

# Armstrong's Just-so story about Consciousness\*

Daniel Stoljar, ANU

## 1. Introduction

D. M. Armstrong's *A Materialist Theory of the Mind* is a prime source of many ideas that are still widely discussed in contemporary philosophy of mind (see Armstrong 1968). Among these are:

- The causal or functionalist analysis of belief as a state apt to cause a certain sort of behaviour; and the correlative analysis of a purpose (i.e., an intention or desire) as an information sensitive mental cause;
- The analysis of perception in terms of belief: perception as the getting of belief about one's immediate surroundings;
- The analysis of other mental states in terms of perception, and so ultimately in terms of belief: sensation as a sort of perception of one's own body, introspection a sort of perception of one's own mind;
- The analysis of a conscious mental state as a state that is the target of a certain sort of introspection, or inner perception; and
- The distinctive two-premise argument for the identity of mental states with physical states of the central nervous system. The first premise of the argument, which follows from his causal analysis, is that mental states are states apt to produce a certain sort of behaviour. The second, empirical, premise is that c-fibers firing and other neurophysiological states are in fact states apt to produce that sort of behaviour. The conclusion is that mental states are physical states.

Perhaps less well known is Armstrong's claim that his book provides "an answer to the question asked again and again by so many biologists and psychologists 'What is the biological function of consciousness?'" (1968, 162) In particular, he suggests at several points

---

\**Acknowledgments:* I presented an earlier version of this paper at the 2018 conference on *A Materialist Theory of the Mind: 50 Years On* held at The University of Sydney. I thank all the members of the audience their for their help, particularly W.G.Lycan. Thanks also to Peter Anstey, David Braddon Mitchell, Justin D'Ambrosio and Frank Jackson for extremely helpful comments as I was writing the paper.

that his views about purpose and about introspection may be combined to provide a “just-so story about introspective consciousness,” as W.G. Lycan later described it (see Lycan 1996).

My topic here is this just-so story—Armstrong’s account of the function of consciousness, as I will usually say. After providing an interpretation of the account (sections 2–3), I will develop two critical points about it. The first (section 4) is that appreciating this aspect of Armstrong’s view blurs the difference between his own perceptual model of introspection, and the anti-perceptual models advanced by such critics as Sydney Shoemaker. The second (section 5) is that the account is implausible since it conflates two interpretations of its central idea, namely, that “a purpose is an information-sensitive mental cause.” In section 6 I will bring these two criticisms together by noting a connection between Armstrong’s position and an apparently different argument by Shoemaker; both fail, I will argue, for the same reason. Section 7 is a brief conclusion.<sup>1</sup>

## 2. Armstrong’s Account

Armstrong’s account is contained in two main chapters of *A Materialist Theory of the Mind* (hereafter MTM), chapter 15 and chapter 6. Chapter 15 summarizes his overall attitude to introspection, which he says is crucial for the argument in the proceeding chapters. The summary concludes with this passage:

If there are to be purposive trains of mental activity, then there must be equally some means by which we become apprised of our current mental state. Only so can we adjust mental behaviour to mental circumstances. For instance, if we are doing a calculation ‘in our head’ we will need to become aware of the current stage in the mental calculation that we have reached. Only if we do become so aware will we know what to do next. So there must be a way of becoming aware of our current mental state, which means that there must be introspection. (1968, 327)

Similar passages occur earlier, in chapter 6, which sets out his account of purposes:

---

<sup>1</sup> It goes without saying that the contemporary literature relevant to the evolutionary function of consciousness is much bigger and more sophisticated now than when Armstrong was writing. Some examples are: Barron and Klein 2016, Godfrey-Smith, 2017, 2019, LeDoux 2019, Polger 2017, Polger and Flanagan 2002, Robinson, Maley, and Piccinini 2015, Rosenthal 2008. Beyond a few comments, I will concentrate here on Armstrong’s own view, and the way it interacts with other parts of his philosophy of mind, rather than attempt to connect what he says with this literature.

If the train of mental states between stimulus and response is to be a purposive affair... there is some method whereby information about the current state of affairs can be fed back to the driving force behind the teleological sequence. But if the teleological sequence in question is a sequence of events in the mind, this means that the agent must become informed of what is currently going on in his mind. (1968, 163)

Introspection is therefore a logical precondition of teleological mental behaviour. So we have given a ‘teleological deduction’ of the existence of introspection. (1968, 163)

It is reasonable to assume that Armstrong’s account of the function of consciousness, whatever it is, is contained in these passages. So what is that account exactly?

I think it is fair to summarize the account as an argument for the conclusion that consciousness has a biological function. The argument may be formulated this way:

- P1. Purposive mental activity has a biological function.
- P2. Consciousness is necessary for purposive mental activity.
- C. Ergo, Consciousness has a biological function.

