Why are people often rational? Saving the causal theory of action

Mihnea Capraru

Since Donald Davidson issued his challenge to anticausalism in 1963, most philosophers have espoused the view that our actions are causally explained by the reasons why we do them. This Davidsonian consensus, however, rests on a faulty argument. Davidson's challenge has been met, in more than one way, by anticausalists such as C. Ginet, G. Wilson, and S. Sehon. Hence I endeavor to support causalism with a stronger argument. Our actions are correlated with our motivating reasons; to wit, we often do what we have reason to do. In yet other words, we are often rational. Our frequent rationality is easily explained if causalism is correct, but looks like a staggering coincidence otherwise. Anticausalism thus appears to be ill-equipped to account for the very existence of rational behavior, and so far no attempts in this direction have succeeded.

Keywords: causal theory of action • motivating reasons • rationalization • correlation and causation • teleological explanation

1 Introduction

When we explain an action by the reasons that motivate it, we thus identify the action's purposes. Since such explanations are teleological, numerous philosophers have adopted, until Donald Davidson argued otherwise, the doctrine of anticausalism, a doctrine that denies that our actions are caused by the reasons[1] why we do them. In 1963, Davidson introduced an influential argument against anticausalism. He argued that only causalism seems equipped to answer an important kind of question: the question which motivating reason, out of an agent's several, the agent has acted on. Take Amelie, who has two motivating reasons to travel to Chicago – to meet Carol ($\text{reason}_{\text{Carol}}$) and to meet Steve ($\text{reason}_{\text{Steve}}$). Although there are two reasons to do it, Amelie travels to Chicago only for one, namely to meet Carol. Causalism can easily explain what makes it so: Amelie is acting on $\text{reason}_{\text{Carol}}$ because she is caused to travel by $\text{reason}_{\text{Carol}}$. Anticausalism, however, appears at first to be out of options.

When Davidson gave this argument, he was supporting three related, but distinct views. Let us distinguish them explicitly.

First, Davidson was arguing for the view that I will call basic causalism. Basic causalism, straightforwardly enough, says that our actions are caused by the reasons why we do them.

Second, Davidson was arguing for a subtler view that I will call essential causalism. Before I formulate essential causalism, consider the relation that obtains only between an action and the reason why the action is done. This relation is not merely the relation of motivating or incentivizing: although Amelie was incentivized to travel to Chicago both by $\text{reason}_{\text{Carol}}$ and by $\text{reason}_{\text{Steve}}$, she only traveled for $\text{reason}_{\text{Carol}}$. The relation at issue is something that $\text{reason}_{\text{Carol}}$ has over and above $\text{reason}_{\text{Steve}}$. Since this relation selects the reason that Amelie acted on, let us call it the reason selector. It is the reason selector that makes a reason into the reason why an action has been done. This, then,

[1] It is somewhat jarring and strange to say that actions are caused (or not) by reasons, as if reasons as such had causal powers. Alas, we don't have a wieldy and theory-neutral circumlocution. Davidson himself identified reasons with pairs of the form ⟨belief, desire⟩, and thought that actions are caused by such pairs of mental states. Perhaps we might say, more generally, that actions are caused (or not) by our having reasons, and leave it open whether our having a reason constitutes a fact, a concrete occurrent mental state, or something else.

is in brief the thesis of essential causalism: that the reason selector is causation.[2]

Let us notice three things. First, essential causalism does not follow trivially from basic causalism: it is logically possible that our actions are always caused by the reasons why we do them (basic causalism) yet that it is not causation in virtue of which these are the reasons why we do them (negation of essential causalism). Second, notice that essential causalism does not reduce to saying that necessarily actions are caused by their reasons. Rather, essential causalism affirms a particular kind of necessity, the necessity that stems from the fact that when an action is done for a reason, this is in virtue of causation. Third, notice that essential causalism is a claim about metaphysics and not about concepts; thus essential causalism is compatible with a non-causal analysis of our ordinary concept of acting for a reason.

Finally, Davidson was arguing for a third view, one that I will call conceptual causalism. Conceptual causalism maintains that explanation by reasons "is a species of ordinary causal explanation" (Davidson 1963: 685); causation figures in the conceptual analysis of acting for a reason: "Two ideas are built into the concept of acting for a reason ...: the idea of cause and the idea of rationality. A reason is a rational cause" (Davidson 2001).

Although the Davidsonian causalist consensus has exerted tremendous influence, its core argument – Davidson's challenge – has turned out to be less compelling than it once seemed. The challenge has been responded to, prima facie plausibly, by anticausalists. Carl Ginet has responded by identifying the reason selector with concurrent intention. On this view, the reason for an action is selected by the intention that has accompanied (but has not necessarily caused) that action (1989; 1997; 2007; 2008; 2016): if Amelie intends to meet Carol by traveling to Chicago, then she is traveling to Chicago for $\text{reason}_{\text{Carol}}$.[3] George Wilson and Scott Sehon, too, have answered the challenge.

[2] This pithy formulation could use a fair amount of refinement in order to exclude potentially deviant causal chains. Ideally the correct types of causal chains should be worked out. There are likely multiple such causal chain types, some of them involved, e. g., in paradigmatically rational and deliberate action, whereas others are involved in action that is still rational, but habitual and unreflective. However, this is not the place for tackling a complex endeavor of this kind.

[3] Although anticausalists can employ this concurrent-intention strategy toward a full-blown attack on causalism, it is important to notice that Ginet himself rejects causalism only for free actions. As for non-free actions, Ginet is willing to entertain not only basic but also essential causalism (2008).

To this end, they have identified the reason selector with the action's (irreducibly) teleological features: if Amelie travels in order to meet Carol, then she travels for $\text{reason}_{\text{Carol}}$ (Wilson 1989; Sehon 1994; 2000; 2005).[4] To be sure, Wilson and Sehon are assuming that there is such a thing as irreducible teleology, which causalists may reasonably deny, especially if they are also naturalists. But Ginet doesn't seem to assume anything naturalistically unacceptable. It seems, therefore, that Ginet has enabled anticausalists to meet Davidson's challenge rather easily.

