



Action and Rationalization

Samuel Asarnow

To cite this article: Samuel Asarnow (2021): Action and Rationalization, Australasian Journal of Philosophy, DOI: [10.1080/00048402.2021.1936093](https://doi.org/10.1080/00048402.2021.1936093)

To link to this article: <https://doi.org/10.1080/00048402.2021.1936093>



Published online: 02 Jun 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Action and Rationalization

Samuel Asarnow

Macalester College

ABSTRACT

According to the 'standard story' in the philosophy of action, actions are those movements of a creature's body that are caused and rationalized by the creature's mental states. The attractions of the causal condition have been widely discussed. The rationalization condition is nearly ubiquitous, but it is notoriously obscure, and its motivation has rarely been made explicit. This paper presents a new argument for including the rationalization condition in the causal theory of action, and sketches a broadly Davidsonian theory of what rationalization is.

ARTICLE HISTORY Received 27 May 2020; Revised 20 April 2021

KEYWORDS action; rationalization; motivating reasons; Davidson; causal theory of action

1. Introduction

One central question in the philosophy of action concerns what concepts we need in order to understand the distinction between intentional actions and other things that agents do. A second question concerns how to understand the idea of a motivating reason for an action, or an agent's purpose or goal in performing an action. The causal-psychological theory of action (henceforth, *the Causal Theory*) answers these questions by appeal to the concept of causation and the concepts of the mental states. Intentional actions are events caused by a creature's mind, and the elements of the creature's mind that did the causing are (or specify) its reasons for acting. This 'standard story' of action [Velleman 2000; Smith 2012] is appealing because it distinguishes actions from other events, illuminates the use of reasons in action explanations, and distinguishes the agent's reasons for acting from the motives that could have been reasons but were not [Davidson 1980a; Dancy 2000: 161].

Most versions of the Causal Theory appeal to additional concepts as well. Following Davidson, many appeal to the controversial idea of *non-deviant* causation [Davidson 1980b; Stout 2011]. Also following Davidson, many hold that actions are caused *and rationalized* by mental states.¹ Unlike non-deviant causation, however, the idea of rationalization has attracted little attention. Its proponents have rarely (if ever) explained what it means and what work it is supposed to do, and those who do without it have rarely (if ever) explained why. The concept of rationalization, while widespread, is thus also somewhat mysterious.

¹ See, most clearly, Davidson [1980c]. He uses the term differently elsewhere [1980a]. Here is a sample of authors who appeal to rationalization: Antony [1989], Moya [1998], Mintoff [2002], Hornsby [2004], Smith [2004b, 2012], O'Brien [2006], Wedgwood [2006], Schlosser [2007, 2011], Hurley [2018], and Wald and Tenenbaum [2018],

The aim of this paper is to shed light on the rationalization condition. I give a novel argument for embracing the rationalization condition, and I sketch a theory of what it is for mental states to rationalize action.

The paper argues for three main points. First, versions of the Causal Theory lacking the rationalization condition face two important ‘subset problems’, both of which arise from the Causal Theory’s identification of actions not with events, but with events under particular descriptions. I argue that the subset problems provide a strong motivation for the rationalization condition.² Second, I develop a design specification for a theory of rationalization, and I argue that many theories of rationalization inspired by the literature face serious problems in satisfying it. Finally, I propose a more satisfactory theory of rationalization (and motivating reasons), inspired by Davidson,

The paper proceeds as follows. Section 1 gives an overview of the Causal Theory. Section 2 considers an argument, inspired by some remarks of Kieran Setiya’s [2007, 2011, 2013], according to which the rationalization condition is both unmotivated and subject to counterexamples.³ Section 3 responds to the first part of the Setiya-inspired argument by exhibiting the two subset problems and showing how the rationalization condition promises to solve them. Section 4 generalizes the second part of the Setiya-inspired argument, providing a design specification that any plausible theory of rationalization must satisfy. Section 5 shows that three theories of rationalization inspired by the literature fail to satisfy the design specification. Section 6 outlines a superior theory of rationalization, and the conclusion considers a pair of questions raised by the theory.

2. The Causal-Psychological Theory

In the spirit of Davidson [1980b], I understand the Causal Theory as two principles. The first concerns bodily actions:

ACTION. An event involving a creature’s body is an intentional action only if it is non-deviantly caused and rationalized by its mental states (for instance, by its beliefs, desires, or intentions).

The second concerns motivating reasons:

REASONS. When a creature *Ds* intentionally, its reasons for *D*-ing are (or are a function of) the mental states that non-deviantly caused and rationalized its *D*-ing.

These principles constitute a minimal core of the Causal Theory, leaving room for disagreement about how to understand mental actions, actions that are not bodily movements, the ontology of motivating reasons, the nature of non-deviant causation, and so on.

3. A Setiya-Inspired Argument

While not all versions of the Causal Theory include the rationalization condition (or an analogue), I am not aware of any arguments against it. However, related discussions by Setiya [2007, 2011, 2013] include ideas that generate such an argument, and provide a natural frame for my discussion. The Setiya-inspired argument supports two

² I suspect that something like the subset problems motivated Davidson’s original appeal to rationalization (see, especially, [1980c]).

³ Related ideas can be found in Enç [2003].

conclusions—that there is no plausible rationale for including the rationalization condition in the Causal Theory, and, moreover, that the rationalization condition is subject to counterexample. After explaining these, I reply to the first conclusion in section 3, and the second conclusion in sections 4–6.

