

# Constitutivism about Instrumental Desire and Introspective Belief

RYAN COX

doi:10.48106/dial.v74.i4.02

Ryan Cox. 2020. "Constitutivism about Instrumental  
Desire and Introspective Belief." *Dialectica* 74(4): 635–664.  
doi:10.48106/dial.v74.i4.02.



# Constitutivism about Instrumental Desire and Introspective Belief

RYAN COX

This essay is about two familiar theses in the philosophy of mind: constitutivism about instrumental desires, and constitutivism about introspective beliefs, and the arguments for and against them. Constitutivism about instrumental desire is the thesis that instrumental desires are at least partly constituted by the desires and means-end beliefs which explain them, and is a thesis which has been championed most prominently by Michael Smith. Constitutivism about introspective belief is the thesis that introspective beliefs are at least partly constituted by the mental states they are about, and is a thesis which has been championed most prominently by Sydney Shoemaker. Despite their similarities, the fortunes of these two theses could not be more opposed: constitutivism about instrumental desire is widely accepted, and constitutivism about introspective belief is widely rejected. Yet, the arguments for both theses are roughly analogous. This essay explores these arguments. I argue that the argument which is widely taken to be the best argument for constitutivism about instrumental desires—what I call the argument from necessitation—does not provide the support for the thesis it is widely taken to provide, and that it fails for much the same reasons that it fails to provide support for constitutivism about introspective belief. Furthermore, I argue that the best argument for constitutivism about instrumental desires—what I will call the argument from cognitive dynamics—is also a good argument, if not equally good, for constitutivism about introspective belief (at least when the thesis is suitably qualified).

This essay is about two familiar theses in the philosophy of mind: *constitutivism about instrumental desire*, and *constitutivism about introspective belief*, and the arguments for and against them. Constitutivism about instrumental desire is the thesis that instrumental desires are at least partly constituted by the desires and means-end beliefs which explain them and is a thesis which has been championed most prominently by Michael Smith (2004). Consti-

tutivism about introspective belief is the thesis that introspective beliefs are at least partly constituted by the mental states they are about and is a thesis which has been championed most prominently by Sydney Shoemaker (1996, 2012). Despite their similarities, the fortunes of these two theses could not be more opposed: constitutivism about instrumental desire is widely accepted, and constitutivism about introspective belief is widely rejected. Yet, the arguments for both theses are roughly analogous. So if one thesis is to be accepted while the other is rejected there must be good reasons for rejecting the arguments for one thesis but not the other.

In this essay, I argue that the *best* argument for constitutivism about instrumental desire—what I will call the *argument from cognitive dynamics*—is also a *good* argument, if not an *equally good* argument, for constitutivism about introspective belief (at least when the thesis is suitably qualified). So, at least with respect to this argument, there are no good reasons for accepting one thesis while rejecting the other. At the same time, however, I argue that the argument which is widely taken to be the best argument for constitutivism about instrumental desire—what I will call the *argument from necessitation*—does not provide the support for the thesis it is widely taken to provide, and that it fails for much the same reasons that it fails to provide support for constitutivism about introspective belief. So, with respect to this argument, there are no good reasons for accepting one thesis while rejecting the other, because this argument does not give us good reasons for accepting either thesis.

These conclusions suggest that the fortunes of constitutivism about instrumental desire and constitutivism about introspective belief are more closely tied together than is often appreciated. Indeed, I hope to bring to bear on the topic of instrumental desire an important lesson which has been learnt in the philosophy of introspection. For philosophers of introspection have shown that the argument from necessitation for introspective belief is unsuccessful. This partly explains why constitutivism about introspective belief is not widely accepted in the way constitutivism about instrumental desire is. Yet an analogous lesson for the argument from necessitation for constitutivism about instrumental desire has not yet been absorbed by those working on the philosophy of instrumental desire. There is a certain irony here, since Shoemaker explicitly draws analogies with the case of instrumental desire—particularly the role played by means-end beliefs and non-instrumental desires in rationalising and explaining further desires—in developing his arguments for constitutivism about introspective belief (1996). Rejecting this argument for

constitutivism about instrumental desire puts this thesis on a less firm footing, and brings its fortunes more closely into line with those of constitutivism about introspective belief. As I will argue, the fortunes of both theses rest on the prospects of the argument from cognitive dynamics. I hope to show that, even if the argument from cognitive dynamics is taken to offer adequate support for constitutivism about instrumental desire while ultimately not offering adequate support for constitutivism about introspective belief, the fortunes of these two theses are tied closer together than is often appreciated.

The essay is structured as follows. Section 1 discusses some preliminary issues concerning how constitutivism about instrumental desire and constitutivism about introspective belief are to be understood. Section 2 considers the argument from necessitation for constitutivism about instrumental desire and the argument from necessitation for constitutivism about introspective belief respectively. I argue that both arguments fail. Section 3 considers the argument from cognitive dynamics for constitutivism about instrumental desire and the argument from cognitive dynamics for constitutivism about introspective belief respectively. I argue that the arguments provide equally good reasons for accepting both theses.

## 1 Locating the Topic

At the outset, we are going to need a way of understanding instrumental desires and introspective beliefs which does not prejudice the case either for or against the respective constitutivist theses. It might be thought, after all, that there isn't much to say about constitutivism about instrumental desire, since instrumental desires just are, by definition, according to some, those desires which are at least partly constituted by the desires and means-end beliefs which explain them. While the term "instrumental desire" is sometimes used this way, there is a more neutral way of understanding instrumental desires which does not prejudice the issue in this way. We can simply say that instrumental desires just are, by definition, those desires which are *rationaly explained* by other desires and means-end beliefs, where we leave open the question of whether the former is partly constituted by the latter. This neutral understanding of instrumental desire is widespread in the literature (Marks 1986, 9; Davis 1986, 69; Schroeder 2004, 5; McDaniel and Bradley 2008, 286; Arpaly and Schroeder 2014, 6). It is arguably this understanding which Hume has in mind when he writes:

Ask a man *why he uses exercise*; he will answer *because he desires to keep his health*. If you then enquire, *why he desires to keep his health*, he will readily reply, *because his sickness is painful*. If you push your enquiries farther, and desire a reason *why he hates pain*, it is impossible he can ever give any. This is an ultimate end, and is never referred to any other object. (Hume EPM, Appendix I)<sup>1</sup>

In this way, Hume distinguishes between what we would call the instrumental desire to exercise, rationally explained by a desire to keep one's health, and the non-instrumental desire to avoid pain, not explained by any further desire. When Smith writes "I will call the desires which are explained by [non-instrumental] desires and means-end beliefs 'instrumental' desires" (2004, 95) he is clearly stating that he means to use "instrumental desires" in this neutral way.<sup>2</sup>

On this way of understanding instrumental desires, it is clearly a substantive question whether instrumental desires are partly constituted by means-end beliefs and other desires. Of course, many theorists accept this substantive thesis about instrumental desires. Here is Smith's assertion of his commitment to constitutivism about instrumental desire:

Instrumental desires are not distinct from the non-instrumental desires and means-end beliefs that explain them, but are rather just the complex state of having such non-instrumental desires and means-end beliefs standing in a suitable relation. (2004, 96)

According to Smith, then, instrumental desires are (at least partly, if not wholly) constituted by non-instrumental desires and means-end beliefs standing in suitable relations. As he puts it: "Instrumental desires are thus better thought of as being nothing over and above the non-instrumental desires and means-end beliefs that explain them" (2004, 96).<sup>3</sup> Here Smith appears to commit himself to a strong form of constitutivism about instrumental desire, one which holds that instrumental desires are "nothing over and above" or are *wholly constituted* by the non-instrumental desires and means-end beliefs that

---

1 Quoted in Smith (2004, 94).

2 See also "[I]nstrumental desires are those that can be explained by non-instrumental desires and means-end beliefs" (Smith 2004, 96–97).

