

Chapter 24

Teleological Explanation

SCOTT SEHON

A teleological explanation of human behavior explains the behavior by citing a state of affairs or goal towards which was directed. Canonical teleological explanations of action comprise mention of an agent, the agent's behavior, a teleological connective, and the state of affairs towards which the behavior was directed. For example,

Jane went to the kitchen in order to get coffee.

In this example, the connective is the phrase "in order to," but explanations can be explicitly teleological with other connectives as well. For example,

Jane went to the kitchen for the purpose of getting coffee.

Jane went to the kitchen to get coffee.

Jane went to the kitchen with the aim of getting coffee.

These teleological explanations of behavior are paradigm cases of explaining an agent's action in terms of her *reasons*, and thus count as reason explanations of behavior.

Like the examples above, many reason explanations of human behavior are in explicitly teleological form. Others are not in explicitly teleological form:

- (1) Jane went to the kitchen because she wanted coffee.
- (2) Hayden shucked corn because his father told him to.
- (3) Josephine went upstairs because her shoes were there.

According to some authors, such explanations should nonetheless be construed teleologically.

That is to say, the explanations can be read as claiming,

- (1a) Jane went to the kitchen in order to satisfy her desire coffee.
- (2a) Hayden shucked corn in order to fulfill his father's request that he do so.
- (3a) Josephine went upstairs in order to get her shoes.

Most philosophers would agree that many common sense explanations of human action, are already in teleological form, and that others, like (1)-(3) can be plausibly construed teleologically, as in (1a)-(3a). After all, acting for reasons or purposes is part of what makes us agents, and teleological explanations capture that directly.

The true controversy concerns whether action explanation is *irreducibly* teleological. Many philosophers of mind would claim that talk of *goal direction* and *purpose* is a useful heuristic, but it is ultimately to be cashed out in other terms. In the first section, I briefly discuss some of these attempts to reduce teleology to other forms of explanation, and mention some of the problems some of these approaches face. In the next section, I turn to positive accounts of teleology, according to which teleological explanations are not reducible to other forms of explanation. In the final section, I consider objections to the view that action explanation is irreducibly teleological.

Reductionist accounts of teleology

Prior to Darwin, purpose seemed to pervade the physical world, even beyond the goal directed behavior of human beings and beyond designed artifacts like hammers. Clearly, it seemed, bird wings were designed for flight, eyes were designed for sight, etc. When asking about the purpose of a hammer, we naturally assumed that some agent *designed* the hammer for some end. Similarly with bird wings and the like, it was a natural thought, even if not inevitable, that there must be a designer. Whatever the merits of these thoughts, prior to Darwin, as an argument for the existence of God, certainly the scene changed considerably after Darwin. As

Darwinian biology shows, wings that fly and eyes that see can arise through a process of reproduction with random variation and natural selection. We can thus explain the prevalence of biological adaptations, and there is no need for a designer with conscious (or unconscious) purposes.

With reason explanation, the item to be explained is individual behavior, rather than a thing like a wing or an eye. But there is nonetheless a similarity. Consider, for example, an artifact like a hammer or a stapler. Just as these artifacts are designed for a particular purpose, one might think of particular bits of human or animal behavior as designed by the agent for its goal. Just as some agent first fashioned a hammer for striking nails, Jane designed her walking for the end of getting coffee. But here too, at least in the case of certain animal behaviors, Darwinian explanation can eliminate the need for genuine talk of design. Why do cats arch their backs in the presence of a perceived threat? There is a selectional explanation. When cats arch their backs, they appear larger and more threatening, and this sometimes scares away potential predators. So cats with that behavioral disposition survived longer and had a higher differential reproductive rate, and thus the disposition came to dominate the population. Why did Fluffy, in particular, arch her back when Rover approached? Well, evolution explains her having the general disposition to do that in circumstances of the appropriate type, and these are circumstances of the appropriate type. Of course many details would need to be worked out. But the basic idea is broadly Reductionist, for the goal directed behavior exhibited by Fluffy can be cashed out in evolutionary terms, and Fluffy's role as the conscious agent or designer of the behavior drops out. The realm of *sui generis* or genuine teleological explanation seems to be much narrower than might have been originally thought. Where we once had apparently irreducibly teleological explanation of birds' wings and all sorts of animal behavior, the realm of *sui generis* teleological explanation now shrinks to the more idiosyncratic and intelligent behavior of human beings and perhaps other higher level rational animals.

But now the question becomes: can this evolutionary approach be applied even to human behavior? Human beings are certainly part of the evolutionary story, and we explain the fact that our eyes are good at seeing in the same evolutionary way that we explain that bird wings are good for flying. And we can likewise give selectional explanations of many of our general behavioral tendencies—from our disposition towards sweet foods, to more complex dispositions like our practice of burying or burning human corpses. One might think that we will be able to follow the evolutionary line of thought even for ordinary actions like Jane's going to the kitchen for coffee. If so, this would take the apparently *sui generis* teleology concerning the behavior of animate agents and would reduce it to purely naturalistic explanation in terms variation and natural selection.