The first argument proceeds by elimination [Setiya 2007: 61–5]. It argues that each of two natural rationales for the rationalization condition relies on an implausible assumption about the relationship between action and normativity. So, the condition is unmotivated.

The first rationale is the guise-of-the-good thesis (henceforth, GUISE), according to which, whenever someone acts, they act because they see something good about their action. In causal-psychological terms, this means that every intentional action is caused, in part, by a belief that the action would be good (or has some other normative property). Given the assumption that such beliefs rationalize actions, one might see the rationalization condition as a way to incorporate GUISE into the Causal Theory.

Setiya argues (and I agree) that GUISE, so construed, is too strong to be credible.⁴ It requires that every action be accompanied by a normative belief (not a mere ‘appearance of goodness’), and that the belief be present at the time of action (not that the creature be disposed to form it on reflection). Counterexamples to GUISE are widely known [Velleman 2000a; Setiya 2007: 63], and it is rejected even by most proponents of the general thesis.

The second rationale that Setiya considers relies on Davidsonian ideas about motivating reasons and rationality. Here is John McDowell’s formulation of the idea [McDowell 1998: 328; cited at Setiya 2007: 64]:

[The] concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be.

One might be led to the rationalization condition as a way to incorporate this influential idea into the Causal Theory.

Setiya argues convincingly that this idea is problematic. It seems to rule out, what is evidently possible, that agents are sometimes highly irrational (and not even ‘approximate[ly]’ rational). And, crucially, the standard thought experiments appealed to in its support are unpersuasive. Consider Nagel’s example of an agent who, driven by thirst, attempts to put a coin into a pencil sharpener (discussed by Nagel [1970: 34]; cited by Setiya [2007: 64]).⁵ This is highly irrational and perhaps, in a sense, unintelligible. But, as long as the agent is sufficiently rational overall to justify the attribution of intentional states, proponents of the Causal Theory will rightly insist that this particular doing is an intentional action: it is not a reflex or an accident. Given that even highly irrational doings can be actions, McDowell’s Davidsonian idea does not motivate the rationalization condition [Setiya 2007: 64–5]. So, the second rationale for the rationalization condition is no more persuasive than the first.

The second Setiya-inspired argument follows from the argument against McDowell. That argument assumed that highly irrational and seemingly unintelligible actions can be intentional. If (as seems plausible) the rationalization condition requires that every

⁴ I discuss weaker versions of GUISE in [2016].

⁵ Compare also Levin [1988] and Quinn [1993].

action is, to some degree, rational and intelligible, then the rationalization condition is implausible.

4. The Job for the Rationalization Condition

Like Setiya, I reject GUISE and accept that actions can be highly irrational. Nevertheless, I disagree with both parts of the Setiya-inspired argument. In this section, I respond to the first argument by presenting a novel motivation for the rationalization condition: it provides a natural solution to two subset problems that versions of the Causal Theory lacking rationalization have no clear way to solve.

4.1 The Leading Idea

Both problems are generated by an influential idea associated with Anscombe and Davidson. According to the causal-psychological tradition, causation is a relation between events, and so actions are constituted by events. But most causal-psychological theorists deny that the things that we do intentionally are individuated by the events that constitute them. Anscombe and Davidson both argued that an intentional action is an event ‘under a description’, since there may be many descriptions of the event constituting an action under which it is not intentional.⁶ Other philosophers mark this idea in other ways. Some identify intentional actions with properties of the events that constitute them, or with action-types that the token actions instantiate [Goldman 1970: ch. 1]. My argument is general, targeting any view on which intentional actions are individuated in a finer-grained way than events are.

For concreteness, I will make two Davidsonian assumptions, although my argument does not turn on them (as I will illustrate in section 4.3). The first assumption is that actions are constituted by bare events (that is, events not under any particular descriptions), but that intentional actions are individuated by pairs of events and (true) descriptions of them. The second assumption is that events are concrete individuals, individuated by regions of spacetime. On this view, an agent’s action is constituted by the event that consists in the region of spacetime in which it is performing the action.

Begin by considering a version of the Causal Theory without rationalization:

ACTION- An event involving a creature’s body is an intentional action only if it is non-deviantly caused by its mental states (for instance, by its beliefs, desires, or intentions).

REASONS- When a creature *Ds* intentionally, its reasons for *D-ing* are (or are a function of) the mental states that non-deviantly caused its *D-ing*.

My argument will be that, because causation is a relation between bare events, it is too coarse-grained a relation to do some important action-theoretic work. **ACTION-** and **REASONS-** thus need to be enriched. Because rationalization is a relation of the appropriate fineness-of-grain, it is the right sort of relation to do the relevant work.

⁶ This idea is discussed by Davidson [1980a, 1980e] and Anscombe [1981, 2000: 11–12], although Anscombe employs it somewhat differently.

4.2 Subset Problem 1: Identifying the Descriptions under Which an Action Is Intentional

The first problem can be illustrated with an example from Davidson [1980a: 4–5]:

Midnight Prowler. A desires to illuminate their bedroom, believes that flipping the switch is a way to do that, and so flips the switch. In so doing, A illuminates the room, awakens a person sleeping in bed, and alerts a prowler in the backyard.

This case describes a single event that has many descriptions: it is a flipping of the switch, an illuminating of the room, an awakening of a sleeper, and so on. (That each describes the same event is famously supported by the plausibility of Davidson's [1980d] event semantics for change sentences.) Yet, crucially, the action is intentional under only some of those descriptions. A did not alert the prowler intentionally.

