fact, Spinoza always speaks of what the person will *try* to do; and, as I will explain shortly, the word 'try' is essentially teleological. But never mind that just now. My present point is that in the propositions 3p12 and 13 and nine others derived from those two Spinoza accepts conditionals with the actual or attempted behavior in the consequent. These propositions say 'If. . . ., we try. . . .' and not 'Only if. . . do we try. . . ', and they say 'We try to do whatever. . . ' and not 'We try to do only what. . . '. Furthermore, in parts 4 and 5 Spinoza is clearly relying on a doctrine of self-interest that is openly teleological and predictive of behavior.

In the paper mentioned in section 1, Parkinson adopts my view about what a good teleological explanation looks like, and comments: 'Bennett (*Linguistic Behaviour*, p. 41) thinks that his theory is an *answer* to Spinoza's views. But as has already been argued (n. 11) that Bennett misunderstands Spinoza's views about final causation.'[2] I do agree that Spinoza ends up saying things that are teleological in the way that Parkinson and I both accept, and I don't think I was clear about that until I returned to Spinoza after studying teleology. But it does not make much difference. These genuinely teleological things that Spinoza says do fall within the scope of his challenge to teleology, and he does nothing to argue that they do not, i.e. nothing to clear himself of the charge of 'reversing the order of nature'. In saying this, I cannot be reaping the fruit of that 'misunderstanding of Spinoza's views about final causation' with which Parkinson charges me: That concerned why Spinoza thought that 'He raised his hand so as to deflect the stone' puts Raise and Deflect in the wrong order; it did not affect my view about *what* the objectionable order is, and that is all I need to make

[1] Interestingly, Boyle's translation actually uses the word 'all' in rendering the passage: '. . . from the nature of which all things which help in [his] preservation necessarily follow'.

[2] Parkinson, *op. cit.*, p. 8 (n. 15).

the challenge apply to Spinoza's own teleological statements. Parkinson apparently construes the challenge as a very limited affair amounting to nothing but the point that Deflect did not cause Raise. There is very little actual teleological talk that would not be fully acceptable to Parkinson's Spinoza: He would serenely accept Braithwaite's proposal (reported in section 2) and would be under no strain in accepting a conatus doctrine that implies that, if somebody thinks that doing F would help him, then he will do F.

Well, I submit that my tougher and less consistent Spinoza is more interesting and deeper. He is also the actual Spinoza. If Parkinson were right, Spinoza would not have needed to insist upon his special notion of appetite, which is so resolutely unteleological. (Remember that to have an appetite for P's being the case is not to be disposed to do whatever will make P the case; it is merely to be in a condition in which one's behavior is apt to make P the case.) Also, I will show in my next section that Spinoza's route to his teleological conatus doctrine, in the pages of the *Ethics*, is a sequence of invalidities. I take these as evidence that Spinoza is in trouble here: he is trying to arrive at something that he has implicitly forbidden to everyone; and so it has to be developed in a twisted, tangled, illegitimate manner. I do not, of course, mean that Spinoza knew that that is what he was doing.

Having responded to Parkinson, I should say that this conflict between us affects only a tiny part of his admirable paper, whose principal aim is to set forth Spinoza's use in his moral philosophy of the doctrine of conatus, which he actually has, i.e. the teleological one to which Parkinson does and I do not think he is entitled. The question of entitlement is marginal to Parkinson's concerns, which presumably explains his not inquiring at all into the provenance of the teleological conatus doctrine in Spinoza's text and thus having nothing to say about the tissue of invalidities to which I now turn.

10. How did the mistake occur?

The word 'conatus' is Latin for 'trying'. And, properly speaking, 'trying' is always a matter of *trying to do x* or *trying to bring it about that P*; that involves behavior that is explained by one's thinking it may have a certain result, which is teleological and involves explaining what happens at one time by reference to what might happen later. So, Spinoza's very choice of name for the doctrine in question suggests that he has been covertly thinking of it as teleological right from the outset, and so he has. Although it is not until 3p12 and 3p13 that we see the teleological conditional openly at work, the basic malfeasance occurs in the moves from 3p4 to 3p6 which the conatus doctrine is originally announced. I will explain how.