This being so, some might already proclaim the demise of the causalist consensus: "The widespread acceptance of the causal theory of action explanation stems apparently from the perceived lack of a developed alternative .... If ... there is a plausible alternative ..., then there is no compelling rationale for the causal theory of action" (Sehon 2000: 79–80). Yet we cannot abandon the consensus lightheartedly. For one thing, causalism has informed numerous important positions in the philosophy of mind, in the theory of action, and on the problem of free will. In particular, each of the following views is imperiled by the threat to causalism: functionalism (see Sehon 1994), compatibilism, event-causal libertarianism, and agent-causal libertarianism (see Ginet 1997).

But the trouble runs even deeper. Not only is causalism presupposed by much of our contemporary philosophy, but it also looks like the best hope for understanding how reason fits within the world. All else being equal, we are more likely to do something if we have a motivating reason to do it. This is not surprising if our motivating reasons tend to cause the actions that they motivate. Why, though, would we so often (and so luckily?) do what we have reasons to do, if our reasons had no causal power over our actions? Anticausalism appears to transform ordinary rational behavior into a miracle. This, then, is the new challenge for anticausalists: to find a compelling way to explain non-causally why our actions regularly fit with our reasons, and are predicted by them.

[4] As with concurrent intentions, anticausalists can employ primitive teleology toward a full-blown attack on causalism. And indeed, Sehon has done just that: he has rejected all three of Davidson's views including basic causalism (2000:68). Wilson, on the other hand, has intimated that basic causalism could be right (1989:183); instead, he has mainly been concerned to reject conceptual causalism.

2 The New Challenge: In Support of Basic Causalism

The New Challenge is, first of all, an argument for basic causalism. Aside from this, in section §5 I will also examine a version of the New Challenge that supports essential causalism. As an argument for basic causalism, the new challenge is, in short, to explain why our actions correlate predictively with our motivating reasons.

This argument is kindred to the No Miracles Argument (NMA) as used to defend realism about beliefs and desires (see J. Fodor 1985:79–80). Hilary Putnam's NMA is a general argument for scientific realism. The general argument is that "realism ... is the only philosophy that doesn't make the success of science a miracle" (Putnam 1975:73). When used to defend realism about beliefs and desires, the NMA goes as follows: When we ascribe beliefs and desires to people, we can predict these people's actions; hence it is likely that when we ascribe beliefs and desires, we describe reality as it is. Notice, now, that this is an argument that beliefs and desires are real. Our anticausalists, however, do not deny that; more to the point, they do not deny that reasons are real. On the contrary, they accept that reasons are real, and that reasons explain actions. What they deny is that reasons cause actions.[5] Thus whereas Fodor is asking the antirealists: why do beliefs and desires often predict actions, if they are not real?, we will ask the anticausalists: if reasons do not cause actions, then why do they serve to predict them?

A similar question is raised by Erasmus Mayr in 2011. Mayr objects to G. H. von Wright's anticausalism, on the grounds that it "gives rise to a mysterious 'parallelism' between intentional actions and the occurrences of the bodily motions which are the results of these actions" (2011:32–33). Notice that Mayr's question is not an instance of the New Challenge, but rather, an instance of the mind–body problem. His question concerns the correlation between actions and bodily motions, the latter defined as 'movings in the intransitive sense' (Mayr 2011:29), i. e., as changes in the po-

[5] As we saw in the Introduction, what anticausalists deny varies from case to case. E. g., Sehon seems to deny that actions are caused by reasons altogether, whereas Ginet denies that free actions are thus caused, but allows causation-by-reasons for actions that aren't free. Others may well allow that all actions are caused by reasons, but deny that it is reasons-causation that rationalizes the actions. In this section I am discussing the New Challenge as it concerns basic causalism, hence I am addressing an anticausalist who denies that actions are caused by reasons altogether.

sitions of bodies. The New Challenge, however, is not to explain mind–body correlations, or in particular the action–motion correlation; the challenge is to explain the action–reason correlation. Notice that the action–reason correlation still obtains even for purely mental actions, such as focusing or imagining, which often occur without bodily motion. (This is not to take a stance on whether an action can be purely mental, but simply to point out that the New Challenge is logically distinct from the mind–body problem.)

Here, now, is the shape that the New Challenge takes for Ginet's concurrent-intention view. On this view, when a desire gives the reason for an action, this is not because the desire causes the action, but merely because the agent intends the desire to be satisfied by the action (Ginet 1989:36–38). The New Challenge now is to answer the following question: if desires do not cause the actions they explain, then why do we so often perform just those actions that will satisfy our desires?

Or consider Wilson's and Sehon's view. On this view, the reason for an action is given by its purpose, and purpose does not reduce to causation. Here, then, is the challenge: if our actions are not caused by our having purposes, then why do our actions and purposes so often agree?

Let us dwell on this point lest we miss it because of its obviousness. Many of us have asked at an early age how cars move. Few of us, however, happened to ask how animals do. Apparently we are born with an inbuilt expectation that inanimate things do not move 'by themselves' but animals do. Cars defeated our expectation, so we asked why; animals did as expected, so we never noticed that they too were mysterious. Aristotle may have perplexed quite a few of his contemporaries when he discovered the very problem of animal locomotion. Something similar occurs with actions. So obvious is it that our actions accord with our reasons, and so deeply is this fact woven into our lives, that we easily overlook that it is indeed a striking fact that calls for an explanation. We are puzzled for instance when we procrastinate, but we are not instinctively puzzled when we do not procrastinate. In fact, both these behaviors need to be explained.

3 Causalism and the Action–Reason Correlation

There are two very broad views on the nature of causation. On one view, causes 'make their effects happen,' or 'produce' them. On the other view – let us label it correlationism – causation consists merely in (simple or sophisticated) correlations between causes and effects. As for myself, I am sympathetic to the first view, but my sympathies are unimportant here. What matters is that both views allow Causalism to respond to the New Challenge.