Let me begin by making some comments about the constituent notions here and about the initial plausibility of both the premises and the inference.

### **3. Commentary**

*Comment 1.* By ‘a purpose’, as we have seen, Armstrong means an “information-sensitive mental cause” (1968, 139). To illustrate, suppose that my purpose is to get a drink, or, as we might say more readily these days, I desire to get a drink. This desire is a mental cause, for Armstrong, because it is a state apt to bring about, i.e. cause, certain behaviours: the desire for a drink will usually cause me to get a drink. The desire is information sensitive, in that its existence and nature will be influenced by what other information I have, that is, by what other beliefs and desires I have. Suppose I learn, and so come to know, that I have a drink, for example, and for some reason hadn’t noticed this before. Then my desire to get a drink will, other things being equal, go away or be extinguished. Being extinguished by the knowledge that it is satisfied is one way for a desire to be information sensitive.

Of course not all desires are sensitive in this simple way. Suppose I desire to be warm; that is, suppose I desire to be in some persistent state—a “continuing condition” Armstrong calls it (1968, 140)—rather than for some event to occur. This desire will not go away if I learn that I am warm. Nevertheless the desire is still sensitive to information in a more general sense. If I desire to be warm, and learn that I am, I will normally be disposed to form new desires: the desire to remain warm, for example.<sup>2</sup>

*Comment 2.* By ‘purposive activity’, Armstrong means a series of actions initiated and sustained by a purpose, in the sense just introduced. As we will see, it is crucial for him that such activity comes in two varieties: purposive mental activity and purposive physical activity.

Purposive physical activity is a series of physical actions initiated and sustained by a purpose. As an example of purposive physical activity, Armstrong gives a case in which, as he says (1968, 143), “I set my face toward the public house”—that is, I decide to go out to the pub, the Nag’s Head Hotel, for example. In such a case, my desire causes a series of physical actions—leaving my front door, walking down Glebe Point Road, turning right on St John’s Road, and so on.

Purposive mental activity is a series of mental actions initiated and sustained by a purpose. As an example of mental purposive activity, he gives a case of mental arithmetic, in which “we are doing a calculation ‘in our head’” (1968, 327). In such a case, my desire, for example, to arrive at the product of two other numbers, requires me to focus first on certain operations, then remember the outcome of one operation, while performing another, and so on. A different case Armstrong gives is one in which we go through various possible actions in imagination, in order to work out their consequences and so their desirability (see, e.g. 1968, 158ff). I may know there are two routes to the Nag’s Head, for example, and consider in imagination first taking one, then the other. To the extent that the first route is imagined to lead to a negative outcome—e.g. running into a certain moustachioed logician outside the Habit Wine Bar—I may decide on the second.

---

<sup>2</sup> Armstrong emphasizes the information sensitivity of purposes or desires, but presumably this is a feature of mental states in general. Take the belief that I have a drink in my hand. If I learn that I do not have a drink in my hand, this belief will normally go away. In that sense, beliefs are information sensitive too, since they are usually affected by whatever else I know.

*Comment 3.* When he speaks of the ‘biological function’ of consciousness and other mental states, I take Armstrong to be committed to a kind of *adaptationism* about the mind, i.e., he is assuming that (e.g.) consciousness is a trait whose current presence in a population has a quite specific historical explanation in terms of natural selection. Admittedly Armstrong is not terribly explicit on this point, and MTM predates the contemporary discussion of adaptationism.<sup>3</sup> But this is nevertheless the most plausible interpretation of what he says. After all, whether consciousness is an adaptation is “the question asked again and again by so many biologists and psychologists.” And it is adaptations, and not functions of other sorts, about which people tell just-so stories.<sup>4</sup>

*Comment 4.* Putting comments 1–3 together, P1 makes an empirical claim about the biological past, since it claims that purposive activity played a distinctive role in the evolutionary history of conscious creatures.

What is Armstrong’s evidence for such a claim? The only relevant passage I know of in MTM is this:

Let us begin by considering what is the biological function of mind. Mental processes biologically speaking are those that fall between stimulus and response. Their biological function is to make the response more sophisticated, and so more efficient. Now, if what goes on in the mind has a directed or purposive character, this may well aid the agent to make the best response...For instance, in a problem situation, various possible responses may be tried out ‘in the imagination’ in order to see which response will best fulfill the agent’s purposes. As a causal result the response may be more efficient. (1968, 163)

Unfortunately, this passage raises more questions than it answers. Make the response more sophisticated than what? Sophisticated in what way? Why does increased sophistication generate greater efficiency? Did greater efficiency *in fact* play any role in the biological past of human beings? And greater efficiency than what exactly? Without an answer to these

---

<sup>3</sup> The classic paper here is Gould and Lewontin 1979; see also Orzack and Forber 2017.