3 The qualification that the non-instrumental desire and the means-end belief must stand in a suitable relation is missing here but we can assume that it is intended.

explain them standing in a suitable relation. Constitutivism about instrumental desire, so understood, is a widely accepted thesis in the philosophy of mind. However, as we have just seen, it is a substantive thesis about instrumental desires, and we will be concerned with the arguments for it in the next two sections.

Introspective beliefs raise their own unique problems for understanding. While there is no temptation in the case of introspective beliefs to hold that they are, by definition, beliefs which are partly constituted by the mental states they are about, it is nonetheless difficult to say what an introspective belief is. It is not enough to say that an introspective belief is just a belief about one's own mental states. For, it is relatively uncontroversial that at least some of our beliefs about our own mental states are arrived at on the basis of inference. And these beliefs are very plausibly thought of as being constitutively distinct from the mental states they are about. They are no more constituted by the mental states they are about than our beliefs about the mental states of others are constituted by the mental states they are about. Rather, introspective beliefs must be understood as beliefs about our own mental states which are arrived at by some special means and are not based on evidence or observation in the way that our beliefs about the mental states of others are.<sup>4</sup> While it is a matter of controversy how introspective beliefs are to be understood, I will assume that some such distinction can be drawn among beliefs about our own mental states and that only some of these beliefs will count as introspective beliefs. Constitutivism about introspective belief holds that *these* beliefs are at least partly constituted by the mental states they are about. Here is Shoemaker's assertion of his commitment to constitutivism about introspective belief:

What I am inclined to say is that second-order belief, and the knowledge it typically embodies, is supervenient on first order beliefs and desires—or rather, it is supervenient on these plus a certain degree of rationality, intelligence, and conceptual capacity. By this I mean that one has the former *in* having the latter—that having the former is nothing over and above having the latter. (1996, 34)

---

<sup>4</sup> For an influential discussion of these features of introspective beliefs see Moran (2001). For scepticism about the existence of introspective beliefs so understood see Cassam (2014).

According to Shoemaker, then, introspective beliefs are at least partly (if not wholly) constituted by the mental states they are about together with “a certain degree of rationality, intelligence, and conceptual capacity”—having the former is “nothing over and above” having the latter. This passage anticipates the argument to be examined in the next section, as Shoemaker moves here from a claim about supervenience or necessitation to a claim about constitution. So we have seen that constitutivism about introspective belief is a substantive thesis about introspective belief. It will be the aim of the next sections of the essay to evaluate arguments for both forms of constitutivism.

Before turning to those arguments a final qualification is in order. In the passage from Shoemaker just quoted, Shoemaker is concerned with a kind of constitutivism about introspective belief which concerns introspective beliefs about *attitudes* in particular. For the most part, in what follows, I will be concerned with forms of constitutivism about introspective belief which are restricted in this way, holding that introspective beliefs about our *attitudes* are partly constituted by the attitudes they are about. Introspective beliefs about phenomenally conscious states raise further issues that I will not be able to address here, and while constitutivism about these introspective beliefs may be defensible, different arguments may be required.<sup>5</sup> There is also an interesting question about how constitutivism might be extended to states which are plausibly thought of as having both cognitive and non-cognitive components, like emotions. While I think that constitutivism can be defended for a wide range of mental states, I will largely set aside such an exploration here, and will focus on the particular case of introspective beliefs about our own attitudes. I will also set aside the difficult question of whether we can or should expect a uniform account of introspection and introspective beliefs which applies to all mental states.<sup>6</sup> Finally, related to these questions is the question of which theories of introspection—theories of the means by which we arrive at introspective beliefs—are compatible with constitutivism about introspective belief and those which are not. Some theories of introspection have implications for the relation between introspective beliefs and the mental states they are about. The self-scanning theory of David Armstrong (1968) has the implication that introspective beliefs are constitutively distinct from the states they are about since the former are caused by the latter. Other theories are neutral about the relation and while they may provide causal

---

<sup>5</sup> See Chalmers (2010) for a discussion of phenomenal beliefs.

<sup>6</sup> See Boyle (2009) and Byrne (2011) for discussion.

explanations of the means by which we arrive at introspective beliefs—by means of answering deliberative questions (Moran 2001), or by means of an ascent-routine (Gordon 1986)—they remain compatible with the possibility that introspective beliefs are partly constituted by the attitudes they are about. I will also set aside this question, as the arguments we will consider for constitutivism about introspective belief do not presuppose any particular theory of introspection.

## 2 The Arguments from Necessitation

In this section I will formulate and evaluate the argument from necessitation for constitutivism about instrumental desire and the argument from necessitation for constitutivism about introspective belief. The arguments belong to a family of arguments, arguments from necessitation or supervenience, which are familiar enough across many areas of philosophy. The crucial step in such arguments is a move from a claim about metaphysical necessitation or supervenience, to a claim about constitution. I will first formulate and motivate each argument and then turn to evaluation.

### 2.1 *For Constitutivism About Instrumental Desire*

For simplicity, we can formulate and motivate the argument from necessitation for constitutivism about instrumental desire by focussing on an arbitrary example which we can take to reveal something general about instrumental desires. Suppose, then, that Jane desires to exercise because she desires to keep her health, and believes that exercising is a means to keeping her health. It follows from our understanding of instrumental desires, and the assumption that this is the “because” of rational explanation, that Jane instrumentally desires to exercise. Why think that Jane’s desiring to exercise is at least partly constituted by her desiring to keep her health and her believing that exercising is a means to keeping her health? The argument from necessitation proceeds in two steps.

The first step establishes that the relation between Jane’s desiring to keep her health, her believing that exercising is a means to keeping her health, *her being fully rational*, and her desiring to exercise is not a merely contingent relation, but is, in some sense, necessary. We can bring this out by reflecting on a claim about necessity like the following:



$N_1$ . Necessarily, if Jane desires to keep her health, believes that exercising is a means to keeping her health, and is fully rational, then she desires to exercise.

I will have more to say about the relevant understanding of “is fully rational” in this claim below, and will connect this to Smith’s claims about means-end beliefs and non-instrumental desires standing in suitable relations. But for now, we can simply observe that on a natural understanding of “is fully rational” such a claim is intuitively plausible. To bring this out we might notice that while it is certainly possible for Jane to desire to keep her health, to believe that exercising is a means to keeping her health, while not desiring to exercise to any degree—after all, she might be less than fully rational—it is not possible for her to desire to keep her health, to believe that exercising is a means to keeping her health, and not desire to exercise to any degree *if she is fully rational*.

The second step establishes the best explanation of the necessary connection is that Jane’s desiring to exercise is at least partly, if not wholly, constituted by her desiring to keep her health, her believing that exercising is a means to keeping her health, and her being fully rational. While it might be tempting to move directly from the claim about necessitation to this conclusion, it is important to see that there are alternative explanations of the necessary connection which need to be considered. The explanations of the necessary connection we need to consider in this case are these:

$H_1$ . Jane’s desiring to exercise is wholly constituted by her desiring to keep her health, her believing that exercising is a means to keeping her health, and her being fully rational.