Let us first look at correlationism. Its oldest and simplest version is Humeanism. If Humeanism is correct, then in all likelihood the matter is already settled. If causation reduces to the constant conjunction of causes and effects, then given the action–reason correlation, causalism seems to follow trivially.

To be sure, the action–reason correlation is not perfect, because we don't act on all of our motivating reasons. But the same goes for virtually every causal generalisation, except perhaps for those of fundamental physics.

Close to Humeanism is the counterfactual theory of causation.[6] Notice, now, that actions correlate with reasons not only in the actual world, but also counterfactually: if people didn't have the reasons they have, then ceteris paribus they wouldn't act in the ways they do. This makes anticausalism implausible on both the Humean and the counterfactual view of causation.

Further varieties of the Correlation view are the interventionist[7] and the probabilistic[8] accounts of causation. If we intervened to ensure that Amelie does not have $\text{reason}_{\text{carol}}$ to travel to Chicago, then ceteris paribus she would not travel there. Recall that we are assuming that Amelie does not, at the actual world, travel to Chicago for $\text{reason}_{\text{steve}}$. Hence if we intervened in a surgically precise manner, as envisioned by interventionists, then: Amelie would no longer have $\text{reason}_{\text{carol}}$ to travel; she would still have $\text{reason}_{\text{steve}}$; but she would still not travel for $\text{reason}_{\text{steve}}$. Likewise, an action is more probable, ceteris paribus, in the presence of the reasons that explain it, than it would be without

[6] David Lewis, 1973. Famously, Hume himself seems to equivocate between the two (1748, Section VII).
[7] Spirtes, Glymour, and Scheines (2001); Pearl (2009).
[8] Reichenbach (1956).

those reasons. It is plausible that on any robust attempt to explain causation in terms of correlation, the action–reason correlation will result in an action–reason causal generalization.

An important reason why many of us balk at correlationist theories of causation is because these theories leave the relevant correlations unexplained. Intuitively it seems that empirical correlations are explained by deeper causal generalizations. But correlationism has it that certain basic correlations just are.[9] Perhaps, e. g., it just happens that all electrons abide by Wolfgang Pauli's Exclusion Principle. Or perhaps the Exclusion Principle can be explained by deeper correlations; but at some point we are said to reach a level of unexplained correlations, correlations that might as well, for all we know, be just cosmic coincidences.

Since correlationism maintains that causal correlations are ultimately unexplained, one may worry prima facie that correlationism is incompatbile with causalism about action.[10] The worry goes as follows: According to causalism, actions are not just caused but also explained by the reasons that motivate them. (It would be a pyrrhic victory to show that actions are caused by their reasons, if we had to pay for it by denying that actions are explained by their reasons.) But – so goes the worry – correlationism entails that causes do not explain their effects. This is because correlationism has it that the most basic cause–effect correlations are unexplained; and if the correlation between C-type causes and E-type effects is unexplained, then – so goes the worry – every particular E-type effect is also unexplained; for otherwise the particular explanations of the E-type effects would sum up to an overall explanation of the C-type–E-type cause–effect correlation.

This worry, however, is surmountable. It is logically open for the correlationist to claim that a) cause–effect correlations are ultimately unexplained, but b) particular effects are explained by the particular causes with which they go together. Let us recall, for illustration, the venerable deductive–nomological (D-N) model of scientific explanations.[11] According to D-N, we scientifically explain an event by deducing the event's occurrence from one or several lawlike generalizations, as

[9] Psillos (2012).

[10] This worry is perhaps present in the thinking of present-day 'Naïve' Action theorists (Wiland 2012; Fritts 2021), whose views will be discussed in section §8.

[11] Hempel and Oppenheim 1948.

well as from certain initial conditions. The lawlike generalizations are ultimately left unexplained, but the events we deduce from them are said to be thus explained. Notice that this is not contradictory: it is the generalizations that are unexplained, not the particular events that follow from them; at the same time, it is the particular events that are explained, and not the generalizations. Nothing is both explained and unexplained.[12] Likewise, the correlationist causal theorist of action, not being bound by the logical empiricists' fear of causation, can replace talk of 'lawlike generalizations' with talk of causal generalizations, and consistently assert that a) causal generalizations are unexplained, but b) particular effects are explained by their particular causes.

As opposed to correlationism, production views maintain that facts about causation are over and above facts about correlation. An important illustration is Wesley Salmon's causal process theory,[13] as well as the closely related conserved-quantity theory (Dowe 2000), conserved trope theory (Ehring 1997), and transference theory (Aronson 1971; Fair 1979). Another important approach is the appeal to powers (Molnar 2003; Shoemaker 1980) or capacities (Cartwright 1989). Such views maintain that causation is not merely a matter of correlation, no matter how sophisticated or hypothetical, but rather, that causes produce their effects. If so, then it is a truism that causation can explain correlation. If causation is production, then causal generalization is production generalization. Why are thunderstorms, for instance, preceded by cumulonimbus clouds? Because storms are produced by stormclouds. Vice versa, why are cumulonimbus clouds regularly followed by thunderstorms? Because that is what they produce ceteris paribus. And similarly, why are motivating reasons often followed by the actions they motivate? Because – if causalism is right – that is what they tend to produce.[14]

[12] To be sure, D-N has been abandoned for good reasons (W. C. Salmon 2006), but those reasons have to do with its inability to model real-world scientific explanations, and not with a lack of internal consistency.

[13] W. Salmon 1984, prefigured by Bertrand Russell in 1948.

[14] Causalism, as we see, can meet the New Challenge regardless which of a wide range of theories of causation turns out to be correct. Does this entail that Causalism is compatible with every one of these views? Not necessarily. Perhaps some of these views can make trouble when it comes to meeting the old challenge, i. e., Davidson's. The old problem of deviant causal chains (Davidson 2015) could eventually compel the causalist to adopt a particularly strong conception of causation, in order to distinguish the rationalizing causal chains from the non-rationalizing ones. If so, then the fate of Causalism would evidently hang on the metaphysical commitments required by the relevant conception of causation. But this discussion would be far outside the scope of our article. Our article is concerned with the New Challenge, on whose score Causalism seems to do entirely well.