The first subset problem is thus the problem of picking out, from the set of all descriptions of this event, the descriptions under which it is intentional. ACTION-alone cannot solve this problem. Causation is a relation between bare events, and all of the relevant descriptions describe the same event. So (extending our language of causation to allow talk of descriptions of events as being caused), anything that caused one caused all of them. So, ACTION- has no way to identify, from the set of all of the descriptions of A's action, those under which it is intentional.

I will say more on how the rationalization condition solves this problem once I introduce my theory of rationalization in section 6. But notice for now that the condition is well-suited to the task. Rationalization is a relation between mental states and actions under descriptions, not bare events, and so it is a relation with the appropriate fineness-of-grain to solve this problem. To illustrate, consider a revised version of ACTION, where E is an event involving A's body, and D is a description of E:

ACTION*. D is an intentional action only if there is a set of A's mental states that non-deviantly caused E and rationalized D.

Plausibly, in *Midnight Prowler*, A's desire to illuminate the room (along with relevant beliefs) 'renders intelligible' their flipping the switch, but not the other descriptions of the action. So, given that gloss of the rationalization condition, ACTION* distinguishes that description of the event from the others, solving the first subset problem.

4.3 Subset Problem 2: Identifying the Mental States that Are Motivating Reasons

The second problem is generated by a familiar observation about motivating reasons. An agent's intentional action is constituted by an event, and that event can be caused (in part) by many of the agent's mental states. Yet, intuitively, not all of those causally efficacious mental states must be (or specify)⁷ the agent's reasons for having acted under any particular description. The second subset problem is the problem of picking out, from the set of the mental states involved in the causal production of the action, the subset that gathers together the agent's reasons for acting under any particular description.

⁷ For simplicity, this section will assume that motivating reasons are mental states, rather than a function of mental states. Nothing in my argument will depend on that assumption.

One way to illustrate the problem is with a case in which an agent does more than one thing intentionally at the same time:⁸

Icy Sidewalk. A walks to the library because they desire to check out a book. A walks on the sunny side of the street, because they want to avoid patches of ice on the shady side.

Intuitively, it seems possible for A's desire to get the book to be their motivating reason for walking to the library, but not their reason for walking on the sunny side of the street. And it seems possible for A's desire to avoid ice to be their reason for walking on the sunny side of the street but not for walking to the library. After all, A's desire to check out the book might not explain A's walking on the sunny side, and A's desire to avoid ice might not explain their walking to the library. Importantly, however, if A's walk to the library happened entirely on the sunny side of the street, those two descriptions describe one event.

The second subset problem is the problem of picking out, from the set of all of the mental states that caused an event (including, in *Icy Sidewalk*, both of the desires mentioned above), which ones are (or represent) the agent's motivating reasons for the action under any given description. It is straightforward to see why REASONS- cannot solve this problem. Since causation is a relation between bare events, REASONS- appears to entail that, necessarily, all of the mental states that caused a particular event are motivating reasons for the action under every description of it. It follows that, in *Icy Sidewalk*, A's desire to avoid patches of ice was one of their reasons for walking to the library. This is problematic.

As above, rationalization is well-suited to solve this problem: it is a relation with the appropriate fineness-of-grain. I give more detail in section 6.2, but, in order to see the idea, consider a revised version of REASONS, where E is an event involving A's body, and D is a description of E:

REASONS*. If D is an intentional action, A's reasons for D are (or are a function of) the mental states

that non-deviantly caused E and rationalized D.

To use the familiar gloss on rationalization, it seems plausible that a desire to avoid ice does not 'render intelligible' walking to the library. REASONS* promises to explain why it is not among A's reasons for walking to the library.

4.4 A More Conservative Solution?

There might be concepts other than rationalization that can be added to the Causal Theory to solve these problems. I do not know of any, although of course such conceptual innovation would be welcome. I want to suggest here, however, that innovation of some kind is needed, as the prospects are dim for a conservative solution that appeals only to concepts in ACTION- and REASONS-.

I am pessimistic because a very natural conservative strategy is a nonstarter. To see why, recall that both problems turn on the observation that causation is too coarse-grained a relation to perform necessary work. A natural conservative strategy is thus to individuate events in a finer-grained way than Davidson did (and I assumed above), so that each description of an action corresponds to (or is identical to) a distinct event. One might, for example, appeal to Kim's theory of events [1976], which

⁸ I discuss a similar case elsewhere [2016].

identifies events with property instantiations (individuated by triples of entities, properties, and times). Kim's theory entails that each description of an action corresponds to a different event, and so allows that, for example, the flipping of the switch might have a different cause than the alerting of the prowler.

Unfortunately, given any plausible theory of causation, this idea goes nowhere. Even if alerting the prowler and flipping the switch are different events, they share common causes. A's alerting the prowler was (in part) caused by their desire to illuminate the room, and A's walking on the sunny side of the street is caused (in part) by their desire to check the book out of the library. This will be true of Davidsonian or non-Davidsonian theories of causation, and even on theories that deny that causation is transitive [Lewis 1973; Schaffer 2005].

This does not show a conservative solution is impossible. But it makes me sufficiently pessimistic that I think it is worth considering whether there is an account of rationalization that avoids the second Setiya-inspired argument. I turn to that question now.

5. A Design Specification for the Rationalization Relation

The lesson of the second Setiya-inspired argument is that the rationalization condition must be qualified in order to be plausible. In this section, I state four constraints that a theory of rationalization (that is, a theory of what it is for a set of mental states to rationalize an action under a description) would have to meet in order to avoid counterexamples. Two derive from the Setiya-inspired argument, and two derive from other platitudes about motivating reasons. Some of the constraints are controversial, but I think that all four are quite plausible, and they are widely enough held in conjunction that it is worth considering whether a theory of rationalization can satisfy all four.