There are obstacles to such a reductionist approach. When we explained the arching of Fluffy's back, the evolutionary explanation cited a disposition that led, in Fluffy's ancestors, to higher differential reproduction. A great many human actions seem, at best, neutral towards differential reproduction: writing philosophy papers, arranging to meet a friend, shoveling snow from the driveway, etc. And some of our actions, even seemingly rational actions, appear to be quite inimical to differential reproduction: pledging celibacy and using birth control are obvious examples. So if the goal directedness of human behaviors is to be cashed out in the language of natural selection, there will be many challenges, and it is, to say the least, controversial whether those challenges can be met. For attempts that are at least roughly in this direction see Wright (1976), Millikan (1984), and, more recently, Okrent (2007).

Other authors attempt to reduce teleology not by tying it directly to evolution, but by giving a more straightforward causal reduction. This tradition begins with Donald Davidson's classic article (1963), "Actions, Reasons, and Causes," in which he maintains that standard reason explanations are to be construed as a form of causal explanation in which appropriate mental

states of the agent are cited as the effective cause of the behavior. So, for example, an explanation like

Jane went to the kitchen in order to get coffee

would be construed as saying something like this:

Jane's desire for coffee, and her belief that coffee could be obtained in the kitchen, caused her going to the kitchen.

Alternatively, one might want to assert that intentions are the causally effective mental state. (See, for example, Mele (1992) for a detailed working out of a broadly Davidsonian approach; for an approach that is still causal, but is within the agent-causation rather than event-causation tradition, see O'Connor (2000).)

On the causalist approach, the apparently distinctive teleological element in reason explanations drops out. Reason explanations, even when in superficially teleological form, are ultimately a species of causal explanation. To say that an agent does something for the sake of some goal just is to say that an appropriate mental state caused the behavior. But this can't be quite right as it stands. Consider an example from Davidson:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. (Davidson 1973, p. 79)

Here it seems that the agent's desire to let go of the rope caused him to do just that. But we don't want to say that the agent let go of the rope intentionally; we don't want to say that this was goal directed behavior. Thus, something more needs to be added to the causalist analysis. The causalist might reply that in the case of the mountain climber, the causal chain was deviant in

some way, and that genuinely goal directed behavior requires that the mental state cause the behavior in the right way.

But now the challenge for the causalist is to spell out “in the right way” in purely causal terms. This is the problem of deviant causal chains. It has been pressed upon the causalist most forcefully in Wilson (1989); for representative replies see Mele (2000), Peacocke (1979), and Bishop (1989); for counters see Sehon (2005). Beyond the problem of deviant causal chains, other arguments have been raised against the causalist reduction of teleology. See, for example, Okrent (2007), Sehon (2005), and Schueler (2003).

Nonreductionist accounts

Among recent authors, Wilson (1989), Schueler (2003), and Sehon (2005) are most explicit in taking action explanation to be irreducibly teleological. (There are other authors, e.g., Anscombe (1957), Stoutland (1976), Dancy (2000), Tanney (2009), who suggest that action explanation is not causal without quite saying that action explanation is irreducibly teleological.) On each of the teleological accounts, reason explanations, even when not in explicitly teleological form, work by citing a goal state as the crucial explanatory factor. In saying that Jane went to the kitchen to get coffee, we are saying that Jane's physical behavior was directed, by Jane, at the objective or goal of getting coffee. Two aspects of this form of explanation are of special note: the fact that the behavior is directed, and the mention of the agent herself as the director of the behavior. On the view that takes reason explanations to be irreducibly teleological, neither of these elements is eliminable. This is, of course, in marked contrast to the causalist view, where the goal-direction itself is cashed out in terms of a specific type of causal chain, and the causal factors are mental states.

On the teleological view, the fact that an agent is directing her behavior to some state of affairs is not grounded in or reducible to some further facts. Goal direction by agents is an ineliminable part of the universe. (Well, you could kill all the agents and thereby eliminate the phenomenon, but insofar as the phenomenon exists, it is not reducible to other features.) But to say that goal direction is ineliminable and in that sense *sui generis* is not to say that the concept is completely isolated from other concepts. Bedau (1992) noted that teleological explanations involve normative notions, and proponents of irreducible teleology explicitly tie teleology to the concepts of rationality and value.

Schueler (2003) claims that when we teleologically explain an agent's action by citing some goal, we assert that the agent assigned that purpose to the behavior as part of a plan. Schueler's account is modeled on the practical syllogism; when we attribute a goal to an agent, we are imputing to the agent a plan that could be the conclusion of a piece of practical reasoning.