I might now be expected to argue directly that anticausalism cannot explain the action–reason correlation. That, however, is not the dialectical import of the New Challenge; the point is not to refute anticausalism once and for all, but rather, to shift the burden of proof back to anticausalism, now that Davidson's original challenge has been met. I will examine, in section §6, several potential anticausalist answers to the New Challenge, and I will explain why they aren't satisfactory for the time being.

4 Wilson and Sehon on the Action–Reason Correlation

In 1989 Wilson looks at a question closely related to that of the action–reason correlation, namely, the question why reasons support counterfactuals – why it is generally true that people wouldn't do what they do, if they didn't have the reasons they have. Wilson agrees that this is a good question, but he maintains that it is simply a fundamental fact that reasons support counterfactuals:

> Now, intuitively, it is a fundamental fact about genuine agency that, if an agent has no objective in view that he believes will or might be promoted by his ϕing, ... then he will not even try to ϕ ... (Wilson 1989:198).

First, Wilson's conditional doesn't seem to be evidently correct. Is it always true that an agent wouldn't have ϕed if the agent hadn't had an objective in view? This seems to rule out absurd, Camusian or Kafkaesque behavior. Perhaps Wilson is presupposing that absurd behavior is merely behavior, and not genuine agency. Or perhaps absurd behavior is regarded by the agent as an objective in itself. Let us therefore assume for discussion that there is a sense in which Wilson's conditional is true; perhaps it is even a conceptual truth. But there is a deeper problem. Reasons support counterfactuals in yet another way, namely:

> Even if an agent has no reason to ϕ, the following is still true: had the agent had strong enough reasons to ϕ, and had these reasons not been defeated by stronger reasons to refrain from ϕ, then ceteris paribus the agent would have probably ϕed.

The above is not just a conceptual truth. In fact, without the qualifications, it would plausibly not be a truth at all: a prima facie counterexample would be procrastination. But it is, with all qualifications, a truth: if another course of action had been the rational one, then often enough people would have taken that other course. Why is this so? Could Wilson maintain that this too is a fundamental fact? Does it just so happen that people are often rational?

To my judgment, at least, human rationality doesn't look like a fundamental fact about the universe. Perhaps divine rationality is different, for those who believe in it, but people don't just act rationally without any further explanation. Suppose, e. g., that yesterday, being sleep-deprived, I irrationally ate a large merengue pie for lunch that I had meant to share with my guests at dinner. Today, having caught up on my sleep, I am behaving rationally and eating things that make sense. Today's rationality is not a fundamental fact; it has its own explanation, namely, that I have slept enough and become reasons-responsive once again.

Sehon seems to sympathize: "From the standpoint of simplicity, it would indeed be nice if teleological explanations were reducible to causal explanations" (2005:172). But Sehon doesn't think we can reduce teleology to causation. Here is his argument, which seems inspired by the classical objection to type physicalism (Putnam 1967; J. A. Fodor 1974):

> The crucial point is that, although there will be a causal story behind each instance of an agent's behaviour conforming to [the principles of rational agency], these instances will not have a common causal story. Each goal-directed behaviour will have a causal story, but the stories will not, from the standpoint of physical science, have natural properties in common (Sehon 2005:218).

Now Sehon is probably right about physical science. It seems rather implausible that physics will some day find the correct definition of rational behaviour. But this doesn't entail that we cannot reduce teleology to causation. Perhaps we can reduce teleological explanation, if not to a common causal story at the level of physics, then for instance to a common story at the level of cognitive science. After all, physics doesn't have a monopoly on causal generalisations, and there is no reason in

principle why cognitive science cannot offer a high-level causal characterisation of the connections between reasons and actions.[15]

There is a parallel question in biology:[16] how to explain that so many biological traits serve their organisms' fitness, unless these traits' existence was caused (in part) by the same factors that determine what counts as fitness, namely, by organism–environment interactions. In biology, of course, the answer is that the traits' existence was ultimately caused (in part) by organism–environment interactions, in the course of natural selection (and often also through phenotypic plasticity). Now we will probably never obtain a general characterisation of natural selection from physics, but we don't take this as a reason to posit primitive biological adaptation that does not reduce to a certain type of causal history. Instead, we conclude that natural selection is multiply realizable.

5 The New Challenge: In Support of Essential Causalism

As promised at the beginning of section §2, the New Challenge comes with two sub-arguments: one for basic causalism, the other for essential causalism. It is now time to introduce the second of these. According to essential causalism, not only are actions caused by reasons, but it is causation that makes it the case that these reasons are why the actions are done. Just as the New-Challenge argument for basic causalism relies on the correlation between actions and reasons, the New-Challenge argument for essential causalism relies on the higher-level correlation between causation itself and any candidate non-causal reason selectors.

It is possible in principle for anticausalists to concede basic causalism, yet to reject essential causalism. To wit, they could concede that our actions are caused by the reasons why we do them, yet deny that causation is the reason selector. Instead, they could identify the reason selector, e. g., with concurrent intention or with irreducible teleology.

[15] Perhaps Sehon doesn't refer by 'physical science' only to physics, but to any form of third-person scientific knowledge, including cognitive science. In that case, however, his premise becomes prima facie implausible; why, indeed, would it be obvious that cognitive science cannot uncover a common causal story about reasons and actions?

[16] Parallel as it is, though, the biological question doesn't need to be construed as concerning teleology.

Let us call this view accidental causalism – 'accidental' because it is not essential. The New Challenge is now to explain why accidental causalism's non-causal reason selector correlates with causation.