The first Setiya-derived constraint is simply that the theory of rationalization must avoid the very strong guise-of-the-good thesis:

No GUISE. The theory must not entail that one of the causes of any action is a belief with a normative content.

The second is that the theory of rationalization should allow that agents are sometimes highly irrational:

Irrationality. The theory must not entail that highly irrational action is impossible.

The theory of rationalization should allow that highly irrational actions, such as trying to put a coin into a pencil sharpener, can be intentional actions.

The third constraint derives from a platitude about motivating reasons: beliefs, desires, and intentions can (at least partly) explain action, and so can be (or represent) motivating reasons.⁹ Given REASONS*, each type of mental state can stand in the rationalization relation:

Pluralism. The theory should be consistent with beliefs, desires, intentions, and perhaps other mental state types being (or representing) motivating reasons.

This constraint is compatible with the idea that beliefs and desires can only rationalize (and so be motivating reasons) in conjunction with each other. It rules out the idea that

⁹ This is defended by Davidson [1980a]. For further discussion, see Pryor [2007], Setiya [2011], and Fogal [2018].

only desires, only intentions, or only beliefs can be motivating reasons. Note that at least two kinds of beliefs can be motivating reasons—means-ends beliefs (such as the belief that doing something would promote one of one’s goals) and normative or evaluative beliefs (such as the belief that something is good).

The final constraint requires that the theory provide guidance about which mental states rationalize which actions:

Informativeness. The theory should entail specific verdicts about at least some cases.

A theory of rationalization need not be fully worked-out to be promising, but it should give verdicts in at least some paradigm cases. For example, it should entail that, in *Midnight Prowler*, A’s mental states rationalize flipping the switch but not alerting the prowler.

6. Unsatisfactory Theories of Rationalization

Satisfying all four constraints is more difficult than one might expect. I will now argue that the constraints rule out three families of theories of rationalization suggested by the recent literature. Proponents of the Causal Theory who accept the constraints must look elsewhere.

6.1 Rationality Theories

A natural idea is to understand the rationalization relation in terms of the idea of rationality.¹⁰ Let A be a creature, let **M** be a set of A’s mental states, and let D be a description of an event involving A’s body. Then a schematic version of such a theory can be stated as follows:

Rationality Theory. **M** rationalizes D just in case, given that A is in **M**, it is rational (in some sense) for A to decide to do D.

Filling in different conceptions of rationality produces different versions of the theory. Perhaps it is rational for A to decide to do D when, given **M**, D maximizes expected utility, or when, given **M**, D’s expected utility is satisfactorily high.

How does this theory fare? Unfortunately, both the maximizing and satisficing versions of this theory clearly violate the Irrationality Constraint, as each requires that intentional actions not be highly irrational.

Several philosophers have suggested (in personal communications) that one might appeal here to an idea of minimal rationality. Perhaps even highly irrational actions, as in Nagel’s example, manifest a minimal degree of rationality, in some sense. While this idea has some promise, I know of no substantive development of the idea of minimal rationality in the literature, nor do I have an intuitive understanding of it.¹¹ Without a substantive theory of minimal rationality, the minimal Rationality Theory does not entail specific verdicts about cases, and so does not satisfy the Informativeness Constraint. With that being said, as I will discuss in section 6, perhaps my preferred theory of rationality can be understood as one way of taking up this suggestion.

¹⁰ For ideas in this spirit, see Antony [1989: 157] and Arpaly and Schroeder [2014: 55–6].

¹¹ For some suggestive ideas, though, see Smith [2009].

6.2 Motivation Theories

A second kind of theory inspired by the literature associates rationalization with motivation.¹² On this view, all and only desires and intentions to perform actions under particular descriptions rationalize those actions. A schematic version of this theory is as follows:

Motivation Theory. M rationalizes D just in case M is a set of appropriate motivational states directed at A's doing D.

I accept that every action is caused by a motivational state. But the Motivation Theory fails to satisfy the Pluralism Constraint, as it does not allow beliefs to rationalize action.

6.3 Belief Theories

A third kind of theory conceives of rationalization in terms of a relation between the content of a belief and the relevant description of an action:

Belief Theory. M rationalizes D just in case M is a set of appropriate beliefs.

Like the Motivation Theory, the Belief Theory fails to satisfy the Pluralism Constraint, and so is unacceptable. However, three different versions of this theory are suggested by the literature, and each fails in an additional way, which it is useful to consider.

All three types of theories appeal to the idea of a normative reason for action. The first is inspired by Raz [1999]:

Belief Theory 1. M rationalizes D just in case M consists of the belief that there is a normative reason for A to do D.

This theory violates the No GUISE Constraint (which Raz rejects). Versions of Belief Theory 1 appealing to other normative concepts also violate it.

Two other types of theories are more subtle. The first, inspired by Audi [1985], appeals to beliefs whose contents are (or correspond to) normative reasons:

Belief Theory 2. M rationalizes D just in case M consists of the belief that P, where P is (or corresponds to) a normative reason for A to do D.

This theory entails that agents never do actions for which there are no normative reasons at all. That is both inconsistent with the Irrationality Constraint and obviously subject to counterexample, for instance, in cases like Williams' 'gin and petrol' case [1981].

A third theory, inspired by Parfit [2011: 37] and perhaps Dancy [2000: 95–6], attempts to avoid such counterexamples:

Belief Theory 3. M rationalizes D just in case M consists of the belief that P, where, if P were true, P would be (or correspond to) a normative reason for A to do D.