Spinoza's argument for his conatus doctrine starts with the no-self-destruction thesis: 'No thing can be destroyed except through an external cause' (3p4). Never mind where that comes from. Our present concern is with what Spinoza infers from it, namely: 'To the extent that one thing can destroy another, they are of a contrary nature, i.e. they cannot be in the same subject' (3p5). There are two ways of taking this: **(1)** It could be saying that if one *thing* can destroy another then they could not both 'be in the same subject'—presumably this means that they could not be parts of a single organism. **(2)** It could be saying that if one *property* can destroy another—presumably this means that a thing's acquiring one would cause it to lose the other—then nothing could instantiate both properties at once.

Of these two readings, **(1)** is favored by Spinoza's use of 'thing' in the proposition and by the idea of one item's 'destroying' the other, but **(2)** is favored by the phrase 'be

in the same subject'. With regard to the credentials of the proposition, the two readings are about on a par: Neither is entailed by 3p4, though each is encouraged by it; and on each reading the proposition is in some danger of being trivially true. Nor do the five subsequent uses of 3p6 resolve the ambiguity. Two of them clearly favor reading **(1)**, two others clearly favor reading **(2)**, while the remaining one is perfectly neutral between them![1] Fortunately, this neutral use is 3p6d, the demonstration of the conatus doctrine, so we can examine how this makes use of 3p5 without having to resolve the latter's ambiguity. This I now do.

The most Spinoza has any claim to be saying in 3p5 is that, if x can destroy y, then they are 'contrary' in the sense that they cannot coexist; i.e. they are **(1)** things that a single organism cannot have as parts or **(2)** properties that a single thing cannot instantiate. In 3p6d, however, Spinoza takes himself to have meant some thing very different from this, namely, that, if x can destroy y, then y is 'opposed' to x in the sense that it will exert itself to reduce the threat from x. This lavish over-interpretation of 3p5 is expressed in Spinoza's concluding that each thing 'tries to persevere in its being', i.e. tries to stay in existence. This has to mean that each thing acts against threats, which goes far beyond 3p5's assertion that if one item threatens another then they are incapable of a kind of coexistence.

Although the two are not simply a conditional and its converse, there is an element of that in the difference between them. Given that x can destroy y, all that 3p5 as originally offered says about their behavior is a conditional with behavior in its antecedent:

(1) For any f, if y does f, then the doing of f will not result in y's coexisting with x.

But what Spinoza makes of this in 3p6d is a conditional with behavior in its consequent:

(2) For any f, if the doing of f would tend to keep y safe from x, then y will do f.

In this analysis I am not relying just on the phrase 'tries to persevere'. Further support is given by Spinoza's saying at the end of the demonstration, though not in the official proposition at the head of it, that each thing, as far as it can, tries to persevere in its being. It is easy to fit 'as far as it can' into **(2)**, the teleological conditional: If P, then y will do f as far as it can. But there is no plausible way of fitting it into **(1)**, the other conditional, the official 3p5 one that has behavior only in the antecedent. The only grammatically possible place for it yields the result: 'If y does f as far as it can, then. . . ', which makes philosophical nonsense.

Having thus invalidly brought something teleological into his doctrinal structure, Spinoza immediately proceeds to deny that he has done any such thing. In 3p7 he says that the so-called conatus. or trying, 'is nothing but the actual essence of thing', and this, properly understood, is an important disclaimer. Although it does not use the term 'appetite', it amounts to the claim that the apparently teleological term 'conatus' really stands only for austere Spinozist appetite: In attributing a self-preserving conatus to an organism, he wants us to believe, we are saying only that it has a nature that will cause it to behave in self-preserving ways. He is not entitled to this. Granted, the basic causal story concerns the organism's intrinsic nature or 'essence', but that is not the whole explanatory story; for Spinoza has also said that the organism will 'try as far as it can to preserve itself', and nothing can save this from meaning something teleological. Later on in the *Ethics*, indeed, he stops even

[1] Favoring reading **(1)** are 3p10d and 4p30d; favoring **(2)** are 3p37d and 4p7d.

gesturing toward Spinozistic 'appetite' in preference to real goals and purposes, or toward 'If he does it, it helps him' in preference to 'If it would help him, he does it'. Apparently, he thinks that the sequence 3p4 through 3p7, has entitled him to a conatus doctrine that will do teleological work for him without being open to his own objections to teleology.