More strongly yet, we are thereby attributing to the agent the normative judgment that her reason for acting represents that which she had “the strongest reason to do,” that which has “the most to be said for it.” (p. 136) Thus teleological explanation is normative, in that each teleological explanation attributes to the agent a fairly strong normative judgment.

In Sehon (2005), I offer a picture that is similar to Schueler’s in some ways but is less modeled on conscious practical reasoning or the relatively rare acts of deliberately assigning a purpose to some behavior. According to my account, since facts about teleology do not reduce to some other set of causal facts, one cannot give definitive truth conditions for teleological explanations in non-teleological terms. However, there is still much one can say about how we in fact make teleological explanations. Loosely following Davidson’s views of interpretation (but not his endorsement of the causal theory of action), I argue that we arrive at teleological explanations as part of an overall attempt to construct a theory of the agent, and a constitutive part of reason explanation is to produce a theory according to which the agent is as rational as possible. Rationality can be judged in different ways, but two axes are particularly relevant to judging a candidate teleological explanation: the degree to which the explanation makes the behavior appropriate for achieving the goal, and the degree to which the goal is of value. We aim to maximize both of these in forming our theory of the person.

In simple and straightforward cases, application of this is almost entirely automatic. Recall Josephine who went upstairs, and suppose that the circumstances were these: another ten year old friend had shown up at the door and asks if Josephine wants to come outside to play; Josephine starts to run excitedly outside, but her father says, “wait! you need shoes! they’re upstairs”. And off she goes up the stairs. Here it is quite obvious that going upstairs would be appropriate to achieving the goal of getting her shoes, and we can easily understand the value that this would have for Josephine. We might also believe various counterfactual conditionals that point in the same direction: If Josephine had believed that her shoes were in the kitchen, she would have gone there instead of going upstairs; had she believed that she already had shoes on,

which perhaps she did prior to her father's admonition, she would have simply gone outside; etc. The general point is that our theory of Josephine is constructed so as to make the most rational sense we can out of her behavior in the actual and in nearby counterfactual circumstances.

Naturally, these rationalizing principles do not constrain our theorizing about the behavior of inanimate things like rocks or planets. Or, to put it the other way around, on any theory according to which a rock was an agent, the rock would either come out as quite irrational, or would have too impoverished a set of goals to count as a genuine agent. If we attribute to the rock one and only one desire, the desire to follow the laws of physics, then of course the rock comes out as always acting in ways appropriate to its one goal. But it is not clear why this one goal would be of value to a rock or anything else. On the view in question, being an agent requires a complex set of goals—a life. We cannot successfully attribute anything of the sort to the rock. So we conclude that the rock is not an agent at all, and no teleological explanation of its behavior will be true.

Prospects and consequences

Davidson (1963) famously raised a challenge for any non-causal account of action explanation. He pointed out that an agent's action might be justified by two distinct reasons, but that the agent in fact acted on only one of the reasons. For example, Jane walks downtown and thereby accomplishes two things of value: she gets out of the office for some fresh air, and she meets a friend as planned; but in fact she may have acted only for the latter reason. Davidson says that the causalist can easily make sense of this scenario: Jane's real reason for action is determined by whichever desire or intention caused her behavior. But the teleological account might seem to have more of a problem. How can the teleologist claim that Jane really acted simply in order to meet her friend, given that the goal of fresh air would also make good rational sense of her behavior?

The teleologist's answer lies in the fact alluded to above: that in seeking rationalizing explanations of the agent's behavior, we aim to make sense of the person both in the actual and nearby counterfactual circumstances. If an agent's behavior is truly directed at a goal, then we expect that, in nearby counterfactual circumstances, the agent would have done what it took to accomplish that goal. So, in the case of Jane, if she really walked downtown in order to meet her friend rather than to get some fresh air, we would expect that, had she believed she could meet the friend by staying in her office, that is what she would have done. In other words, different teleological explanations of the same behavior support different counterfactuals, and this will be enough to answer Davidson's challenge.

Even if the teleological account can meet Davidson's challenge, there is a perhaps more serious worry waiting in the wings, namely that the teleological account seems to run afoul of Occam's razor. There are many differences of opinion about the details, but it is fairly clear that our theorizing about the world is constrained by some version of the following principle:

- (S) Given two theories, it is unreasonable to believe one that leaves significantly more unexplained mysteries.

If teleological explanation is irreducible, then the teleologist has no account of why teleological explanations work, or why, more specifically, the behavior of certain systems is such that the rationalizing mode of teleological explanation works consistently over time. A reductionist, by contrast, simply reduces teleological explanations to some form of causal or selectional explanation, and thus has no further mysteries with teleologically explicable behavior. Thus, it would seem that the reductionist theory leaves fewer unexplained mysteries, and should therefore be seen as preferable to the teleological account.