Given Parfit's own theory of rationality, this theory would violate the Irrationality Constraint in addition to the Pluralism Constraint. But, strikingly, it also fails to avoid some counterexamples that it has been thought to avoid. While it capably handles the 'gin and petrol' case, in which a person acts on a false belief that, if

¹² This idea is in the spirit of Goldman [1970: 54–5] and Shepherd [2014].

true, would be a reason, it cannot obviously handle cases like Allan Gibbard [1999]’s case of Caligula. On many influential views [Gibbard 1999, Bratman 2009], when Caligula harms an innocent because he desires pleasure and believes that harming an innocent will bring him pleasure, he has no belief such that, if it were true, it would be a normative reason for him to harm.¹³ Yet his belief and desire rationalize his action.

7. The Davidsonian Theory of Rationalization

Do these arguments support pessimism about one or more of these constraints, or about appeal to the rationalization condition itself? I think not. In this section, I sketch a theory that can satisfy all four constraints. I cannot offer a full elaboration or defence of the theory here, and so my discussion is something of a promissory note: if a theory like this one can be made plausible, then the subset problems provide substantial motivation for the rationalization condition.

The theory of rationalization that I prefer is inspired by Davidson’s early discussion of rationalization [1980c: 85–6]:

When is an action (described in a particular way) reasonable in the light of specific beliefs and pro attitudes? One way to approach the matter is through a rather abstract account of practical reasoning. ... [I]f someone acts with an intention [i.e., acts intentionally], he must have attitudes and beliefs from which, had he been aware of them and had the time, he could have reasoned that his action was desirable ... If we can characterize the reasoning that would serve, we will in effect have described the logical relations between descriptions of beliefs and desires, and the description of an action, when the former give the reasons with which the latter was performed.

In contemporary terms, Davidson’s idea is that mental states rationalize a creature’s action when the creature could have reasoned from those mental states to a decision to perform the action.¹⁴ The agent need not have engaged in the relevant piece of reasoning (or any reasoning whatsoever); the point is that they could have done so. Davidson’s theory can thus be put schematically:

Davidsonian Theory 1. **M** rationalizes **D** just in case **A** could reason from **M** to the decision to do **D**.

Filling out this theory requires specifying what is meant by ‘**A** could reason from **M** to the decision to do **D**’, and it could be spelled out in a variety of different ways.

I will now show how to develop Davidson’s ideas into a theory that can satisfy the four constraints. My version of the Davidsonian Theory draws on a widely discussed conception of practical reasoning.¹⁵ The main idea is that there are norms of reasoning that permit various possible or actual transitions of thought. Norms of theoretical reasoning govern transitions from doxastic states to doxastic states, while norms of practical reasoning govern transitions from sets of mental states (such as pairs of beliefs and desires) to decisions to act. Examples of permissible theoretical reasoning patterns include enumerative induction, statistical inference, abduction, and

¹³ This will be disputed by philosophers, such as Schroeder [2007], who accept this consequence. Those philosophers should nonetheless reject Belief Theory 3, on Pluralism Constraint grounds.

¹⁴ Davidson [1980c: 84] at this time identified pro-attitudes with judgments that specific actions are desirable.

¹⁵ This picture of the reasoning can be found in Broome [2013], as well as McHugh and Way [2016b]. It is also closely associated with the ‘Reasoning View’ about normative reasons: see Hieronymi [2011], Silverstein [2016], McHugh and Way [2016a], and Way [2017]. For my version of the Reasoning View, see Asarnow [2016, 2017].

conjunction introduction, while (as I discuss below), examples of permissible practical reasoning patterns include instrumental reasoning and enkratic reasoning. The norms of both theoretical and practical reasoning are defeasible or non-monotonic, in that they may permit the transition from a set of premise states to a conclusion state without permitting the transition from some strictly larger set of premises states to the same conclusion state.¹⁶ In part because of this, engaging in a permissible pattern of reasoning is necessary but not sufficient for reasoning well, in a more robust sense.¹⁷ One reasons permissibly in this sense, but badly in a more robust sense, when, for example, one infers that the population will resemble a sample, despite evidence that the sample is not representative.

Given this idea, I can state a more precise version of the Davidsonian Theory:

Davidsonian Theory 2. **M** rationalizes **D** just in case the norms of reasoning permit the transition from **M** to a decision to do **D**.

This is one natural interpretation of Davidson's idea, and I argue that it can satisfy all four constraints while solving both subset problems.

7.1 The No GUISE, Pluralism, and Irrationality Constraints

My argument turns on three assumptions about reasoning. Two concern which patterns of practical reasoning are permissible, and both have been defended at length by John Broome [2013: chs. 14, 16]. Together, they allow Davidsonian Theory 2 to satisfy No GUISE and Pluralism.

The first concerns enkratic reasoning:

Enkratic Reasoning Assumption. The norms of practical reasoning permit reasoning from normative beliefs and means-ends beliefs to decisions to act.

On one plausible view, the norms permit reasoning from a belief that you ought to do **E**, and a belief that doing **M** is a way of doing **E**, to a decision to do **M**. But I leave open exactly what kinds of beliefs (normative and instrumental) can be premises in enkratic reasoning.

The second assumption concerns instrumental reasoning:

Instrumental Reasoning Assumption. The norms of practical reasoning permit reasoning from intentions or desires and means-ends beliefs to decisions to act.