<sup>4</sup> This is certainly true of the work referred to in footnote 1, for example.

questions, it is hard to avoid the conclusion that the passage, and MTM more generally, provides no support at all for P1.

If MTM offers no support for P1, one might be tempted to dismiss the argument set out above. One might also be tempted to suppose that the argument can't possibly capture the account presented in MTM. But I think both reactions are mistaken. First, while I have presented Armstrong's account as an argument, I don't mean that he offers an explicit defence of its first premise. On the contrary, I am sure he is not trying to provide an account of the function of mental activity or of the mind in general; rather he is providing an account of the function of consciousness *given* the premise that purposive mental activity has a function. Second, while Armstrong does not defend that premise, one might think it fair enough to assume that it is true, especially if we go along with the adaptationism I am assuming Armstrong is working with. After all, adaptationism presumably entails that if we have a thing of sufficient complexity, we may provisionally suppose it to have some biological function, even if we don't know what that function is. Moreover, as we will see later, by purposive mental activity, Armstrong means something similar to what other writers refer to as rationality or reasoning, and it is not at all implausible that rationality or reasoning has a biological function.<sup>5</sup> I propose therefore not to reject Armstrong's argument simply because he himself provides no argument for P1.<sup>6</sup>

*Comment 4.* By "consciousness" Armstrong means, as he says, "the awareness of our own state of mind" (1968, 85). He goes on: "the technical term for such awareness of our own mental state is 'introspection' or 'introspective awareness.'" It is in this sense that consciousness and introspection come together for Armstrong, and this is why he is happy to move back and forth from one to the other. For him, one is in a conscious mental state if and only if one is introspectively aware of one's being in that state; to put it differently, one is in a conscious mental state if and only if one is conscious in a certain sort of way of one's being in that state.

---

<sup>5</sup> A recent prominent recent discussion of this issue is Mercier and Sperber 2017.

<sup>6</sup> In this sense, Armstrong's account is an instance of what Robinson, Maley and Piccinini (2015) have recently called the 'strong reading' of an adaptationist account. They characterize an account of this sort as follows: "On a strong reading...consciousness is necessary for certain mental processes. That is, the relevant mental processes cannot take place without consciousness. If so, then it is impossible to have nonconscious functional duplicates of conscious beings. Instead, there must be functions performed by conscious beings that a nonconscious being could not perform. In order for such claims to stand, a nonconscious organism must be unable to perform the same function; otherwise consciousness would not be needed to obtain the adaptive advantage" (2015, 368).

The notion of consciousness Armstrong adopts here is what Norman Malcolm in this later debate with Armstrong (see Armstrong and Malcolm 1984) called the ‘transitive sense’, in which one is conscious *of* something. But Armstrong also uses ‘conscious’ in an intransitive sense to apply to the state in question. In this sense, a state is (intransitively) conscious if and only if the subject of the state is aware in some way (i.e., conscious in the transitive sense in some way) of being in that state. Since these notions are inter-defined, for most purposes it makes no difference which one we have in mind.

*Comment 5.* Armstrong focuses on a notion of consciousness tied to introspection—introspective consciousness, as Lycan says—but he does not deny that there are other notions of consciousness; indeed his later work on this topic prefigures the pluralism about consciousness that is dominant in the literature today (see Armstrong 1981; see Block 1997 and for more recent discussion, see Pautz and Stoljar 2019). In that later work, for example, Armstrong distinguishes perceptual consciousness, in which you are conscious of what you perceive, i.e., physical objects and their perceptually available properties, and introspective consciousness, in which you are conscious of (e.g.) your perceiving something.

Nevertheless, introspective consciousness is in his view the central or basic case, the sort of consciousness that explains the rest. This is supposed to follow from reflection on the case of the long-distance truck driver, who figures prominently in Armstrong’s later discussions of consciousness (see Armstrong 1981, 1984), but who makes his first oblique appearance in MTM:

This is something that can happen when one is driving very long distances in monotonous conditions. One can ‘come to’ at some point and realize that one has driven many miles without consciousness of the driving, or perhaps of anything else. One has kept the car on the road, changed gears even, or used the brake but all in a state of ‘automatism’. (1968, 93)

As I understand it, Armstrong’s reasoning concerning this example is as follows:

consciousness in the central sense is what is missing before the truck-driver comes to; what is missing before the truck-driver comes to is introspective consciousness; hence introspective consciousness is consciousness in the central sense.