Let me illustrate. $\text{Reason}_{\text{Carol}}$ was why Amelie travelled to Chicago, so accidental causalism accepts that $\text{reason}_{\text{Carol}}$ caused Amelie to travel. Nevertheless, accidental causalism maintains that it was not causation in virtue of which $\text{reason}_{\text{Carol}}$ was why Amelie travelled; instead, this was in virtue of some other, non-causal reason selector. Now if accidental causalism is right, then this reason selector, albeit non-causal, occurs at places where causation, too, occurs. Here is why:

1. The reason selector occurs between actions and the reasons why they were done. (definition)
2. An action is caused by the reason why it was done. (basic causalism, accepted by accidental causalism)

∴ The reason selector occurs between actions and the reasons they were caused by.

Here then is the New Challenge as addressed to anticausalists who accept only basic but not essential causalism: they must explain why their non-causal reason selector occurs at the same places as causation. Pending a convincing answer, it seems safe to conclude that if basic causalism is right, then so is essential causalism.

6 Potential Answers to the New Challenge

Let me now address three strategies that anticausalists could adopt.

First, they could maintain that actions correlate with reasons not because of a causal connection, but merely as a matter of defeasible, empirical generalization. This, to be sure, is not logically impossible. It is, however, improbable, because it appears to turn the action–reason correlation into a cosmic coincidence.

Second, anticausalists could adopt a form of Leibnizian parallelism, and maintain that actions are not caused by reasons, but rather, that actions and reasons are produced by parallel causal chains

that trace back to a remote common cause. (For Leibniz himself, this remote common cause would be God.) This would allow actions to correlate with reasons without being caused by them. Such a speculative reply, however, would violate Occam's Razor, and therefore it would seem rather improbable compared to regular causalism.

As an alternative to Leibnizian parallelism, anticausalists could invoke a more proximate kind of common cause – perhaps, e. g., some yet to be discovered theoretical entities of cognitive science. If this common cause were to cause first reasons and then the corresponding actions, then this would explain why reasons predict actions. In such a case the New Challenge could be met. Meanwhile, however, the challenge stands.

Third, anticausalists could shift the locus of causation from reasons to reason selectors. That is, Ginet-style anticausalists could maintain that actions are caused not by reasons, but rather by intentions. These anticausalists could then explain that our actions accord with our reasons because a) our actions are caused by our intentions, and b) our intentions are formed in light of our reasons. Similarly, Wilson–Sehon-style anticausalists could maintain that our actions are caused not by our reasons, but rather by our purposes, or perhaps by ourselves as purposeful agents. These anticausalists could explain that our actions accord with our reasons because a) our actions are produced by our purposes, and b) our purposes are chosen in light of our reasons.

If anticausalists chose this latter strategy, however, they would shift not only the locus of causation, but also the target of the New Challenge. We would no longer ask why actions accord with reasons, but why intentions or purposes do. We have represented anticausalists as explaining that purposes or intentions accord with reasons because they were formed 'in light' of them. Light, though, is a metaphor, and behind the metaphor must lie an explanation. At first it may look attractive to reply that the light metaphor stands for causation, i. e., that intentions or purposes accord with reasons because they are caused by them. This reply, however, is not open to anticausalists, because it would make reasons the ultimate causes of actions, and thus it would bring back causalism.

7 Sense-Making Neo-Wittgensteinianism

Predavidsonian anticausalism flourished at a time when many philosophers saw metaphysical inquiry as fundamentally misguided. Such philosophers might have rejected our New Challenge by protesting that when we explain actions by reasons, we simply aren't trying to explain why things are as they are – why our reasons and actions correlate as they do. Instead, when we explain actions by reasons we aim to 'make sense' of the actions, not the kind of sense we make of falling bodies when we find what determines their paths, but rather the kind of sense that is proprietary to rational agents, the kind that relates to other such concepts as 'meaning' (of an effort), 'purpose,' or 'significance.' This Wittgensteinian outlook has been revived recently by Julia Tanney (2009) and by Giuseppina d'Oro (2012; 2019). According to them, when we explain an action by giving reasons we strive not to find that action's causes, but rather to "[enable] the one who is puzzled to see the action in a new, sense-making light" (Tanney 2009:100), "to make sense of [the] action by establishing a relation of rational fit between the premises and conclusions of a practical argument" (D'Oro 2012:216). Davidson's challenge is to be solved by the following principle: An action is explained by that particular reason which makes the most sense of it.

First, let me point out that this has a very unfortunate consequence. It renders us far more rational than we are. It entails that everything we do, we do for the best of reasons. Little room is left, for instance, for akrasia. We must now believe that procrastination, or alcohol abuse, are not irrational actions done for short-sighted reasons, and done in spite of the better reasons against them. Instead, weak-willed actions must now be seen as simply the actions that best make sense in the light of everything that we know about the agents involved. These Panglossian implications are difficult to square with what we ordinarily believe about rationality and about the widespread lack thereof. It is true enough, as I repeatedly emphasize, that people are often rational; but they are not always so.

Let us now return to the New Challenge. I maintain that Neo-Wittgensteinians are not exempt from it. Why is it even possible to make sense of so many actions by giving reasons? I have yet to see

a compelling rationalization of the periodic table of elements, or of Pascal's triangle, or even of why Germanic languages want every sentence to have a syntactic subject, whereas Romance languages can live without it. To be sure, there is no shortage of explanations in chemistry, combinatorial analysis, etc. Such explanations, however, are not teleological. It is primarily in the realm of human action that teleological explanation flourishes. Why, then, is human action amenable to teleological explanation, if not because actions are caused by the reasons that motivate them?

It is true enough that not everyone must do every job, and in particular, not everyone must do the metaphysical job of explaining why actions tend to make sense, i. e., why they correlate with reasons. But it is one thing to refuse to explain the correlation, and it is another thing to render the correlation a much deeper mystery than it needs to be.

Consider, for comparison, the following analogy. Carnivores tend to have claws, and this calls for an explanation. Regardless, if I'm not an evolutionary biologist but rather a geologist, then it seems perfectly fair for me to decline to explain it myself. If I then make claims about geology, however, claims that render it a cosmic coincidence that carnivores have claws, then my geological claims are at least as improbable as the cosmic coincidence that they imply.