11. Could Spinoza have made such a mistake?

Failing to distinguish a conditional from its converse is a bad mistake. So is confusing the sense of 'contrary' that operates in 3p5 with the sense of 'opposed' that is needed for 3p6d. Some of Spinoza's admirers will think that these mistakes are so bad that he cannot have been guilty of making them. I see that attitude toward Spinoza as a solid obstacle to understanding his work. As I hope my new book[1] will show, to learn a lot from the *Ethics* one needs a firm general idea of what kinds of help it can give and what kinds it cannot; and that requires a just appreciation of Spinoza's own strengths and weaknesses. Above all, it has to be understood that Spinoza's mind was strong, deep, wide-ranging, tough, brave, and original but not quick and not sharp. Leibniz's kind of nimble acuity was altogether foreign to Spinoza, and Leibniz himself is on record with a wry comment about the invalidity of some of Spinoza's demonstrations.[2]

Often, the trouble is merely expository: Spinoza would assert a conditional when he meant a biconditional or label as a 'definition' (to be read left to right) a biconditional that turns out to be a substantive thesis that can be used in argument from right to left. Sometimes, however, it is not bad writing but error. When, for example, Spinoza moves from:

If (x resembles y, and *x thinks that* Fy) then Fx

to

If (*x thinks that* x resembles y, and Fy) then Fx,

as he demonstrably does when he purports to rely on 3p27 in 4p68s, this is incompetence. Spinoza was a genius and one of the most challenging and instructive philosophers who ever wrote, but there is a certain kind of logical competence that he lacked—falling short not only of Leibniz, who is supreme in this respect, but also of the other major figures in early modern philosophy.

With that said, I should add that the conditional conversion that is my present topic is not as blundering as my diagnosis has made it appear. One of the hazards of philosophical debate is that in philosophy, unlike some other disciplines, the best techniques for exposing error tend to make the error look elementary and its perpetrator stupid. We show something to be wrong by boiling it down to some patent absurdity, and we may tend to forget that the absurdity was perpetrated in a thick, difficult context, not in the extracted form in which we expose it. So it is with Spinoza's switch from conatus to teleology. (I certainly hope so. If the error was gross, then we who care about Spinoza's thought must be correspondingly dense. It took me more than twenty years to discover what had gone wrong in the conatus doctrine, and others seem to have been even slower.) Although it is true that the core mistake is the conversion of a conditional, generated by malpractice with 'contrary' and 'opposed', these mistakes are disguised—rendered easy to make and hard to discover—by the complex philosophical context in which they occur.

[1] J. Bennett, *A Study of Spinoza's Ethics* (forthcoming).

[2] G. W. Leibniz, *Sämtliche Schriften und Briefe* (published by the Berlin Akademie Verlag), Series 2, Vol. 1, pp. 379f.

12. The context of the error

I will sketch the context of Spinoza's mistake about conatus and teleology, trying to show that my analysis of the situation is not greatly to his discredit. This is not to persuade myself that he deserves my admiration, for I have never begun to doubt that. Nor is it to defend his reputation in the minds of others, for that is too well established to be affected by anything I might say. The point is just that one's settled admiration for a philosopher can naturally affect what interpretations of him one is prepared to consider seriously, and I want to persuade Spinoza's admirers not to shut their minds against my analysis of his muddle about conatus and teleology.

The context of Spinoza's conatus-teleology muddle is created by the intersection of three big thrusts in his thought. Let us look at them one by one.

First, having had the insight to see the prima facie problem involved in our ordinary notions of goal or purpose or end, and not seeing how to solve it, Spinoza concluded that these notions—in the form in which the common person has them—must be jettisoned. Never forget that philosophers who retained these notions had that advantage because they had seen less than Spinoza did, not more, as they overlooked the problem rather than seeing the solution.

Second, Spinoza thought it to be a universal truth that humans are always self-interested. There is plenty of evidence of egoism, and Spinoza was not one to shrink from taking a widespread tendency to be a universal truth. He will have found further support in the use he could make of psychological egoism in his moral theory. Projecting from his own character and attitudes, as all sincere moral philosophers do, he wanted a moral system with a coolly unsentimental input and a morally upright and even noble output; and he thought he could achieve this remarkable result if he had egoism as his chief premise.

But the doctrine of egoism had to be freed from the taint of teleology; and Spinoza also had a need, created by his intellectual temperament, for the egoism to be shown to be somehow necessary, deeply rooted in the nature of reality. (So indeed it is, but he was in no position to give the right, evolutionary reason why something like self-interest runs strongly through the behavioral patterns of all organisms.) This double need was satisfied, Spinoza thought, by what comes next.