$H_2$ . Jane’s desiring to keep her health, her believing that exercising is a means to keeping her health, and her being fully rational is partly constituted by her desiring to exercise.

$H_3$ . There is something that Jane’s desiring to exercise is wholly constituted by and which Jane’s desiring to keep her health, her believing that exercising is a means to keeping her health and her being fully rational is partly constituted by.

These explanations correspond to the familiar options for explaining necessary connections of this form. If there is a necessary connection between something's being an *F* and its being both a *G* and an *H*, then we might explain this necessary connection by holding that its being an *F* is wholly constituted by its being a *G* and an *H*, or by holding that its being a *G* and an *H* is partly constituted by its being an *F*, or by holding that there is something else, such that its being a *G* and an *H* is partly constituted by and its being an *F* is wholly constituted by.

Of these explanations,  $H_1$  certainly looks to be the best. There is no candidate for the kind of third-factor required by  $H_3$ , and, at least initially, it is hard to see why Jane's desiring to keep her health, her believing that exercising is a means to keeping her health, and her being fully rational, would be partly constituted by her desiring to exercise: none of these conditions seem to be independently partly constituted by her desiring to exercise, and it is hard to see how, jointly, they could be partly constituted by her desiring to exercise. So we may tentatively conclude that  $H_1$  provides the best explanation of the necessary connection.

These two steps, then, provide motivation for the two premises of the argument from necessitation for constitutivism about instrumental desire. We can think of the argument as proceeding as follows:

- $P_1$  There is a necessary connection between (i) Jane's desiring to exercise and (ii) her desiring to keep her health, her believing that exercising is a means to keeping her health, and her being fully rational.
- $P_2$  The best explanation of this necessary connection is that Jane's desiring to exercise is wholly constituted by her desiring to keep her health, her believing that exercising is a means to keeping her health, and her being fully rational.
- $C_1$  Jane's desiring to exercise is wholly constituted by her desiring to keep her health, her believing that exercising is a means to keeping her health, and her being fully rational.

Since the example we have focused on here was entirely arbitrary, the same reasoning can be followed in arguing for constitutivism about instrumental desire as a general thesis. At least initially, then, the argument from necessitation provides a good case for constitutivism about instrumental desire. I will now explicate an analogous argument for constitutivism about introspective belief before turning to objections.

## 2.2 *For Constitutivism About Introspective Belief*

For simplicity, we can formulate and motivate the argument from necessitation for constitutivism about introspective belief by focusing on an arbitrary example which we can take to reveal something general about introspective beliefs. Suppose that Jane introspectively believes that she believes it is about to rain. Why think that Jane's believing that she believes it is about to rain is not constitutively distinct from her believing that it is about to rain? The argument proceeds in two steps.

The first step establishes that the relation between Jane's believing that she believes that it is about to rain and her believing that it is about to rain, her having some interest in the question of whether she believes that it is about to rain, and her being fully rational, is not a merely contingent relation, but is, in some sense, necessary. We can bring this out by reflecting on a claim about necessity like the following:

N<sub>2</sub>. Necessarily, if Jane believes that it is about to rain, understands and has some interest in the question of whether she believes that it is about to rain, and is fully rational, then she believes that she believes that it is about to rain.

A few clarifications are in order here. Recall Shoemaker's claim that introspective beliefs are supervenient on "a certain degree of rationality, intelligence, and conceptual capacity." This suggests a claim like the following: necessarily, if Jane believes that it is about to rain, has a certain degree of rationality, intelligence, and conceptual capacity, she believes that it is about to rain. Shoemaker adds these claims about intelligence and conceptual capacity here in order to avoid problems stemming from small children and animals who may have the relevant attitudes, have a certain degree of rationality, and yet not even be able to understand the question of whether they have the attitudes in question. I have captured this element of Shoemaker's view with the claim about understanding in N<sub>2</sub>. However, I have added the further claim that Jane must *have some interest in the question of whether she believes that it is about to rain*. Arguably, Shoemaker's conditions are too weak. It seems to be possible for Jane to believe that it is about to rain, for her to understand the question of whether she believes that it is about to rain, for her to be fully rational, and yet for her not to believe that she believes that it is about to rain *if she*

*has no interest in the question of whether she believes that it is about to rain.*<sup>7</sup> Quite generally, it seems that we will not believe that we have some attitude or another when we have no interest in the question of whether we have that attitude. With these clarifications in order, we can see that this claim is intuitively plausible on a natural understanding of “is fully rational.” To bring this out we might notice that while it is certainly possible for Jane to believe that it is about to rain, to have some interest in the question of whether she believes that it is about to rain, while not believing that she believes that it is about to rain—after all, she might be less than fully rational—it is not possible for her to believe that it is about to rain, to have some interest in the question of whether she believes that it is about to rain, and to not believe that she believes that it is about to rain *if she is fully rational*.

The second step in the argument establishes that the best explanation of this necessary connection is that Jane’s believing that she believes that it is about to rain is at least partly, if not wholly, constituted by her believing that it is about to rain, her having some interest in the question of whether she believes that it is about to rain, and her being fully rational. Again, we must consider the alternative explanations. The explanations of the necessary connections we need to consider in this case are these:

H<sub>1</sub>. Jane’s believing that she believes that it is about to rain is wholly constituted by her believing that it is about to rain, her understanding and taking an interest in the question of whether she believes that it is about to rain, and her being fully rational.

H<sub>2</sub>. Jane’s believing that it is about to rain, her understanding and taking an interest in the question of whether she believes that it is about to rain, and her being fully rational is partly constituted by her believing that she believes that it is about to rain.

H<sub>3</sub>. There is something that Jane’s believing that she believes that it is about to rain is wholly constituted by and which Jane’s believing that it is about to rain, her understanding and having an interest in the question of whether she believes that it is about to rain, and her being fully rational is partly constituted by.

---

<sup>7</sup> See Stoljar (2019) for a discussion of further ways of refining such claims. For the purposes of the argument, all that matters is that there is some non-trivial, finite, mental condition which necessitates the introspective belief.

Of these explanations,  $H_1$  certainly looks to be the best. There is no candidate for the kind of third-factor required by  $H_3$ , and, at least initially, it is hard to see why Jane's believing that it is about to rain, her understanding and taking an interest in the question of whether she believes that it is about to rain, and her being fully rational, would be partly constituted by her believing that she believes that it is about to rain: none of these conditions seem to be independently partly constituted by her believing that she believes that it is about to rain, and it is hard to see how, jointly, they could be partly constituted by her believing that she believes that it is about to rain. So we may tentatively conclude that  $H_1$  provides the best explanation of the necessary connection.

These two steps, then, provide motivation for the two premises of the argument from necessitation for constitutivism about introspective belief. We can think of the argument as proceeding as follows:

- $P_1$  There is a necessary connection between (i) Jane's believing that she believes that it is about to rain and (ii) her believing that it is about to rain, her understanding and having some interest in the question of whether she believes that it is about to rain, and her being fully rational.
- $P_2$  The best explanation of this necessary connection is that Jane's believing that she believes that it is about to rain is wholly constituted by her believing that it is about to rain, her understanding and having some interest in the question of whether she believes that it is about to rain, and her being fully rational.
- $C_1$  Jane's believing that she believes that it is about to rain is wholly constituted by her believing that it is about to rain, her understanding and having some interest in the question of whether she believes that it is about to rain, and her being fully rational.