Back, now, to our problem. We could, in principle, take the following approach: On one hand, we could adopt the Tanney–d'Oro line and maintain that actions are always explained by the reasons that best make sense of them; at the same time, we could remain neutral on the question whether these reasons also cause the actions they explain. This would not be completely unreasonable, as far as the New Challenge is concerned, and setting aside the other objections to the sense-making view that I present in this section. We would merely give an account of how it is that a reason explains an action when it does so, and leave open the question why widespread rationality is possible at all.

If, however, we did take a stance on causalism and if we rejected it, then we would threaten to render the action–reason correlation a cosmic coincidence. Neo-Wittgensteinians do not merely refrain from accepting causalism; they appear to deny causalism outright, even in the basic sense. Tanney, in particular, denies that mental terms function to refer (2009, p. 105). If reasons-explanations

do not refer, then ipso facto they do not refer to causes. D'Oro, in turn, maintains that "reasons are not causes" (2019, p. 97). This means, however, that the New Challenge becomes their burden too.

Perhaps Neo-Wittgensteinians could resort to a radically antirealist theory. According to this radical view, an agent's motivating reasons do not exist independently of rationalizations. Rather, reasons are constructed by third parties (or even by the agent) in the process of sense-making. Small wonder, then, that people's actions correlate with their reasons, if we assume that these reasons are postulated after the fact, in the light of actions that were already performed![17]

This radical antirealism cannot stand. Any correct Neo-Wittgensteinian theory must explain, or at least leave room for explaining, not only why we can make sense of actions in hindsight, but also why we can predict these actions based on people's motivations. These predictions, to be sure, are fallible, but they are a lot better than chance. Why are actions often predicted precisely by the reason attributions that would best make sense of them?

Perhaps the antirealist can reply that people willingly act in the most sense-making ways, i. e., in the ways they will be able to rationalize by constructing the best reasons, even when they wouldn't have cared about those reasons otherwise.

This would be an interesting thought, but also a self-defeating one. It would merely shift the

[17] Something similar to this view is imputed to D'Oro by Al Mele (2013). Mele contends that D'Oro cannot distinguish between reasons in reality and reasons in fiction. D'Oro rejects the charge in 2019, but only halfheartedly. She admits that there are facts of the matter only about real people's mental processes, and not also about those of fictional people, but she claims that these mental processes are irrelevant to reasons-explanations. Her argument is that reasons have a normative dimension, whereas concrete mental processes do not. This argument is not convincing prima facie. Normativity comes in kinds. There is moral normativity, but also prudential normativity, etc. As Ruth Millikan has argued beginning in 1984, there are even norms that are purely natural; for instance, the mean of a statistical distribution is also normal in a sense. D'Oro's argument, in particular, appears to be inspired by Hume's line of separation between 'is' and 'ought.' But the Humean separation does not exclude conditionals of the following form: 'if you want to satisfy your pro-attitude(s), then it would be instrumentally rational for you act as your beliefs advise'; e. g., 'if you want to eat a slice of tiramisu, then it would be instrumentally rational for you to go to the cake shop where you believe they sell it.' It is from such instrumental-rationality conditionals that motivating reasons inherit their normativity. Such conditionals can be deduced from a description of the world, and do not entail anything about whether, e. g., eating tiramisu is intrinsically good, or an end in itself, or a duty, or right, or just, etc. Information about people's interests and mental processes entails information about what is instrumentally rational for them to do. This, indeed, is why it is possible for real people to have secret motivations, in a way in which this is not possible for Sherlock Holmes. (It is possible for Sherlock to have secret motivations that we only read about in chapter five, but he cannot have secret motivations that we never read about at all, because he does not have interests or mental processes over and above the fiction.)

problem from simple reasons – e. g., finding food – to sophisticated reasons such as enacting a coherent narrative, living a meaningful life, or behaving in rationally accountable ways. The simple reasons are now said to be constructed, but what becomes of the sophisticated ones? For instance, by virtue of what is the agent acting in order to enact a coherent narrative? If such higher-level reasons are still constructions, then why do they predict the actions? At pain of infinite regress, we must allow at least the higher-level reasons to count as robustly real, and not just as constructed.

A final worry about Neo-Wittgensteinianism is that it seems to clash with the way in which we judge actions morally. This paper, to be sure, is not about justifying reasons, but about motivating ones. The latter, however, are not irrelevant to moral judgment. When we morally or legally evaluate a deed, we often take into account the reasons that motivated it – whether it is explained by need, greed, revenge, jealousy, hatred, sadism, etc. We distinguish intent crimes, premeditated crimes, crimes of passion, etc. According to Neo-Wittgensteinianism, however, motivating reasons are only related to actions by virtue of third-party sense-making. It seems possible, on this view, that if two distinct third parties had different information, different Bayesian priors, or different general understandings of the world, then these two third parties would best make sense of an action in different ways. Then, however, an action's motivation would become relative to whom we ask, and it seems unjust to judge people's deeds based on such relative interpretations. If we adopted Neo-Wittgensteinianism, then apparently we would need to either ignore motivation when morally judging deeds, or to adopt a form of judge-sensitive moral relativism.

8 The Naïve Action Theory

Another contemporary approach that attempts to overcome Davidson's Challenge is the Naïve Action Theory (Thompson 2008; Wiland 2012; Fritts 2021). This theory takes its name from the 'naïve', prima facie non-mentalistic action explanations frequently volunteered in ordinary conversation:

A: 'Why are you starting the car?'

B: 'Because I'm visiting grandma.'

Notice that B explains one action, starting the car, by another action – visiting grandma – without explicitly invoking any desires, plans, intentions, or in general, pro-attitudes. This accounts for the 'naïve' aspect of the theory. The theory is, furthermore, non-causal, because the action explained, starting the car, is not caused by the explaining action – visiting gradma.