I think—unoriginally; see Lycan and Ryder 2003—that this reasoning is unpersuasive. While introspective or second-order consciousness *is* missing in the truck driver case so too is

consciousness in a first-order sense. Armstrong is right that, prior to coming to, the truck driver fails to be introspectively aware of his own mental states. But the driver also fails to attend to the road or the gears or the breaks, even though presumably he was aware of such things. If you think, as many do, that attention, or attention to a sufficient degree, is one of the factors that turns mere awareness into consciousness in at least one good sense, the truck driver example does not show the centrality of Armstrong's introspective notion, as opposed to an alternative first-order attention-based one.

*Comment 6.* While Armstrong overreaches on the centrality of introspective consciousness, I think we should ignore this too in our investigation of the argument presented above. For suppose introspective consciousness as Armstrong defines it is merely one of several different varieties of consciousness, rather than the main or basic one; it is still reasonable to focus on the function of consciousness of this variety.

The reason is that this version of consciousness raises in a very straightforward way the very question that Armstrong says is asked again and again, namely the question of the function of consciousness. What prompts that question is the apparent possibility that various cognitive functions can go on in the absence of consciousness: on the face of it, one can believe, remember, want, perceive and act, and yet one does not *consciously* believe, remember, want, perceive or act. This concern is just as urgent in the case of introspective consciousness as in the case of varieties.<sup>7</sup> In sum, what Armstrong is trying to establish is the interesting conclusion that consciousness in his sense does not come apart from some other cognitive functions; it does not detract from the interestingness of that conclusion that there are analogous issues about consciousness in other senses.

*Comment 7.* 'Necessity' as it occurs in P2 is intended to capture Armstrong's repeated use of modals in the passages above. Take his remark, "if there are to be purposive trains of mental activity, then there must be equally some means by which we become apprised of our current mental state" (1968, 327, emphasis added). I take 'must' here to have wide-scope over the conditional; hence what is being said is that, necessarily, if there is purposive mental activity, there is a means by which we become apprised of our current mental state, i.e., there is

---

<sup>7</sup> Contemporary work is often focused, not on introspective consciousness but on phenomenal consciousness (as Block 1997 calls it). Nevertheless, questions about the evolution of consciousness are still prompted by the idea that consciousness in this sense may be removed or altered even though the function may remain.



consciousness. Further, I take ‘necessarily’ here to be logical or metaphysical necessity, i.e., necessity in the widest sense—it is hard to interpret Armstrong’s use of words like ‘deduction’ in any other way. Hence what is being said is that, in all possible worlds, if there is purposive mental activity, so too there is consciousness.<sup>8</sup>

*Comment 8.* Putting comments 4–7 together, P2 has nothing to do with the contingent causal history of the world, and *a fortiori* nothing to do with the evolutionary past of conscious creatures. Rather it is a modal claim: necessarily, if you undergo purposive mental activity, the states in question must be conscious in the introspective sense.

Why does Armstrong believe P2? The basic idea, as I understand it, concerns the distinction between purposive mental activity and purposive physical activity introduced above. In the case of purposive physical activity, we need to keep track of what is going on in the world, and adjust our mental states accordingly. In the case of purposive mental activity, by contrast, we need to keep track of what is going on in the mind. But keeping track of what is going on in the mind is a matter of being introspectively aware of what is going on in the mind, and this in turn requires consciousness. Later we will consider whether the contrast between purposive physical activity and purposive mental activity is really as Armstrong says it is.

*Comment 9.* We have seen the reasons for supposing that both premises of this argument are true; do they entail the conclusion?

The answer is no. Suppose the heart has a biological function, to pump blood say; and suppose that, necessarily being located in space is necessary for pumping blood. It does not follow that being located in space has a biological function: being located is a precondition of something that has a function, but does not have the relevant function itself. More generally, A can have a biological function, and B can be necessary the existence of A, and yet B might fail to have any function. Hence Armstrong’s premises do not entail their conclusion.

In view of the fact that the argument is invalid one might yet again be inclined to reject it. But again I am not going to make this criticism in what follows. For one thing, it may be that, while the premises do not entail the conclusion, they nevertheless support it; for example, perhaps the best explanation of the two premises is that consciousness has a

---

<sup>8</sup> I will consider later whether a weaker notion of necessity would suit Armstrong’s purposes.

function. Moreover, it may be that the premises can be enriched in such a way that the conclusion does follow; for example, it might be that what is true is not simply that being conscious is necessitated by purposive mental activity but that it explains why purposive mental activity has the function it has.

#### 4. Perceptual and Non-perceptual Theories of Introspection

So far I've set out Armstrong's just-so story as an argument for the conclusion that consciousness has a biological function. And I've noted that, while one might question both the first premise of that argument and its passage to the conclusion, this will not be my approach. For the rest of the paper, therefore, I will focus on Armstrong's second premise—i.e. P2—the premise that consciousness is necessary for purposive mental activity.