Since the example we have focused on here was entirely arbitrary, the same reasoning can be followed in arguing for constitutivism about introspective belief as a general thesis. At least initially, then, the argument from necessitation provides a good case for constitutivism about introspective belief.

### 2.3 *Evaluating the Arguments*

We can now turn to the evaluation of the arguments from necessitation. Whether the arguments are successful turns crucially on how the notion of rationality is understood. So far I have presented the arguments without com-

ment on how rationality is to be understood. I will now argue that there are two understandings of “being fully rational” which are relevant to the arguments from necessitation: an evaluative sense and a dispositional sense. When the arguments are understood in terms of the former, their first premises are true, but their second premises are false. When the arguments are understood in terms of the latter, their first premises are false. The arguments get whatever force they have from equivocating on these two understandings of “being fully rational.” This objection to the arguments from necessitation is due, in its essentials, to Amy Kind, who makes the objection in connection with the argument from necessitation for constitutivism about introspective belief (2003).<sup>8</sup> While the objection has been generally appreciated in the philosophical literature on introspection (Gertler 2010), it has not been generally appreciated in the philosophical literature on instrumental desires.<sup>9</sup>

It is natural to think about the arguments above on an evaluative understanding of “being fully rational.” On this understanding, someone is fully rational if and only if they are not in violation of the principles of rationality, that is, if and only if they fully conform to the principles of rationality. On this understanding, someone is less than fully rational if they do not fully conform to the principles of rationality. It is very plausible that there is a principle of rationality which requires you to desire the means if you desire some end and believe that the means are a means to that end. Similarly, it is very plausible that there is a principle of rationality which requires you to believe that you have some attitude if you have that attitude and you understand and have some interest in the question of whether you have it. This can be brought out by reflecting on the necessitation claims with this understanding of “being fully rational” made fully explicit:

$N_1'$ . Necessarily, if Jane desires to keep her health, believes that exercising is a means to keeping her health, and fully conforms to the principles of rationality, then she desires to exercise.

$N_2'$ . Necessarily, if Jane believes that it is about to rain, understands and has some interest in the question of whether she believes that

---

8 While the objection, in its essentials, is due to Kind, the specific development of the objection made here is original to this essay.

9 In both Kind's (2003) and Gertler's (2010) discussions, there is an appeal to causation, and a contrast between causation and constitution, which is not made in the presentation of the objection here.

it is about to rain, and fully conforms to the principles of rationality, then she believes that she believes that it is about to rain.

Since it is arguably this understanding of “being fully rational” which we evaluated the original necessitation claims with, it is not surprising that they both appear to be plausible when this understanding is made fully explicit.

The problem arises for the arguments from necessitation when we turn to the evaluation of their second premises on this understanding. To see the problem in the case of instrumental desire, notice that if Jane desires to keep her health and believes that exercising is a means to keeping her health, then, in order to fully conform to the principles of rationality, she must desire to exercise. But then, given that she desires to keep her health and believes that exercising is a means to keeping her health, *if* she fully conforms to the principles of rationality, this must be at least partly because she desires to exercise. So, her fully conforming to the principles of rationality is partly constituted by her desiring to exercise. And if her fully conforming to the principles of rationality is partly constituted by her desiring to exercise, her desiring to exercise cannot be even partly constituted by her conforming to the principles of rationality. So, on this understanding of “being fully rational,”  $H_1$  is not the best explanation of the necessary connection and the argument from necessitation fails.

To see the problem in the case of introspective belief, notice that if Jane believes that it is about to rain, and understands and has some interest in the question of whether she believes that it is about to rain, then, in order to fully conform to the principles of rationality, she must believe that she believes that it is about to rain. But then, given that she believes that it is about to rain, and understands and has some interest in the question of whether she believes that it is about to rain, *if* she fully conforms to the principles of rationality, this must be at least partly because she believes that she believes that it is about to rain. So, her fully conforming to the principles of rationality is partly constituted by her believing that she believes that it is about to rain. And if her fully conforming to the principles of rationality is partly constituted by her believing that she believes that it is about to rain, her believing that she believes that it is about to rain cannot be even partly constituted by her fully conforming to the principles of rationality. So on this understanding  $H_1$  is not the best explanation of the necessary connection and the argument from necessitation fails.

An analogy might help to drive home the crucial point here. Suppose that you are legally required to pay your taxes before the end of the financial year every financial year. Then, you fully conform to the law only if you pay your taxes before the end of the financial year every year. Now, if it is the end of the financial year, then *if* you fully conform to the law, this must be at least partly because you have paid your taxes. So, your fully conforming to the law is partly constituted by your paying your taxes. And if your fully conforming to the law is partly constituted by your paying your taxes, then your paying your taxes cannot be, even partly, constituted by your fully conforming to the law. As this analogy demonstrates, while conforming to certain norms may, along with other conditions, necessitate certain further conditions, it is your conforming to the norms which is partly constituted by those further conditions, and not vice versa. It is not surprising then that on the evaluative understanding of “being fully rational” the second premises of the arguments from necessitation are false.

At this stage a proponent of the arguments from necessitation may argue that this is not the understanding of “being fully rational” they had in mind. They may instead appeal to a dispositional understanding of “being fully rational.” On this understanding, being fully rational is being disposed to conform to the principles of rationality (perhaps along with there being no barrier to one’s manifesting this disposition). Someone who was fully disposed to conform to the principles of rationality, where there is no barrier to their manifesting this disposition, would very plausibly *come to desire* to exercise if they desired to keep their health and believed that exercising was a means to keeping their health. Someone who was fully disposed to conform to the principles of rationality, where there was no barrier to their manifesting this disposition, would very plausibly *come to believe* that they believed that it was about to rain if they believed that it was about to rain and understood and had some interest in the question of whether they believed that it was about to rain. The trouble with this understanding is that the first premises of the arguments from necessitation seem to be false if they are understood in terms of it. To see the trouble, we can reflect on the necessitation claims with this understanding of “being fully rational” made fully explicit:

$N_1$  ". Necessarily, if Jane desires to keep her health, believes that exercising is a means to keeping her health, is disposed to conform to the principles of rationality, and there is no barrier to the manifestation of her disposition, then she desires to exercise.



$N_2''$ . Necessarily, if Jane believes that it is about to rain, understands and has some interest in the question of whether she believes that it is about to rain, is disposed to conform to the principles of rationality, and there is no barrier to the manifestation of her disposition, then she believes that she believes that it is about to rain.

The problem here is that even when all barriers to the manifestation of a disposition are removed, there is no metaphysically necessary connection between having the disposition, the triggering conditions for the disposition obtaining, and the disposition manifesting. At best, the relation between having a disposition, the triggering conditions for that disposition obtaining, and the manifestation of the disposition, is one of *nomological necessity*. On this understanding of “being fully rational” the first premises of the arguments from necessitation are false, and there is no necessary connection between being in certain psychological conditions, being fully rational, and being in some further psychological condition.

It might be thought at this point that we might add some further condition to the antecedents of the conditionals above to avoid this problem. My response to this suggestion, however, is that there is nothing that can be added that will not either (i) fall prey to considerations like those just given against the current proposal or (ii) fall prey to considerations like those given against the evaluative understanding of “being fully rational.” Suppose someone suggests that we need only add to the antecedents of the conditionals that the dispositions be manifested. Then we would have:

$N_1'''$ . Necessarily, if Jane desires to keep her health, believes that exercising is a means to keeping her health, is disposed to conform to the principles of rationality, and manifests this disposition, then she desires to exercise.