Again, there is room for disagreement, but, on one plausible view, the norms permit reasoning from an intention to do **E**, and a belief that doing **M** is a way of doing **E**, to a decision to do **M**.

Given these assumptions, Davidsonian Theory 2 satisfies Pluralism, as it allows that normative beliefs, means-ends beliefs, and desires can rationalize action and thus count as motivating reasons. And it satisfies No GUISE because it counts instrumental reasoning as permissible reasoning.

A third assumption allows Davidsonian Theory 2 to satisfy Irrationality—reasoning permissibly does not guarantee much rationality:

¹⁶ I make this idea precise in [2017].

¹⁷ Other necessary conditions might include, for example, that the agent is responsive to the norms of reasoning [Broome 2013: 245–7], or that the agent does not believe any 'defeaters' for the pattern of reasoning.

Reasoning ≠ Rationality Assumption. It is possible for A to be highly irrational in doing D, even though the event described by D was caused by mental states **M** and the norms of reasoning permit the transition from **M** to a decision to do D.

This assumption reflects the defeasibility of the norms of good reasoning: reasoning well (in this specific sense) does not guarantee being rational, or reaching the best conclusion. Recall the example discussed above, in which a person reasons in a way permitted by the norms (inferring that a population resembles a sample) despite reasoning very badly, in the sense that they ignore evidence that the sample is not representative. Similarly, someone might reason from an intention and a means-ends belief to a decision to take the means to their intended end, despite their possessing other mental states that make it highly irrational for them to take that means.

This assumption allows the Davidsonian Theory to satisfy the Irrationality Constraint. Highly irrational action is possible when the agent engages in a permissible pattern of reasoning, but possesses other mental states that make acting in that way highly irrational. Indeed, some natural interpretations of Nagel's example are along these lines. Perhaps this person irrationally formed the belief that their thirst will be satisfied by putting the coin in the pencil sharpener, despite other beliefs to the contrary. Or perhaps their thirst caused them to irrationally form an intrinsic desire to put the coin in the pencil sharpener, a desire that doesn't cohere with their other mental states [Smith 2004a]. In either case, their action is caused by mental states that instantiate a permissible pattern of reasoning (say, instrumental reasoning from a desire and a means-ends belief), but the agent is highly irrational to act on that belief and desire, given the rest of their mind.

7.2 The Informativeness Constraint

Can the Davidsonian Theory satisfy the Informativeness Constraint? Above, I complained that a minimal version of the Rationality Theory could not do so, at least in the absence of a substantive theory of minimal rationality. Given the three assumptions from above, I argue that Davidsonian Theory 2 can satisfy the constraint.

Begin with the first subset problem. Applied to Midnight Prowler, this is the problem of explaining the difference between the person's flipping of the switch and their alerting the prowler. The rationalization condition, interpreted in terms of the Davidsonian Theory, solves this problem. The person's action, in this case, is caused in part by a desire to illuminate the room and a belief that flipping the switch is a means to doing that. The Davidsonian Theory entails that those mental states rationalize flipping the switch, because the Instrumental Reasoning Assumption guarantees that the norms of reasoning endorse the transition from that belief and that desire to the decision to flip the switch. By contrast, the norms of reasoning do not permit the transition from any of the causally efficacious mental states to a decision to alert the prowler. So, the Davidsonian Theory has an explanation of why *flipping the switch* is a description under which the action is intentional, but *alerting the prowler* is not.

Now turn to the second subset problem. Applied to Icy Sidewalk, this is the problem of explaining why it is possible that the agent's desire to check out the book is (or specifies) their reason for walking to the library, but not for walking on the sunny side of the street, even though those two descriptions describe the same event. The Davidsonian Theory explains how this is possible. Plausibly, the norms of reasoning

permit the transition from the desire to check out the book and an appropriate means belief to the decision to walk to the library, but not to the decision to walk on the sunny side of the street. So, the Davidsonian Theory entails that the desire rationalizes (and so is, or specifies, a motivating reason) for one but not the other. Of course, the Davidsonian Theory is compatible with the possibility that, in some cases, the agent's desire to walk to the library rationalizes walking on the sunny side of the street (if, for instance, their walking was caused in part by a belief that walking on the sunny side of the street is a way to go to the library). But the Davidsonian Theory leaves open the possibility that it is not.

A fully informative version of the Davidsonian Theory would offer a theoretical basis for the three assumptions on which I rely here. However, because the assumptions are widely accepted and (in my view) quite plausible, I think that they allow the Davidsonian Theory to be sufficiently informative for present purposes.

Notably, those who are antecedently sympathetic to the minimal version of the Rationality Theory are welcome to see the Davidsonian Theory as a version of such a theory. While the Reasoning \neq Rationality Assumption ensures that being motivated by states that instantiate a permissible pattern of reasoning is compatible with being highly irrational, it is consistent with the idea that engaging in such a pattern is a kind of (very) minimal rationality. Those who are inclined to understand instantiating a permissible pattern of reasoning as a kind of minimal rationality are welcome to understand the Davidsonian Theory as one way to implement the minimal Rationality Theory.

8. Conclusion

This paper has argued for three conclusions. First, the subset problems motivate enriching the Causal Theory with some concept beyond those of non-deviant causation and the mental states. Second, *pace* Setiya, the rationalization condition can do the job, at least if understood in terms of the Davidsonian Theory. Finally, no other satisfactory theory of rationalization is readily available. In the absence of other solutions to the subset problems, I take these three conclusions to provide substantial support for the 'Davidsonian Causal Theory'—ACTION* and REASONS* supplemented with Davidsonian Theory 2.