Third, he discovered the argument stretching from 3p4d through 3p6d. Wrong as this argument is, its ingenuity should not be underestimated. It starts with the argument for the thesis (3p4) that nothing can, unaided, cause its own destruction: if something destroyed itself, it would have a nature that was causally (and thus, for Spinoza, logically) sufficient for its own nonexistence; but that would be an inconsistent nature, which nothing can possibly have (3p4d). The conclusion of the argument is false, as Spinoza might have come to suspect while wrestling with the fact of suicide (4p20s), and so, of course, the argument is faulty. Where it says that, if a thing could destroy itself, its nature would be logically sufficient for its non-existence, the argument ought to say only that if a thing could destroy itself, its nature *at one time* would be *causally* sufficient for its nonexistence *later*. But since Spinoza conflated logical with causal necessity and, associated with that, did not generally attach weight to temporal differences, he was not well placed to see anything wrong with 3p4d. That argument is thus not a merely perverse contrivance; on the contrary, it is just what Spinoza ought to say, given his causal/logical conflation and his inattention to temporal differences. Just as his rejection of teleology arose from a real insight, so his

no-self-destruction thesis arose from a brilliant exploitation of his own philosophical assumptions and attitudes.[1]

Next comes the move to 3p5. I will expound this on the basis of one resolution of the ambiguity noted in section 10; the whole story could, perhaps less plausibly, be reconstructed in terms of the other reading. The idea is that if nothing can destroy itself then nothing can have two parts one of which will destroy the other. It does not quite follow, since—as Spinoza well knew—the health of a whole may require occasional destruction or atrophy of some of its parts; but in the given context, with so much at stake, the mistake is a natural one. What needs more explaining is the final mistake in the sequence, namely, moving from the premise that if x can destroy y then they are 'contrary' in that sense, to the conclusion that if x can destroy y, then y is 'opposed' to x in the sense that it will make war on x, so to speak. Perhaps it is just a conflation of 'contrary' in one sense with 'opposed' in another, but there may be more to it than that. Here is a guess about what more there is.

We are to interpret 3p5 as applying to two things x and y, which are themselves 'individuals' but which are also fit to be parts of larger 'individuals', and Spinoza is taking the proposition to imply that if x can destroy y then they cannot both be parts of a single individual that is not vastly greater than either of them is. That stipulation about relative size is needed: The universe itself is an 'individual' in Spinoza's sense, and 3p5 must not imply that if x can destroy y then they cannot coexist in the same universe. And similar considerations apply at smaller sizes. For example, if x is a person who can destroy person y, they could still belong to the same universe and even the same nation; but Spinoza might say that their belonging to the same village or to the same family would tend toward creating the impossible situation of an individual (a village or a family) that could destroy itself without outside aid. Now, approaching 3p5 in that manner, Spinoza could reasonably take it to imply that if x could destroy y then they must always be at a distance from one another, because if they came too close they would threaten to unite within a single individual that could contain both only at the risk of being self-destructible. From that he might infer that y could be depended upon, if necessary, to keep x at a safe distance; and from that he might drift into thinking, in 3p6, that y could be relied upon to do whatever would reduce the threat from x—perhaps keeping it at arm's length but perhaps instead launching a preemptive strike against it. That would be a bad mistake, but it would have more structure to it than a mere confusion would.

My overall point is just that in the sequence 3p4 through 3p6 Spinoza is arguing intricately and ingeniously and is playing for high stakes. What is at issue is the establishment of a deeply rooted egoism that is no way teleological! In such a context, even a wonderful philosopher is likely to make bad mistakes.[2]

---

[1] One might think that Kant is criticizing this argument here: 'The principle that realities never logically conflict with each other is entirely true as regards the relation of concepts, but has no meaning in regard to nature. For real conflict does take place; there are cases where A + B = 0, that is, where two realities combined in one subject cancel one another's effects' (*Critique of Pure Reason* A 273, quoted with omissions and one correction [minus changed to plus]). But, although this scores a direct hit on Spinoza's 3p4d and 3p5d, its intended target is Leibniz, whom it misses. See G. H. R. Parkinson's helpful paper, 'Kant as a Critic of Leibniz', *Revue Internationale de Philosophie* 136–37 (1981), pp. 302–14, at pp. 310ff.

[2] I am indebted for good help with this paper to my colleagues William P. Alston and C. L. Hardin.