$N_2'''$ . Necessarily, if Jane believes that it is about to rain, and has some interest in the question of whether she believes that it is about to rain, is disposed to conform to the principles of rationality, and manifests this disposition, then she believes that she believes that it is about to rain.

But this suggestion faces a problem analogous to that faced by the view which understands “being fully rational” in the evaluative sense. Given that Jane

desires to keep her health, believes that exercising is a means to keeping her health, and is disposed to conform to the principles of rationality, *if* she manifests her disposition, this must be in part because she desires to exercise. To manifest this disposition is, in part, to come to desire to exercise. So, her manifesting this disposition is partly constituted by her desiring to exercise. And if her manifesting this disposition is partly constituted by her desiring to exercise, her desiring to exercise cannot be even partly constituted by her manifesting this disposition (and so cannot be wholly constituted by her desiring to keep her health, believing that exercising is a means to keeping her health, being disposed to conform to the principles of rationality, and manifesting this disposition). I suspect that Smith's appeal to the idea that a means-end belief and a non-instrumental desire must stand "in a suitable relation" in his statement of constitutivism about instrumental desire is an attempt to straddle the gap here between the triggering conditions of a disposition obtaining and the disposition manifesting. As we have just seen, however, if we understand the claim in terms of the triggering conditions being met—the means-end belief and the non-instrumental desire are appropriately related just when they trigger the relevant disposition—then Smith's position falls prey to considerations like those given against the dispositional understanding, and if we understand it in terms of the disposition manifesting—the means-end belief and the non-instrumental desire are appropriately related just when the relevant disposition manifests—then it falls prey to the considerations just given.

As initially compelling as the arguments from necessitation may seem, they are ultimately unsuccessful. As I said earlier, this is old news in the literature on introspective belief. There, it is widely conceded that Sydney Shoemaker may be right about the necessitation claim, at least if it is understood on the evaluative sense of "rational," but it is held that nothing follows from this vis-à-vis the constitutive theory of introspective belief. The problem is that Shoemaker moves too quickly from the necessitation claim to the constitution claim. To see this, consider the passage quoted earlier from Shoemaker. There Shoemaker moves from a supervenience claim in the first sentence to the constitution claim in the second. Indeed, he says that he means the same thing by both claims. But, at best, he has only argued for the metaphysical

necessitation or supervenience claim, and the constitution claim does not immediately follow, as we have seen.<sup>10</sup>

It is doubtful that Shoemaker had the evaluative sense of “rational” in mind, however. In one of the only places where Shoemaker gives us any clues about the sense of rationality he has in mind, he writes: “The fact that the person is rational might be compared to the fact that the powder in the bomb was dry” (1996, 32). This certainly suggests that Shoemaker had the dispositional sense in mind. And, as we have seen, it is plausible that *if* the necessitation claim were true on the dispositional sense of “rational,” then the constitution claim would follow. It is widely agreed, however, that Shoemaker’s arguments for the necessitation claim do not establish the necessitation claim. Those who think that the necessitation claim is true, think that it is true on the evaluative sense of “rational,” but for independent reasons.<sup>11</sup> I suspect that no argument could establish the necessitation claim on the dispositional sense of “rational,” so I am sceptical about the prospects of the argument from necessitation. Nonetheless, as I will now argue, there is a better argument for constitutivism about introspective belief available, one which is implicit in Shoemaker’s work but which gets overshadowed by the argument from necessitation.

### 3 The Arguments from Cognitive Dynamics

In this section, I will formulate and evaluate the argument from cognitive dynamics for constitutivism about instrumental desire and the argument from cognitive dynamics for constitutivism about introspective belief. I will argue that the former is a good argument for constitutivism about instrumental desire, and, in light of the conclusions of the previous section, the *best* argument for this thesis. Then I will argue that the latter is a good argument, if not an equally good argument, for constitutivism about introspective belief.

#### 3.1 *For Constitutivism About Instrumental Desire*

The argument from cognitive dynamics for constitutivism about instrumental desire begins from an observation about the cognitive dynamics of instrumen-

<sup>10</sup> The supervenience claim is the conclusion of Shoemaker’s famous argument from self-blindness. See Shoemaker (1996, 47–48).

<sup>11</sup> Kieran Setiya argues that a nearby necessitation claim is true, and that this is what Shoemaker has correctly drawn attention to (Setiya 2011). But what gets necessitated, according to Setiya, is a capacity for introspective belief, not introspective belief itself.

tal desires. The observation is that they systematically come and go with the desires and means-end beliefs which explain them.

D<sub>1</sub>. (i) If someone comes to desire to  $\Psi$  and to believe that  $\Phi$ -ing is a means to  $\Psi$ -ing, they will come to desire to  $\Phi$  if they are fully rational and (ii) if someone merely instrumentally desires to  $\Phi$  and they cease desiring to  $\Psi$  or cease believing that  $\Phi$ -ing is a means to  $\Psi$ -ing, then they will cease desiring to  $\Phi$  if they are fully rational.

These are claims about the cognitive dynamics of particular desires. They are the cognitive dynamics of someone who is rational, that is, someone who is disposed to conform to the principles of rationality. These cognitive dynamics are partly constitutive of what it is to be disposed to conform to the principles of rationality. It is because we observe these cognitive dynamics that we believe that we are rational in this sense.

How could a cognitive system exhibit these dynamics? To see how, let's assume a broadly functionalist picture of beliefs and desires. On this picture, for  $X$  to believe that  $P$  is for  $X$  to be in some state or other which plays the believing-that- $P$ -role, and for  $X$  to desire to  $\Phi$  is for  $X$  to be in some state or other which plays the desiring-to- $\Phi$ -role. We could then re-describe the cognitive dynamics above in terms of ceasing to be in a state which plays the desiring-to- $\Phi$ -role under certain conditions, and coming to be in a state which plays the desiring-to- $\Phi$ -role under certain conditions. But now two importantly different hypotheses arise concerning the relations between these states.

According to one hypothesis, call it the *causal hypothesis*, the dynamics are explained by the fact that when someone comes to be in a state which plays the desiring-to- $\Phi$ -role, and someone comes to be in a state which plays the believing-that- $\Psi$ -ing-is-a-means-to- $\Phi$ -ing-role, their coming to be in these states jointly *causes* them—by means of their manifesting a rational disposition—to come to be in a state which plays the desiring-to- $\Psi$ -role. Similarly, someone's ceasing to be in a state which plays the believing-that- $\Psi$ -ing-is-a-means-to- $\Phi$ -ing-role or ceasing to be in a state which plays the desiring-to- $\Phi$ -role will cause them to cease being in a state which plays the desiring-to- $\Psi$ -role. Rational dispositions can then be thought of as ordinary causal dispositions, where the triggering conditions are thought of as causes of the manifestations. On this hypothesis the cognitive dynamics are explained by various causal transactions between constitutively distinct states or events involving consti-

tutively distinct states (they must be constitutively distinct in order to stand in causal relations).