I want to conclude this discussion by reflecting on two questions that the Davidsonian Causal Theory raises. The first concerns whether the theory is unacceptable on naturalistic grounds, given that it makes reference to the idea of the norms of reasoning. Many proponents of the causal-psychological theory of action see philosophical theorizing about action as broadly continuous with (other kinds of) scientific theorizing about agency, and see the Causal Theory as a theory, not only of human action, but of the actions of other creatures as well. One might worry that those broadly naturalistic commitments rule out appeal to norms of reasoning.

There are two reasons to think that this objection is unpersuasive. One is simply that it is not obvious that any appeal to normative facts renders a theory naturalistically unacceptable [Davidson 1980a; Gibbard 2012]. Arguably, any scientific enterprise requires appeal to normative facts, and the Davidsonian Causal Theory is compatible with various naturalistic views in metaethics, including quasi-realism and naturalistic reductionism. Second, Schroeder [2003] has persuasively argued that facts about the norms of reasoning should not be understood as normative facts. So, naturalistically

inclined philosophers of action should not reject the Davidsonian Causal Theory too hastily.

The second question concerns the relationship between motivating reasons and normative reasons. Interestingly, the theory of motivating reasons generated by Davidsonian Theory 2 bears some resemblance to a family of theories of normative reasons that analyse facts about normative reasons in terms of the norms of practical reasoning, often called the Reasoning View about normative reasons.¹⁸ According to Davidsonian Theory 2, an agent's motivating reasons for having done D are (or are a function of) those mental states that caused the event that D describes and that could have been the premises of a good piece of reasoning concluding in the decision to do D. According to a simplified version of the Reasoning View about normative reasons, a normative reason for A to do D is the content of a possible mental state in a possible piece of good reasoning concluding in the decision to do D. Philosophers inclined to think that there is a close relationship between normative reasons and motivating reasons may find this a source of appeal for the Reasoning View about normative reasons.¹⁹ Importantly, however, the Davidsonian Theory about rationalization is independent of the Reasoning View about normative reasons, and it can be accepted by philosophers with many different views about normative reasons.²⁰

Disclosure Statement

No potential conflict of interest was reported by the author.

References

- Anscombe, G.E.M. 1981. Under a Description, in her *Metaphysics and the Philosophy of Mind: Collected Papers, Volume II*, Oxford: Basil Blackwell: 208–19.
- Anscombe, G.E.M. 2000. *Intention*, 2nd edn, Cambridge, MA: Harvard University Press.
- Antony, Louise 1989. Anomalous Monism and the Problem of Explanatory Force, *The Philosophical Review* 98/2: 153–87.
- Arpaly, Nomy and Timothy Schroeder 2014. *In Praise of Desire*, New York: Oxford University Press.
- Asarnow, Samuel 2016. Rational Internalism, *Ethics* 127/1: 147–78.
- Asarnow, Samuel 2017. The Reasoning View and Defeasible Practical Reasoning, *Philosophy and Phenomenological Research* 95/3: 614–36.
- Audi, Robert 1985. Rationalization and Rationality, *Synthese* 65/2: 159–84.
- Bratman, Michael E. 2009. Intention, Practical Rationality, and Self-Governance, *Ethics* 119/3: 411–43.
- Broome, John 2013. *Rationality through Reasoning*, Malden, MA: Wiley Blackwell.
- Dancy, Jonathan. 2000. *Practical Reality*, Oxford: Oxford University Press.
- Davidson, Donald 1980a. Actions, Reasons, and Causes, in his *Essays on Actions and Events*, Oxford: Clarendon Press: 3–20.
- Davidson, Donald 1980b. Freedom to Act, in his *Essays on Actions and Events*, Oxford: Clarendon Press: 63–82.
- Davidson, Donald 1980c. Intending, in his *Essays on Actions and Events*, Oxford: Clarendon Press: 83–102.

¹⁸ This is developed by, e.g., Hieronymi [2011], McHugh and Way [2016a], Silverstein [2016], and Way [2017].

¹⁹ I defend this view elsewhere [2016, 2017].

²⁰ For helpful feedback and conversation, I thank Facundo Alonso, Franz Altner, Michael Bratman, Sarah Hannan, Carlos Núñez, Grant Rozeboom, RJ Leland, Samuel Murray, Grace Paterson, Herlinde Pauer-Studer, David Taylor, several anonymous referees, and the editor-in-chief of this journal. Thanks as well to audiences at the University of Manitoba and the University of Vienna, where this material was presented.