In a moment I will consider whether P2 is persuasive on its own terms. But first, I want to point out that Armstrong's commitment to it creates difficulties for what has become the standard way of drawing the distinction between perceptual and non-perceptual theories of introspection.

We have seen that the notion of consciousness for Armstrong is introspective consciousness: a state is conscious in this sense if and only if the subject of the state is aware of being in it in the introspective way. But what is the introspective way? Armstrong himself offers a distinctive answer to this question. For him, the subject needs to be aware of the state in a *perceptual* way. It is for this reason that Armstrong holds an *inner sense* view of introspection.

What is it for a person to be aware of something in a perceptual way? For Armstrong, to perceive something in general is to be disposed to form a certain belief about it. This is quite different from other theories of perception that tend to see perception as a state related to but distinct from belief. On relational views, for example, perception is a relation between a person and a physical object. And on contemporary representational views, even when they trace their origins to MTM, perception is often thought of as a state whose functional role is distinct from that of belief.<sup>9</sup>

In light of the fact the relevant notion of perception is so belief-like for Armstrong, one might suspect that the distinction between his own perceptual view of introspection and a non-perceptual view will not be worth drawing. But that is not true to Armstrong's rhetoric,

---

<sup>9</sup> See Siegel 2016 and the references therein for some discussion on these distinctions between theories of perception.

which continually emphasizes the notion of inner sense. Nor is it true to the literature surrounding Armstrong's book (and indeed other literature on introspection), which often classifies his view as an inner sense view, and draws a distinction between it and other views.

How then is the distinction to be drawn? The main suggestion here is due to Sydney Shoemaker (1996), who draws the distinction in terms of the following thesis, which he calls *the impossibility of self-blindness*.

- (1) Necessarily, if you are in a mental state  $M$ , and various background conditions obtain, and you are rational, you will believe you are in  $M$ .

To illustrate the idea behind (1), take the mental state of being in extreme pain. If (1) is true of extreme pain, it is a necessary truth—using again the widest sense of necessity—that if you are in that state, and you are rational and various background conditions obtain, you will believe that you are in that state.

What does Shoemaker mean by 'rationality' and what background conditions does he have in mind? By 'rationality' I take him to be talking about a property of an agent whose combinations of mental states at a time, and whose sequences of mental states over time, conform to the requirements of rationality, where these requirements are understood in a relatively well-known way: other things being equal, do not have overtly contradictory beliefs; other things being equal, believe the obvious consequences of your beliefs; other things being equal, desire to  $F$  if you believe that  $F$  is how to do  $G$ , and you desire to  $G$ ; and so on. As regards background conditions, Shoemaker I think has in mind the following: (a) being psychologically capable of forming the belief that you are in pain, that is, understanding the proposition that you are in pain, and being able to believe it; (b) being such that it matters to you that you are in pain—sometimes you might be in pain and be psychologically capable of forming the belief, but it just doesn't matter to you that you are, for example, because you are too busy doing or thinking about other things; (c) having no counter-evidence (provided, for example, by testimony from a friend or a doctor). Putting all this together, it is plausible (whether or not it is ultimately correct) that (1) is true of extreme pain: necessarily, given these conditions, if you are in extreme pain you will believe that you are, so long as you are rational.

While (1) is plausible in the case of extreme pain, it is not plausible in case of logically analogous conditions. Consider, for example, extreme debt. Even if counterparts of Shoemaker's background conditions hold, and you are rational, it does not follow that you

will believe you are in extreme debt if you are. You might be on your way to meet your accountant, for example, and have not yet heard the bad news. If so, (1) is not true of extreme debt. (Remember that (1) asserts that something is necessarily true, and ‘necessity’ here is to be understood, again, in the widest sense.)

Extreme debt is not a psychological condition, but there are psychological states that behave more like it than like extreme pain. Consider the state of knowing when and how to contract ‘want’ and ‘to’ to ‘wanna’ in English. According to linguists, speakers of English know when and how to do this. They know it is appropriate, for example, in the sentence ‘Who does Mary want to go to the game with?’ but they will resist it in the (superficially similar but structurally different) sentence ‘Who does Mary want to go to the game with her?’ (The sentence ‘Who does Mary wanna go to the game with?’ is a normal part of speech, but ‘Who does Mary wanna go to the game with her?’ seems ungrammatical.)<sup>10</sup> Presumably there is a rule that captures when and how speakers do this, a rule they both know and abide by. But it is not plausible that speakers of English know that they know this rule, nor is it plausible that they are able to attend to the rule or to formulate it in words. Likewise, it is not plausible that they can teach it to each other, since that would require knowing, attending to, and formulating the rule. (Third parties such as linguists may know the rule, of course, or at least may hope one day to figure it out; a classic paper is Lasnik and Maito 1984; see also Boeckx 2000) But suppose now you are an English speaker and thereby abide by the rule in question—call it *R*, whatever it is. It may be that you know *R*, that variations of Shoemaker’s background conditions may obtain, and that you are rational, and yet you may nevertheless not believe that you know *R*—you may lack certain crucial pieces of evidence that you know to be relevant, for example. If so, (1) is not true of knowing rule *R*.