According to another hypothesis, call it the *constitutive hypothesis*, the dynamics are explained by the fact that when someone comes to be in a state which plays the desiring-to- $\Phi$ -role and someone comes to be in a state which plays the believing-that- $\Psi$ -ing-is-a-means-to- $\Phi$ -ing-role, their coming to be in these states constitutes—by means of their manifesting a rational disposition—their coming to be in a state which plays the desiring-to- $\Psi$ -role. It does so because the former states together *constitute* a state which plays the desiring-to- $\Phi$ -role. Similarly, someone's ceasing to be in a state which plays the believing-that- $\Psi$ -ing-is-a-means-to- $\Phi$ -ing-role or ceasing to be in a state which plays the desiring-to- $\Phi$ -role just is their ceasing to be in a state which plays the desiring-to- $\Psi$ -role. Rational dispositions, on this hypothesis, are not ordinary causal dispositions. They are what we might call constitutive dispositions, since the triggering conditions bear a constitutive relation to the manifestations. On this hypothesis the cognitive dynamics are explained in terms of states which play particular roles jointly constituting states which play other roles. The relevant thing about the constitutive hypothesis is that, if it is true, then instrumental desires are not distinct from the desires and means-end beliefs which explain them. This is because instrumental desires are partly constituted by states which partly constitute the corresponding desires and means-end beliefs which explain them. So if there is an argument for the constitutive hypothesis, there is an argument for constitutivism about instrumental desire.

Each of these hypotheses is clearly an empirical hypothesis. If cognitive science were so advanced that we could determine which states play which roles, then, in principle we could settle the question of whether the states are constitutively distinct and causally related or constitutively non-distinct and constitutively rather than causally related. But we are far from being able to answer the question this way. The best we have, and the best we may ever have, is indirect evidence for one hypothesis over the other based on arguments to the best explanation of the observed cognitive dynamics. Let's consider the evidence for and against, then.

Perhaps the weakest consideration in favour of the constitutive hypothesis, but a consideration nonetheless, comes from the relative cognitive efficiency of having the states which play the role of certain desires being constituted by the states which play the roles of other desires and means-end beliefs, rather than having the former be distinct from and caused by the latter. A cognitive

system requires far fewer distinct states and fewer dependencies between them in order to have a wide range of instrumental desires on the constitutive hypothesis. To put the point in slogan form: *the constitutive hypothesis is cognitively more efficient than the causal hypothesis.*

A stronger consideration in favour of the constitutive hypothesis comes from the observation that the causal roles which are definitive of desires and means-end beliefs, along with the principles of rationality, predict that by merely having those desires and means-end beliefs, and being rational, the agent will be disposed to act *as if* she desired the means. If desiring the means were a matter of coming to be in a distinct state which plays the role of desiring the means, as the causal hypothesis holds, then the disposition to act as if one desired the means would be over-determined. There's nothing by way of the agent's dispositions to act that being in this state would contribute which is not already contributed by their being in these other states and their being rational. The state is motivationally redundant.<sup>12</sup> So, to put the point in slogan form: *the constitutive hypothesis avoids the prediction that instrumental desires are motivationally redundant.*

Perhaps the strongest consideration in favour of the constitutive hypothesis, the one that I am willing to put the most weight on, begins with an observation about the strength of the dependence of instrumental desires on means-end beliefs and other desires. As we saw earlier, one claim about the cognitive dynamics of instrumental desires is that if someone merely instrumentally desires to  $\Phi$  and they cease desiring to  $\Psi$  or cease believing that  $\Phi$ -ing is a means to  $\Psi$ -ing, then they will cease desiring to  $\Phi$ . While this claim is compatible with both the constitutive and causal hypotheses, the constitutive hypothesis has a far better explanation of it. Indeed, the explanation comes for free on the constitutive hypothesis, since it is no surprise that when one ceases to be in either of the states which jointly constitutes the state which plays the desiring-to- $\Phi$ -role that one will cease desiring to  $\Phi$ . The causal hypothesis requires the auxiliary hypothesis here that when the state which plays the desiring-to- $\Phi$ -role is caused and explained by the states which play the other roles, it will remain causally dependent on those states. Of course, it is possible that, simply as a matter of fact, such a state will remain causally dependent on these other states. But this claim has to be added as an auxiliary hypothesis to the causal hypothesis, thus making the hypothesis more complicated than the constitutive hypothesis. And, moreover, there is every reason to think

---

12 See Arpaly and Schroeder (2014, 9) for a similar observation.

that this causal dependence would sometimes break down, giving rise to stray instrumental desires, desires which are no longer dependent on the means-end beliefs and desires which caused them in the first place. But if this phenomenon exists in our psychology, it remains unobserved. The constitutive hypothesis correctly predicts that there will be no stray instrumental desires. To sum up the point of this paragraph in slogan form: *instrumental desires are deeply dependent on other mental states for their existence and the constitutive hypothesis best explains this.*

It is basically this consideration which motivates the argument in the following passage from Smith:

It is a striking fact that instrumental desires disappear immediately an agent loses either the relevant non-instrumental desire or means-end belief [...]. Yet there is no reason why this should be so if an instrumental desire were merely a desire that has a non-instrumental desire and a means-end belief somewhere in its causal history. Why should a desire disappear when (say) the desire that caused it, way back when, disappears? Instrumental desires are thus better thought of as being nothing over and above the non-instrumental desires and means-end beliefs that explain them. (2004, 96)

Smith begins here with the observation that instrumental desires are deeply dependent on the desires and means-end beliefs which explain them. He then argues against the causal hypothesis and for the constitutive hypothesis on the basis of the fact that the latter provides a better explanation of the observation than the former. Of course, Smith's argument against the causal hypothesis is too fast. An instrumental desire could remain causally dependent on another desire and a means-end belief in the way that a light's being on remains causally dependent on the light switch's being turned on. It could be that the relation between an instrumental desire and the desire and means-end belief which explains it is like this. But the constitutive hypothesis nonetheless provides a better explanation of the deep dependence between an instrumental desire and the desire and means-end belief which explains it.

The best argument, then, for constitutivism about instrumental desire is the argument from cognitive dynamics. Unlike the argument from necessitation for constitutivism about instrumental desire, which would decisively establish constitutivism about instrumental desire if it were cogent, the argument from cognitive dynamics makes constitutivism about instrumental

desire the conclusion of an ordinary argument to the best explanation. But it isn't really surprising that this should be so, since constitutivism about instrumental desire is most plausibly thought of as a contingent hypothesis about instrumental desires.

### 3.2 *For Constitutivism About Introspective Belief*

If the argument from cognitive dynamics for constitutivism about instrumental desire provides good support for constitutivism about instrumental desire, then perhaps an analogous argument from cognitive dynamics could provide good support for constitutivism about introspective belief. In this section I argue that it does.

The argument from cognitive dynamics for constitutivism about introspective belief begins from an observation about the cognitive dynamics of introspective beliefs. The observation is that introspective beliefs systematically come and go with the mental states—or in the case under consideration, the attitudes—that they are about.

D<sub>2</sub>. (i) If someone has some interest in the question of whether they  $\Psi$  that P, and they come to  $\Psi$  that P<sup>13</sup>, then they will come to believe that they  $\Psi$  that P if they are fully rational, and (ii) if someone introspectively believes that they  $\Psi$  that P, and they cease  $\Psi$ -ing that P, then they will cease believing that they  $\Psi$  that P if they are fully rational.

These are claims about the cognitive dynamics of particular beliefs. They are the cognitive dynamics of someone who is rational, that is, someone who is disposed to conform to the principles of rationality. These cognitive dynamics are partly constitutive of what it is to be disposed to conform to the principles of rationality. It is because we observe these cognitive dynamics that we believe that we are rational in this sense.