In answer to this objection, let's first compare another case. The standard model in physics postulates the following:

(SP) Four fundamental and independent forces: strong force, weak force, electromagnetic force, and gravity.

Suppose Joe, a sophomore in Philosophy of Science, tells us that he thinks this is wrong and that there is just one force. I.e., he is an advocate of Joe's physics:

(JP) Just one force which suffices to explain the nature of the four 'subforces' as well as the relationships between them.

Should we just accept that (JP) is a simpler theory than (SP) and reject the standard model? Of course not. Joe has not actually provided a physical theory that works. All we know is that *if* Joe can come up with such a theory, then we would have a simplicity argument strongly in its favor.

Now compare two theories, the teleological account (T) and the reductionist account (Rd):

(T) Principles of physical science plus independent rationality principles underlying teleological explanation.

(Rd) Just the principles of physical science, which suffice to explain why the principles underlying teleological explanation hold.

Should we immediately conclude that (Rd) is simpler, and that we should reject (TR)? No. All we know is that *if* philosophers can come up with such a reductionist account, then there is a simplicity argument in its favor. And this is indeed motivation for seeking such an account. However, nonreductionists have argued against the likelihood of finding any such reductive account of teleology. And at least some of us would agree with George Wilson's words: "the evidence points to more than infelicity or incompleteness in the various causalist proposals—it points, that is, to a global breakdown in the whole project of reduction" (1989, p. 258). If there is no successful recipe for reduction, then there is no simplicity argument against nonreductionist views. It would be desirable to have a reduction of teleology in the same sense in which it would be desirable to have a physical theory that unifies the four forces. But if the world does not

cooperate, and no unifying theory accounts for the facts as we know them, then it is a mistake to affirm reductionism, and we must live with teleological explanation of human action as an ineliminable feature of the world.

See also THE CAUSAL THEORY OF ACTION (6), INTENTION (15), MENTAL CAUSATION AND EPIPHENOMENALISM (19), REASONS AND CAUSES (20), DEVIANT CAUSAL CHAINS (22), ACTION AND AGENCY IN EVOLUTIONARY PSYCHOLOGY (52), DAVIDSON (75).

References

Anscombe, G.: *Intention* (Ithaca, NY: Cornell University Press, 1957).

Bedau, M.: 'Goal-directed Systems and the Good', *The Monist* 75 (1992), 34-49.

Bishop, J.: *Natural Agency: An Essay on the Causal Theory of Action* (Cambridge: Cambridge University Press, 1989).

Dancy, J.: *Practical Reality* (Oxford: Oxford University Press, 2000).

Davidson, D.: 'Actions, Reasons and Causes', *Journal of Philosophy* 60 (1963), 685-699.

Davidson, D.: 'Freedom to act', In Honderich (ed) *Essays on Freedom of Action* (London: Routledge and Kegan Paul, 1973), pp. 137-156. Reprinted in Davidson (1980). Page citations from the latter.

Davidson, D.: *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).

Mele, A.: *Springs of Action: Understanding Intentional Behavior* (New York: Oxford University Press 1992)

Mele, A.: 'Goal-directed Action: Teleological Explanations, Causal theories, and Deviance', *Philosophical Perspectives* 14 (2000), 279-300.

Millikan, R.G.: *Language, Thought, and Other Biological Categories: New Foundations for Realism* (Cambridge, MA: MIT Press, 1984).

- O'Connor, T.: *Persons and Causes: The Metaphysics of Free Will* (New York: Oxford University Press, 2000).
- Okrent, M.: *Rational Animals: The Teleological Roots of Intentionality* (Athens, OH : Ohio University Press, 2007).
- Peacocke, C.: *Holistic Explanation: Action, Space, Interpretation* (New York: Oxford University Press, 1979)
- Schueler, G. F.: *Reasons and Purposes: Human Rationality and the Teleological Explanation of Action* (Oxford : Clarendon Press, 2003).
- Sehon, S.: *Teleological Realism: Mind, Agency, and Explanation* (Cambridge, MA: MIT Press, 2005).
- Stoutland, F.: 'The Causation of Behavior', in Hintikka (ed) *Essays on Wittgenstein in Honor of G.H. von Wright* (Amsterdam: North Hollandsche, 1976), pp. 286-325.
- Tanney, J.: 'Reasons as Non-Causal, Context-Placing Explanations', in Sandis (ed) *New Essays on the Explanation of Action* (New York: Palgrave MacMillan, 2009).
- Wilson, G.: *The Intentionality of Human Action* (Stanford, CA: Stanford University Press, 1989).
- Wright, L.: *Teleological Explanations: an Etiological Analysis of Goals and Functions* (Berkeley: University of California Press, 1976).