Proponents of the naïve action theory maintain that this theory can answer Davidson's challenge (Wiland 2012, p. 157). Wiland seems to regard this as intuitively obvious, but in fact the claim is in need of argumentation. In 2021 Megan Fritts takes on this task. She maintains that reasons explanations are grounded in counterfactuals. Using our earlier example, if Amelie travels only for $\text{reason}_{\text{carol}}$, and not also for $\text{reason}_{\text{steve}}$, then this is because:

1. Amelie travels to Chicago,
2. she does so because she is meeting Carol, and
3. if she weren't meeting Carol, then she wouldn't be traveling to Chicago.

It is thus that Fritts intends to account for the difference between $\text{reason}_{\text{carol}}$, which explains Amelie's traveling to Chicago, and $\text{reason}_{\text{steve}}$, which could also have motivated the traveling, but has not in fact done so. Amelie's action counterfactually depends on her having $\text{reason}_{\text{carol}}$, but does not so depend on her having $\text{reason}_{\text{steve}}$.

As with the counterfactual theory of causation, Fritts's counterfactual theory of reasons-explanation is open to counterexamples involving preemption, or what one may call 'backup reasons.' As Fritts illustrates:

> When asked why I am purchasing eggs ... I respond, "I'm making omelets for dinner tonight." As it happens, however, ... I am actually purchasing eggs because I have a romantic interest in the person who runs the egg stand, and I want to see them. The reason I gave for my egg-purchasing was not a lie—I do plan to make omelets for dinner,

> and I mistakenly believe this is my reason for going to the egg stand. In this scenario, however, my chicken farmer crush, and not my future omelet-making, is my reason for visiting the egg stand. However, in the closest possible world in which I knew my crush would not be working the stand, I would still go and purchase eggs because my "backup reason," the omelets, would kick in. ... Seeing my crush, therefore, would not count as a reason that I went to the egg stand. (Fritts 2021, pp. 12698–99)

Let us label the two potential motivations $\text{reason}_{\text{farmer}}$ and $\text{reason}_{\text{omelets}}$. In these terms, Fritts rebuts the above counterexample as follows. If, at the actual world, she goes to the egg stand only for $\text{reason}_{\text{farmer}}$, and not also for $\text{reason}_{\text{omelets}}$, then at the nearest possible world at which the farmer is not at the stand, she does not go there at all: since she does not actually go there for $\text{reason}_{\text{omelets}}$, then something additional would need to change, aside from the farmer being absent, for $\text{reason}_{\text{omelets}}$ to motivate her to go there. That additional change would ensure that we are no longer at the nearest possible world at which the farmer is absent.

This rebuttal is effective enough as far as the original counterexample is concerned. But we can introduce a stronger counterexample that resists the rebuttal:

Every Thursday the supermarket sells avocadoes at a 30% discount. Today is Avocado Thursday. I want avocadoes on discount, hence I go to the supermarket. I am also running out of Greek yogurt, and this could have been a good enough reason to go there, had I considered it. But I never even think about the yogurt, because Avocado Thursday comes to mind first and foremost, and once I make up my mind to go for the avocadoes, there seems to be no point in deliberating any further.

What, however, would I have done had I learned that Avocado Thursday has been canceled because of disruptions to international shipping lanes? I would have done what I always do when I don't know whether to go to the supermarket: I would have consulted my checklist. On that list one item reads: 'Am I running low on Greek yogurt?' Hence at the nearest possible world at which $\text{reason}_{\text{avocado}}$ does not motivate me to go to the supermarket, I go there anyway because of

$\text{reason}_{\text{yogurt}}$. It is therefore false to say that I wouldn't have gone to the supermarket if I hadn't been buying discounted avocadoes. It seems therefore that the Naïve Action theory cannot answer Davidson's Challege. At the very least, the theory's counterfactual-based articulation cannot answer Davidson, and there does not currently seem to be another well-articulated reply in the literature.

Having examined how the Naïve Theory fares against Davidson's Challenge, let us now hold it up to our New Challenge. According to he Naïve Theory, one action A can be rationalized by another action B. This can occur either because A is a means toward B, or because A is a temporal proper part of B. Evidently this constitutes a rejection of basic causalism, because present means are not caused by future purposes – especially given that some of those purposes are eventually missed – nor are present temporal parts caused by present+future temporal wholes.[18] The New Challenge now becomes: by virtue of what do our actions self-assemble into rationally coherent wholes, if there is no causal relation between what we have reasons to do and what we end up doing? Why, for instance, do I consistently end up buying eggs when in the process of making omelets? It is logically possible to be engaged in the process of making omelets, but to fail for lack of ingredients; and yet, I reliably succeed at making omelets, because I reliably procure eggs when I have a reason to do so. Again we find the usual kind of robust correlation between the rationalizer (in this case the action or process of making omelets) and the rationalized (in this case the action of procuring eggs). There is no conceptual or a priori necessity for my actions to rationally cohere this way. My rationality calls for a further explanation, an explanation which, pending alternatives, is the familiar causal one.

9 Conclusion

Anticausalist philosophers of action have found plausible answers to Davidson's challenge. But as this paper has argued, they must do more. They must reach beyond the relation between one action and one reason, and explain the correlation that obtains between actions and reasons overall.

[18] The Naïve Action Theory could still be rendered causal on pain of positing backwards causation, but the theory's adherents do not do so.

Anticausalists who reject basic causalism need to explain why our actions correlate with our reasons, why we so often do what we have reason to do. If they accept basic causalism but still reject essential causalism, then they must explain the correlation, implicit in their view, between causation itself and their non-causal reason selector.

As yet I have said nothing to defend the view that I called conceptual causalism, the view that explanation by reasons is conceptually causal, i. e., that when we explain an action by a reason, we conceive of the action as caused by that reason. This is the strongest formulation of Davidsonian causalism, and I will not defend it because I don't know whether it is true. Perhaps, in this conceptual sense, the anticausalists are right. It seems entirely plausible that explanation by reasons could turn out to be irreducibly sense-making or teleological – albeit only at the level of common-sense concepts. Perhaps, indeed, when we explain an action by citing a reason, we pack into our explanation two distinct ingredients: a) that the reason makes sense of, or gives purpose to the action, and b) that the reason somehow accounts for how the action came to pass. Notice that the second ingredient is logically general. If so, then explanation by reasons is not conceptually causal, and it takes further reflection to settle on causation as the way that reasons account for how actions come to pass. At the same time, the first ingredient explains the intuition that reasons explanations contain – or imply – something over and above mere causation. Causes in general do not rationalize their effects, but reasons, in particular, do.