Now, let me be upfront here about an important disanalogy with the observation about the cognitive dynamics of instrumental desires. The observation just given is likely to strike many as highly controversial. Failures of introspective belief are the norm. Failures of instrumental desire are the exception. I have two responses to this kind of pessimism about introspective belief. The

<sup>13</sup> The antecedent of this conditional may need strengthening. Perhaps one needs to consider the question of whether one  $\Psi$ -s that P also.



first is that it vastly overstates the case. To say that failures of introspective belief are the norm is to overlook the wide range of cases where introspective belief is utterly unproblematic. Over a vast range of mundane beliefs, desires, and other mental states, I have utterly unproblematic introspective access. If failures of introspective belief were the norm here, our mental life would be in serious trouble. The second response is that nothing in the argument from cognitive dynamics depends on an overly optimistic view of our capacity for introspective belief. It may well be that failures of local rationality are far more common in the case of introspective beliefs, but as long as introspective beliefs have some of the features I draw attention to below, the argument from cognitive dynamics will go through. This is a point I will return to after presenting the rest of the argument.

How could a cognitive system exhibit the dynamics above? Again, assuming a broadly functionalist picture of beliefs and desires, we can recast the observation in the following terms. When someone comes to be in a state which plays the  $\Psi$ -ing-that-P-role, they come to be in a state which plays the believing-that-one- $\Psi$ -s-that-P-/-to- $\Phi$ -role, insofar as they are rational and have some interest in the question of whether they  $\Psi$  that P. When someone ceases to be in a state which plays the  $\Psi$ -ing-that-P-role, they cease to be in a state which plays the believing-that-one- $\Psi$ -s-that-P-role. Again, two importantly different hypotheses arise concerning the identity of these states.

According to one hypothesis, call it the *causal hypothesis*, the dynamics are explained by the fact that when someone comes to be in a state which plays the  $\Psi$ -ing-that-P-role, they are caused to come to be in a state which plays the believing-that-one- $\Psi$ -s-that-P-role, insofar as they are rational and have some interest in the question of whether they  $\Psi$ -that-P. And someone ceases to be in a state which plays the  $\Psi$ -ing-that-P-role, they are caused to cease to be in a state which plays the believing-that-one- $\Psi$ -s-that-P-role. On this hypothesis the cognitive dynamics are explained by various causal transactions between constitutively distinct states or events involving these states.

According to another hypothesis, call it the *constitutive hypothesis*, the dynamics are explained by the fact that when someone comes to be in a state which plays the  $\Psi$ -ing-that-P-role, they thereby come to be in a state which plays the believing-that-one- $\Psi$ -s-that-P-role, since the former state plays the latter role. And when someone ceases to be in a state which plays the  $\Psi$ -ing-that-P-role, they thereby cease being in a state which plays the believing-that-one- $\Psi$ -s-that-P-role, since it was the former state which played

the latter role. On this hypothesis the cognitive dynamics are explained by the states which play the first-order roles playing the second-order roles.<sup>14</sup>

Each of these hypotheses is clearly an empirical hypothesis. What arguments can be given for and against? Not surprisingly, the considerations are perfectly analogous to those given in the argument from cognitive dynamics for constitutivism about instrumental desire.

Perhaps the weakest consideration in favour of the constitutive hypothesis, but a consideration nonetheless, comes from the relative cognitive efficiency of having the states which play the roles of particular mental states also play the role of beliefs about those mental states, rather than having the former be distinct from and caused by the latter. A cognitive system requires far fewer distinct states and fewer dependencies between them in order to have a wide range of introspective beliefs on the constitutive hypothesis. This consideration has considerable bite in contemporary contexts where doubt has arisen, both on the basis of philosophical and empirical enquiry, concerning the claim that we have a distinct perception-like capacity for inner-sense, one which causally detects our mental states and outputs introspective beliefs. This view goes hand in hand with the causal hypothesis. The constitutive hypothesis cuts out the middle-man, and requires no distinct perception-like capacity for inner-sense.<sup>15</sup> To sum up the points of this paragraph in slogan form: *the constitutive hypothesis is cognitively more efficient than the causal hypothesis.*

A stronger consideration in favour of the constitutive hypothesis comes from the observation that the causal roles which are definitive of many mental states, along with the principles of rationality, predict that by merely being in those mental states, and being rational, an agent will be disposed to act *as if* she believes that she is in those mental states.<sup>16</sup> In particular, it has been observed that if you are in pain, say, and you are rational, you will be disposed, partly in virtue of the fact that you are in pain, to say “I am in pain.”

---

14 See Shoemaker (1996, 33–34, 242–244).

15 This might be a little bit unfair to the causal hypothesis. Since there is a version of the causal hypothesis which cuts out the middle-man too, and requires no distinct perception-like capacity for inner-sense. My point here is that once we are on the lookout for cognitively efficient hypotheses about introspective beliefs, the constitutive hypothesis wins hands down.

16 This point is well made by Shoemaker (1996). While Shoemaker makes the point in the context of arguing for the necessitation or supervenience claim, I am here making it in the context of the cognitive dynamics argument. Shoemaker sometimes says things which suggest that he might have something like the cognitive dynamics argument in mind. This is a point I will come to in the text.

This point has been made over and over by expressivists in the philosophy of introspection, who take it to show that statements like “I am in pain” do not report mental states but merely express them. But we needn’t understand the claim in this manner, since it is possible that such statements both report and express mental states, because the mental states they express double, according to the constitutive hypothesis, as beliefs about those mental states. If introspectively believing that you are in a particular mental state were a matter of coming to be in a distinct state which plays the role of a belief that you are in some mental state, as the causal hypothesis holds, then the disposition to act as if you believed that you were in that mental state would be over-determined. There’s nothing by way of the agent’s dispositions to act that being in this distinct state would contribute which is not already contributed by their being in the first-order state, their having an interest in the question of whether they are in that state, and their being fully rational. The state is motivationally redundant.<sup>17</sup> So, to put the point in slogan form: *the constitutive hypothesis avoids the prediction that introspective beliefs are motivationally redundant.*

Perhaps the strongest consideration in favour of the constitutive hypothesis, the one that I am willing to put the most weight on, begins with an observation about the strength of the dependence of introspective beliefs on the mental states they are about. As we saw earlier, one claim about the cognitive dynamics of introspective belief is that if someone introspectively believes that they  $\Psi$  that P, and they cease  $\Psi$ -ing that P, then they will cease believing that they  $\Psi$  that P.<sup>18</sup> While this claim is compatible with both the constitutive and causal hypotheses, the constitutive hypothesis has a far better explanation of it. Indeed, the explanation comes for free on the constitutive hypothesis, since it is no surprise that when one ceases to be in the state which plays the believing-that-one- $\Psi$ -s-that-P-role, one will cease believing that one  $\Psi$ s that P. The causal hypothesis requires the auxiliary hypothesis here that when the state which plays the believing-that-one- $\Psi$ -s-that-P-role is caused and explained by the state which plays the  $\Psi$ -that-P-role, it will remain causally

---

<sup>17</sup> To be clear: I am not denying that introspective beliefs themselves are motivationally redundant, that they make no difference to the cognitive functioning of the mental states they are about; I am only claiming that a distinct state which had these consequences would be redundant insofar as being in the mental states in question, having an interest in the question of whether you are in those mental states, and being fully rational would already have the consequences for one’s cognitive life we take introspective beliefs to have.