How does the thesis of the impossibility of self-blindness—that is, (1)—serve to distinguish anti-perceptual theories from perceptual theories of introspection? Shoemaker’s suggestion, I think, is that anti-perceptual theories entail that (1) is true with respect to a range of mental states, whereas perceptual theories entail it is false for all. In other words, perceptual theories want to say that there is a sense in which every psychological state behaves like extreme debt or knowing the rule of ‘wanna’-contraction in English: we can

---

<sup>10</sup> The main structural difference between these sentences might be brought out as follows. ‘Who does Mary want to go to the game with?’ has the form, ‘Who is the *x* such that Mary wants to go to the game with *x*?’ while, ‘Who does Mary want to go to the game with her?’ has the form, ‘Who is the *x* such that Mary wants *x* to go to the game with her’. In the first case, Mary is the implicit subject of the infinitive ‘to go to the game’; in the second, she isn’t.

come to know about them alright, but from an epistemological point of view the way in which we do so is not different from the way in which we know many other things.

Shoemaker motivates this suggestion in part by contrasting (1) with various stronger theses that Armstrong himself criticizes in chapter 6 of *MTM*. One such thesis is:

(2) Necessarily, if you are in a mental state *M*, you will believe you are in *M*.

If (2) is true of extreme pain, then, if you are in extreme pain you will believe that you are. Armstrong argues against this by pointing out that the realizer of extreme pain is modally distinct from the realizer of the belief that you are in extreme pain. Like many others I find this argument unpersuasive. Suppose  $S_1$  is physical state that is the realizer of extreme pain, and  $S_2$  is a physical state that is the realizer of believing that you are in extreme pain. It may be that  $S_1$  and  $S_2$  are modally distinct in the sense that it is possible that you are in  $S_1$  and not in  $S_2$ . But it does not follow (even if it is true) that it is possible to be in extreme pain and not believe that you are. After all, it may be that, in the situation in which you are in  $S_1$  and not in  $S_2$ ,  $S_1$  is no longer the realizer of pain! Conclusion: that the realizers of various states are modally distinct does not entail that the realized states are modally distinct.<sup>11</sup>

Even if Armstrong's specific argument against (2) is unpersuasive, he is nevertheless right to reject it, even for the case of extreme pain. If (2) is true, any agent who is in extreme pain will believe that they are. But this is implausible: the cognitive requirements of being in pain and of believing that one is in pain are quite different, which strongly suggests that the first may be met without the second being met. If so, (2) is not true even in the case of extreme pain. But of course, even if (2) is false, it might be that a weaker thesis, such as (1), is true in that case. That is in fact Shoemaker's suggestion.

In summary, it is plausible that Shoemaker has drawn a clear distinction between perceptual theories and non-perceptual theories of introspection. But now the problem that I want to identify for Armstrong's account of the biological function of consciousness emerges. For look again at Armstrong's second premise, P2. That premise says that if you engage in purposive mental activity, you will need to be conscious of the constituent states of that activity, at least in certain cases. The problem is that, on the face of it, Armstrong's purposive mental activity includes Shoemaker's rationality. When Armstrong talks, for example, of

---

<sup>11</sup> The classic presentation of this point is the last few paragraphs of Lewis 1972; see also Shoemaker 1996. For some more recent discussion, see Stoljar 2018.

adjusting “mental behaviour to mental circumstances”, it is hard to understand what he says, unless he is talking about rational chains of mental states, rational combinations of mental states unfolding over time. But this is precisely what Shoemaker has in mind when he talks of ‘rationality’. However, if it is the case that such rational adjustment of mental behavior to mental circumstances requires consciousness, and if consciousness is just the belief that you are in the relevant state, then it would seem that according to Armstrong (1) is true of the purposes involved in mental activity: they are such that, if you are rational, and if various background conditions apply, you must believe that you have them if you do. But if (1) is true of purposes, then Armstrong stands revealed as holding a non-perceptual theory of introspection. Despite what he says, Armstrong is not an inner sense theorist, at least if we operate with Shoemaker’s test for being an inner sense theorist.