- Davidson, Donald 1980d. The Logical Form of Action Sentences, in his *Essays on Actions and Events*, Oxford: Clarendon Press: 105–21.
- Davidson, Donald 1980e. The Individuation of Events, in his *Essays on Actions and Events*, Oxford: Clarendon Press: 163–80.
- Enç, Berent 2003. *How We Act: Causes, Reasons, and Intentions*, Oxford: Clarendon Press.
- Fogal, Daniel 2018. Deflationary Pluralism about Motivating Reasons, in *The Factive Turn in Epistemology*, ed. Veli Mitova, Cambridge: Cambridge University Press: 193–218.
- Gibbard, Allan 1999. Morality as Consistency in Living: Korsgaard's Kantian Lectures, *Ethics* 110/1: 140–64.
- Gibbard, Allan 2012. *Meaning and Normativity*, Oxford: Oxford University Press.
- Goldman, Alvin I. 1970. *A Theory of Human Action*, Englewood Cliffs, NJ: Prentice-Hall.
- Hieronimi, Pamela 2011. Reasons for Action, *Proceedings of the Aristotelian Society* 111/3/3: 407–27.
- Hornsby, Jennifer. 2004. Agency and Actions, in *Agency and Action*, ed. John Hyman and Helen Steward, Cambridge: Cambridge University Press: 1–24.
- Hurley, Paul 2018. Consequentialism and the Standard Story of Action, *The Journal of Ethics* 22/1: 25–44.
- Kim, Jaegwon 1976. Events as Property Exemplifications, in *Action Theory: Proceedings of the Winnipeg Conference on Human Action, Held at Winnipeg, Manitoba, Canada, 9–11 May 1975*, ed. Myles Brand and Douglas Walton, Dordrecht: Springer: 159–77.
- Levin, Janet 1988. Must Reasons Be Rational? *Philosophy of Science* 55/2: 199–217.
- Lewis, David 1973. Causation, *The Journal of Philosophy* 70/17: 556–67.
- McDowell, John. 1998. Functionalism and Anomalous Monism, in his *Mind, Value, and Reality*, Cambridge, MA: Harvard University Press: 325–40.
- McHugh, Conor and Jonathan Way. 2016a. Fittingness First, *Ethics* 126/3: 575–606.
- McHugh, Conor and Jonathan Way 2016b. What Is Good Reasoning?, *Philosophy and Phenomenological Research* 96/1: 153–74.
- Mintoff, Joe 2002. How Can Intentions Make Actions Rational? *Canadian Journal of Philosophy* 32/3: 331–54.
- Moya, Carlos J. 1998. Reason and Causation in Davidson's Theory of Action Explanation, *Critica: Revista Hispanoamericana de Filosofía* 30/89: 29–43.
- Nagel, Thomas 1970. *The Possibility of Altruism*, Oxford: Clarendon Press.
- O'Brien, Lilian 2006. Davidson on Justification and Rationalization, *Daimon Revista de Filosofía* 37: 171–8.
- Parfit, Derek 2011. *On What Matters, Volume One*, ed. Samuel Scheffler, Oxford: Oxford University Press.
- Pryor, James 2007. Reasons and That-Clauses, *Philosophical Issues* 17/1: 217–44.
- Quinn, Warren 1993. Putting Rationality in Its Place, in his *Morality and Action*, ed. Philippa Foot, Cambridge: Cambridge University Press: 228–55.
- Raz, Joseph 1999. Agency, Reason, and the Good, in his *Engaging Reason: On the Theory of Value and Action*, Oxford: Oxford University Press: 22–45.
- Schaffer, Jonathan 2005. Contrastive Causation, *The Philosophical Review* 114/3: 327–58.
- Schlosser, Markus E. 2007. Basic Deviance Reconsidered, *Analysis* 67/3: 186–94.
- Schlosser, Markus E. 2011. Agency, Ownership, and the Standard Theory, in *New Waves in Philosophy of Action*, ed. Jesús H. Aguilar, Andrei A. Buckareff, and Keith Frankish, Basingstoke: Palgrave Macmillan: 13–31.
- Schroeder, Mark 2007. *Slaves of the Passions*, Oxford: Oxford University Press.
- Schroeder, Timothy 2003. Donald Davidson's Theory of Mind Is Non-Normative, *Philosophers' Imprint* 3/1: 1–14.
- Setiya, Kieran 2007. *Reasons without Rationalism*, Princeton: Princeton University Press.
- Setiya, Kieran 2011. Reasons and Causes, *European Journal of Philosophy* 19/1: 129–57.
- Setiya, Kieran 2013. Causality in Action, *Analysis* 73/3: 501–12.
- Shepherd, Joshua 2014. Causalism and Intentional Omission, *American Philosophical Quarterly* 51/1: 15–26.
- Silverstein, Matthew 2016. Reducing Reasons, *Journal of Ethics & Social Philosophy* 10/1: 1–22.
- Smith, Michael 2004a. The Possibility of Philosophy of Action, in his *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics*, Cambridge: Cambridge University Press: 155–77.

- Smith, Michael 2004b. The Structure of Orthonomy, in *Agency and Action*, ed. John Hyman and Helen Steward, Cambridge: Cambridge University Press: 165–194.
- Smith, Michael 2009. The Explanatory Role of Being Rational, in *Reasons for Action*, ed. David Sobel and Steven Wall, Cambridge: Cambridge University Press: 58–80.
- Smith, Michael 2012. Four Objections to the Standard Story of Action (and Four Replies), *Philosophical Issues* 22: 387–401.
- Stout, Rowland 2011. Deviant Causal Chains, in *A Companion to the Philosophy of Action*, ed. Timothy O'Connor and Constantine Sandis, Malden, MA: Wiley-Blackwell: 159–65.
- Velleman, J. David 2000. What Happens When Someone Acts?, in his *The Possibility of Practical Reason*, Oxford: Clarendon Press: 123–43.
- Wald, Benjamin and Sergio Tenenbaum 2018. Reasons and Action Explanation, in *The Oxford Handbook of Reasons and Normativity*, ed. Daniel Star, Oxford: Oxford University Press: 214–32.
- Way, Jonathan 2017. Reasons as Premises of Good Reasoning, *Pacific Philosophical Quarterly* 98/2: 251–70.
- Wedgwood, Ralph 2006. The Normative Force of Reasoning, *Noûs* 40/4: 660–86.
- Williams, Bernard 1981. Internal and External Reasons, in his *Moral Luck: Philosophical Papers 1973–1980*, Cambridge: Cambridge University Press: 101–13.