<sup>18</sup> This was pointed out to me by Daniel Nolan.

dependent on this state. Of course, it is possible that, simply as a matter of fact, such a state will remain causally dependent on this other state. But this claim has to be added as an auxiliary hypothesis to the causal hypothesis, thus making the hypothesis more complicated than the constitutive hypothesis. And, moreover, there is every reason to think that this causal dependence would sometimes break down, giving rise to stray introspective beliefs, *introspective* beliefs which are no longer dependent on the mental states which caused them in the first place. But if this phenomenon exists in our psychology, it remains unobserved. The constitutive hypothesis correctly predicts that there will be no stray introspective beliefs. And this is a significant point in its favour. So, to sum up the points of this paragraph in slogan form: *introspective beliefs are deeply dependent on other mental states for their existence and the constitutive hypothesis best explains this.*

While Shoemaker is more closely associated with the argument from necessitation, it is clear that he also has something like the argument from cognitive dynamics in mind. Indeed, I think that, to the extent that he does have the latter in mind, this is the best argument he has for constitutivism about introspective belief. Consider the following passage from a recent paper of Shoemaker's defending constitutivism about introspective belief:

One might, indeed, wonder whether there is any need to postulate standing second-order beliefs that self-ascribe available first-order beliefs. It goes with having the available first-order belief that *p* that if the question whether one believes that *p* arises, one will judge that one does—one will assent to the proposition that one believes that *p*. But this seems to be the result of one's having the belief that *p*, not the result of one's having a second-order belief whose cognitive dynamics is independent of that of the belief that *p*, in the way that the cognitive dynamics of one's belief about another person's belief is independent of that of the other person's belief. It would seem inefficient for our psychology to involve the storage of standing second-order beliefs ascribing available first-order beliefs, if there is nothing for these second-order beliefs to do that is not done by the first-order beliefs themselves. (2012, 247)


These remarks combine elements of all three of the considerations I have given above. Shoemaker makes a claim about efficiency, there is also a claim

about redundancy, and he speaks of the predicted cognitive independence of introspective beliefs from the mental states they are about on the causal hypothesis. I have teased out these considerations and argued that together they add up to a reasonable case for the constitutive hypothesis.

The best argument, then, for constitutivism about introspective belief is the argument from cognitive dynamics. Unlike the argument from necessitation for constitutivism about introspective belief, which would decisively establish constitutivism about introspective belief if it were cogent, the argument from cognitive dynamics makes constitutivism about introspective belief the conclusion of an ordinary argument to the best explanation. But it isn't really surprising that this should be so, since constitutivism about introspective belief is most plausibly thought of as a contingent hypothesis about introspective beliefs.

To end, let me return to the obvious line of criticism which may be raised against the argument from cognitive dynamics for constitutivism about introspective belief. The criticism is basically that it depends on far too rosy a picture of introspective belief. But we are now in a position to see that it does not. The consideration about efficiency requires only that we have a significant number of introspective beliefs, so that considerations of efficiency come into play. It doesn't require that we approximate omniscience and infallibility. The considerations about redundancy, likewise, only require that we have a significant number of introspective beliefs, so that considerations about redundancy come into play. And, finally, considerations about deep dependence do not require that we are rarely in error about our own mental states. This is an important point. We may have many false beliefs about our own mental states. We may often be in error about our own mental states. But as long as those beliefs about our own mental states arrived at by introspection remain deeply dependent on the mental states they are about—that is, as long as there are no stray *introspective* beliefs—the point about dependence holds.\*

Ryan Cox

 0000-0002-1381-448X

The University of Sydney

---

\* Thanks to Michael Smith for discussion on a very early version of this paper. Thanks also to members of the ANU Philosophy of Mind Work in Progress Group for their helpful feedback on an earlier version of this paper. Finally, thanks to several reviewers for *Dialectica* for their helpful comments.

ryan.cox@sydney.edu.au

## References

- ARMSTRONG, David M. 1968. *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.
- ARPALY, Nomy, and Timothy SCHROEDER. 2014. In *Praise of Desire*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199348169.001.0001.
- BOYLE, Matthew. 2009. "Two Kinds of Self-Knowledge." *Philosophy and Phenomenological Research* 78(1): 133–164, doi:10.1111/j.1933-1592.2008.00235.x.
- BYRNE, Alex. 2011. "Transparency, Belief, Intention." *Proceedings of the Aristotelian Society, Supplementary Volume* 85: 201–221, doi:10.1111/j.1467-8349.2011.00203.x.
- CASSAM, Quassim. 2014. *Self-Knowledge for Humans*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199657575.001.0001.
- CHALMERS, David J. 2010. *The Character of Consciousness*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780195311105.001.0001.
- DAVIS, Wayne A. 1986. "The Two Senses of Desire." In *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting*, edited by Joel MARKS, pp. 63–82. Chicago, Illinois: Precedent Publishing, Inc.
- GERTLER, Brie. 2010. *Self-Knowledge*. London: Routledge.
- GORDON, Robert M. 1986. "Folk Psychology as Simulation." *Mind and Language* 1(2): 158–171, doi:10.1111/j.1468-0017.1986.tb00324.x.
- HUME, David. 1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. 3rd ed. Oxford: Oxford University Press. Edited by L.A. Selby-Bigge; revised and with notes by P.H. Nidditch.
- KIND, Amy. 2003. "Shoemaker, Self-Blindness and Moore's Paradox." *The Philosophical Quarterly* 53(210): 39–48, doi:10.1111/1467-9213.00294.
- MARKS, Joel. 1986. "Introduction: On the Need for a Theory of Desire." In *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting*, edited by Joel MARKS, pp. 1–15. Chicago, Illinois: Precedent Publishing, Inc.
- MCDANIEL, Kris, and Ben BRADLEY. 2008. "Desires." *Mind* 117(466): 267–302, doi:10.1093/mind/fzno44.
- MORAN, Richard. 2001. *Authority and Estrangement. An Essay on Self-Knowledge*. Princeton, New Jersey: Princeton University Press.
- SCHROEDER, Timothy. 2004. "Functions from Regulation." *The Monist* 87(1): 115–135, doi:10.5840/monist20048717.
- SETIYA, Kieran. 2011. "Knowledge of Intention." In *Essays on Anscombe's Intention*, edited by Anton FORD, Jennifer HORNSBY, and Frederick STOUTLAND, pp. 170–197. Cambridge, Massachusetts: Harvard University Press.
- SHOEMAKER, Sydney S. 1996. *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, doi:10.1017/CBO9780511624674.

- . 2012. “Self-Intimation and Second-Order Belief.” In *Introspection and Consciousness*, edited by Declan SMITHIES and Daniel STOLJAR, pp. 239–258. Oxford: Oxford University Press, doi:[10.1093/acprof:oso/9780199744794.001.0001](https://doi.org/10.1093/acprof:oso/9780199744794.001.0001).
- SMITH, Michael A. 2004. “Instrumental Desires, Instrumental Rationality.” *Proceedings of the Aristotelian Society, Supplementary Volume* 78: 93–109, doi:[10.1111/j.0309-7013.2004.00117.x](https://doi.org/10.1111/j.0309-7013.2004.00117.x).
- STOLJAR, Daniel. 2019. “Evans on Transparency: A Rationalist Account.” *Philosophical Studies* 176(8): 2067–2085, doi:[10.1007/s11098-018-1111-x](https://doi.org/10.1007/s11098-018-1111-x).