There are various ways that one might respond to this result. First, one might suggest that the distinction between perceptual and non-perceptual models is without foundation and that properly understood there is no opposition here: Armstrong wants to reject (2) or something like it, Shoemaker wants to reject (1) or something like it. However, while there is certainly some truth to this, it is hard to see it as an attractive response for Armstrong who as I have noted continually emphasizes the perceptual character of introspection.

Second, one might suggest that what is really important to Armstrong in this part of his work is his well-known hypothesis of the self-scanner, that is, the idea that, as a matter of fact, there is a causal system instantiated in the brain that reliably permits it to record its own states. However, while Armstrong certainly is committed to this, it is not *all* he is committed to. For by itself the existence of a self-scanner is consistent with a denial of a perceptual theory—at least if we operate with Shoemaker’s account of the perceptual/non-perceptual distinction.<sup>12</sup> But as we have seen, Armstrong repeatedly emphasizes that his own account is a perceptual one.

Third, one might draw the distinction between perceptual and non-perceptual theories of introspection in a way different from Shoemaker. For example, one might argue that perceptual theories portray introspection as like perception, not on a representational theory of perception, but instead on a relational theory. However, even if this can be done, it is again unclear that it will help Armstrong. The reason this time is that it is probable that Armstrong

---

<sup>12</sup> Indeed, as I read him Shoemaker makes this point himself in this discussion of Armstrong’s broad perceptual model, particular when he separates what he calls the ‘causal condition’—roughly the idea of a self-scanner—from the ‘independence condition’—roughly the idea that self-blindness is possible; see Shoemaker 1996, 224.

will turn up on the wrong side of the divide. So once again we seem to confront the uncomfortable truth that Armstrong is not the perceptual theorist he claims to be.

Finally, one might try to respond to the point by downplaying it, that is, by suggesting that purposive mental activity is an unusual phenomenon. I suspect this would be Armstrong's own response to the issue, but the problem with it is that purposive mental activity is not unusual. In fact, any sort of purposive activity at all will include purposive mental activity of some sort or another. However, this point will emerge more naturally when we turn to the second criticism of Armstrong's position that I want to develop.

### 5. Two Notions of Information Sensitivity

We have seen that Armstrong's argument about the function of consciousness has a surprising side effect: its second premise entails that he does not hold an inner sense theory of introspection, at least on Shoemaker's understanding of what it is to hold such a theory. But of course, this point is merely *ad hominem*; it does not by itself say much about whether the premise is true. So that is question I want to consider now.

As we saw earlier, the consideration Armstrong offers for P2 is that purposes are information-sensitive mental causes, and that information sensitivity requires consciousness.

However, the problem with this line of thought is that there are two ways to interpret the notion of information-sensitivity, a first-order way and a second order way. To illustrate, consider again the desire to get a drink. On the first-order interpretation, this state is information sensitive just in case if you learn you have a drink, the desire to get a drink will go away, other things being equal—we may put by saying the state has first-order information sensitivity. On the second order interpretation, by contrast, the state is information sensitive just in case if you learn that you have a drink, and *if you are introspectively aware of learning this*, the desire will go away, other things being equal. Or consider a non-purposive state that is presumably also information sensitive, such as the belief that p. This state is first-order information sensitive (as we may say) just in case, if you learn that p is false, the belief that p will go away, other things being equal. By contrast, the state is second-order information sensitive just in case, if you learn that p is false, *and if you are introspectively aware of learning this*, the belief that p will go away, other things being equal.

Once we have this distinction before us, however, the problem for P2 is straightforward. If we understand information sensitivity in the first-order way, P2 is false: it is not true that introspective consciousness is necessary for purposive mental activity, for one can engage in such activity without being aware of the relevant states. If we understand it in

the second-order way, however, then P2 is unjustified, since none of Armstrong's examples motivate the idea that mental states are information sensitive in this second-order way.

In the case of mental arithmetic, for example, it is true that one's next step depends in an information sensitive way on the step before, but there does not need to be introspective consciousness at any point; on the contrary, in such a case one seems firmly focused on mathematical rather than psychological issues. The same thing is true in the case in which I imagine two routes to the Nag's Head. I might consider taking the first route, which leads past the Habit Wine Bar, and resolve after deliberation to take the longer route via Ross St. No introspection need occur here at all. It is true that there needs to be information sensitivity in the sense that the mental states will fall into a rational sequence, but nothing beyond that is required.