References

Aronson, Jerrold L. (1971). "On the Grammar of 'Cause'." In: Synthese 22, pp. 414–430.

Cartwright, Nancy (1989). Nature's capacities and their measurement. New York: Oxford University Press.

D'Oro, Giuseppina (2012). "Reasons and Causes: The Philosophical Battle and The Meta-philosophical War." In: Australasian Journal of Philosophy 90.2, pp. 207–221.

D'Oro, Giuseppina (2019). "On an imaginary dialogue between a causalist and an anti-causalist." In: Explanation in action theory and historiography: causal and teleological approaches. Ed. by Severin Schroeder. Routledge, pp. 97–111.

Davidson, Donald (1963). "Actions, Reasons, and Causes." In: Journal of Philosophy 60.23, pp. 685–700.

— (2001). "Psychology as Philosophy." In: Essays on Actions and Events. Oxford University Press, pp. 229–244.

— (2015). "Freedom to Act." In: Essays on Freedom of Action (Routledge Revivals). Routledge, pp. 137–156.

Dowe, Phil (2000). Physical Causation. Cambridge: Cambridge University Press.

Ehring, Douglas (1997). Causation and Persistence: A Theory of Causation. Oxford University Press.

Fair, David (1979). "Causation and the Flow of Energy." In: Erkenntnis 14, pp. 219–250.

Fodor, Jerry (1985). "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum." In: Mind 94.373, pp. 76–100.

Fodor, Jerry A. (Oct. 1974). "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." In: Synthese 28.2, pp. 97–115.

Fritts, Megan (2021). "Reasons Explanations (of Actions) as Structural Explanations." In: Synthese 199, pp. 12683–12704.

Ginet, Carl (1989). "Reasons Explanation of Action: An Incompatibilist Account." In: Philosophical Perspectives 3, pp. 17–46.

— (1997). "Freedom, Responsibility, and Agency." In: The Journal of Ethics 1, pp. 85–98.

— (2007). "An Action Can be Both Uncaused and Up to the Agent." In: Intentionality, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy. Ed. by Christoph Lumer and Sandro Nannini. Aldershot, UK: Ashgate, pp. 243–55.

Ginet, Carl (2008). “In Defense of a Non-Causal Account of Reasons Explanations.” In: Journal of Ethics 12, pp. 229–237.

— (2016). “Reasons Explanation: Further Defense of a Non-Causal Account.” In: The Journal of Ethics 20, pp. 219–228.

Hempel, Carl Gustav and Paul Oppenheim (1948). “Studies in the Logic of Explanation.” In: Philosophy of Science 15, pp. 135–175.

Hume, David (1748). An Enquiry Concerning Human Understanding.

Lewis, David (1973). “Causation.” In: Journal of Philosophy 70, pp. 556–567.

Mayr, Erasmus (2011). Understanding Human Agency. Oxford University Press.

Mele, Alfred (2013). “Actions, explanations, and causes.” In: Reasons and causes: causalism and non-causalism in the philosophy of action. Ed. by G. D’Oro. Palgrave-Macmillan.

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

Molnar, George (2003). Powers: A Study in Metaphysics. Ed. by Stephen Mumford. New York: Oxford University Press.

Pearl, Judea (2009). Causality. Cambridge university press.

Psillos, Stathis (2012). “Regularity Theories.” In: The Oxford Handbook of Causation. Ed. by Helen Beebee, Christopher Hitchcock, and Peter Menzies. Oxford University Press, pp. 131–157.

Putnam, Hilary (1967). “Psychological Predicates.” In: Art, Mind, and Religion. Ed. by W. H. Capitan and D. D. Merrill. University of Pittsburgh Press, pp. 37–48.

— (1975). “What Is Mathematical Truth?” In: Philosophical Papers, Volume 1: Mathematics, Matter and Method. Cambridge University Press, pp. 60–78.

Reichenbach, Hans (1956). The Direction of Time. Berkeley and Los Angeles: University of California Press.

Russell, Bertrand (1948). Human Knowledge: Its Scope and Limits. New York: Simon and Schuster.

Salmon, Wesley (1984). Scientific Explanation and the Causal Structure of the World. Princeton: Princeton University Press.

Salmon, Wesley C. (2006). "Famous counterexamples to the deductive-nomological model." In: Four decades of scientific explanation. University of Pittsburgh Press, pp. 46–50.

Sehon, Scott R. (1994). "Teleology and the Nature of Mental States." In: American Philosophical Quarterly 31.1, pp. 63–72.

— (2000). "An Argument Against the Causal Theory of Action Explanation." In: Philosophy and Phenomenological Research 60.1, pp. 67–85.

— (2005). Teleological Realism: Mind, Agency, and Explanation. Bradford Books. MIT Press.

Shoemaker, Sydney (1980). "Causality and Properties." In: Time and Cause. Ed. by Peter van Inwagen. D. Reidel, pp. 109–35.

Spirtes, Peter, Clark Glymour, and Richard Scheines (2001). Causation, prediction, and search. MIT press.

Tanney, Julia (2009). "Reasons as Non-causal, Context-placing Explanations." In: New Essays on the Explanation of Action. Ed. by Constantine Sandis. Palgrave Macmillan, pp. 94–111.

Thompson, Michael (2008). Life and Action: Elementary Structures of Practice and Practical Thought. Harvard University Press.

Wiland, Eric (2012). Reasons. Continuum.

Wilson, George M. (1989). The Intentionality of Human Action. Stanford, California: Stanford University Press.