Indeed, the point that nothing in these examples demonstrates the need for second-order information sensitivity can be strengthened if we bring to the surface what seems to be a false presupposition in Armstrong's entire discussion of the distinction between mental and physical purposive activity. As we have noticed, Armstrong does not think that consciousness is necessary for physical purposive activity; it is only necessary for mental purposive activity. But this betrays a misunderstanding of this distinction. There are trains of purposive activity concerning many different things: mathematics, pubs, logicians, moustaches, and so on. Some issue in physical activity, some don't. But almost all of these involve mental activity of some sort, and none of them require the mental activity or the states that constitute this activity to be conscious. If not, there is no support here for P2.

At certain points, Armstrong himself comes close to seeing the basic problem with his account. In response to his own discussion, for example, he points out that "[o]n occasion we can solve quite complex problems during sleep, or while our mind seems to be otherwise occupied. If this is possible, where is the need for awareness of our own mental state in purposive mental activity?" (1968, 164). That is a good question, but Armstrong's response is implausible. What he says is that in such cases we are aware, but not aware that we are aware. In other words, his claim is that in such cases we have a lack of third-order awareness but not a lack of second-order awareness. But while it is indeed plausible that in deep sleep we do not have third-order awareness, this does little to undermine the very plausible suggestion that we lack second-order awareness too.<sup>13</sup>

---

<sup>13</sup> Might it be insisted on Armstrong's behalf that, since second-order awareness is part of the functional role of purposive mental activity, if you don't have awareness of that sort you don't have the activity in question? I



I have argued that P2 as Armstrong formulates it, and therefore his argument in general, is unpersuasive. Is it possible to adjust the premise so that it becomes more plausible?

One possibility is to focus on the strength of the necessity claim in P2. Above I interpreted it widely, as involving all possible worlds. But one might try to interpret it more weakly. On this interpretation, P2 says not that consciousness is metaphysically necessary for purposive mental activity but only that it is empirically necessary for it.

However, the problem with this proposal is that, if Armstrong's account is developed in this way, the speculative ('just-so') element in his position becomes extreme. Earlier I said that Armstrong's account was premised on the empirical assumption the mental activity had a function, noting that this was a reasonable assumption in the circumstances. But if his account is premised on the speculation that consciousness is required for mental activity to perform its function whatever it is, we lose the sense in which he has an answer to the question of the evolution of consciousness at all.

A different possibility is to adjust the notion of 'purposive mental activity' so that what is intended is purposive *conscious* activity: a series of conscious states initiated and sustained by a purpose. However, the problem with this proposal is that while P2 is now plausible, P1 begs the question. From this point of view, Armstrong's P1 amounts to the claim that conscious mental activity has a function, and that is precisely the thing the argument aims to establish.

## 6. The McGinn-Shoemaker Argument

In a certain sense the problem for Armstrong's account we have just been looking at was predictable from the beginning. As I noted above, the reason that people are interested in the evolution of consciousness is because of the apparent possibility that various mental states, events and process can occur without being conscious. What Armstrong is trying to do is rule out that possibility. His idea is that there is a particular sort of mental process that cannot go on without consciousness as a matter of necessity, namely, purposive mental activity. But the problem is that this does not seem to be true; we seem to be able to engage in purposive mental activity *whether or not* we are conscious in Armstrong's sense. For all Armstrong has said, therefore, the question asked again and again remains unanswered.

---

think the problem with this suggestion is two-fold. First, it is hard to see it as attractive to Armstrong, who in general insists on the point that introspective awareness can come apart from first-order states. Second, it is not plausible on its own terms, since it remains the case that there is no justification for building this into the functional role of purposive mental activity in the way that he does. I am indebted here to David Braddon Mitchell.

At this point, various avenues are open to us. One is to leave Armstrong aside and focus for its own sake on what the function of consciousness is. It goes without saying that I am not going to attempt that here. Instead, let me close the discussion by bringing together the two critical points I have made about Armstrong's account: that it undermines his claim to be an inner sense theorist; and that it conflates two notions of information sensitivity. The connection between these points come into sharper relief if we compare Armstrong's discussion with a more contemporary issue in philosophy of introspection.

The issue I have in mind returns us to Shoemaker's thesis of the impossibility of self-blindness, namely (1): necessarily, if you are in a mental state *M*, and various background conditions obtain, and you are rational, you will believe you are in *M*. As I noted Shoemaker argues for this thesis in various ways—partly on the basis of examples, and partly by comparing it with other stronger principles that are implausible. But he also presents an argument for the thesis that is striking in its similarity to Armstrong's discussion.

The main idea of this argument, which is based on a suggestion due to Colin McGinn, is that in order to abide by principles of rationality it must be that one knows what belief states one is in (see McGinn 1982, 20, and Shoemaker 1996, 2009). Suppose that for purposes of illustration we focus on what might be called a 'no-contradiction rule'—viz., it is rationally required that if one believes *p*, one will not also believe not-*p*, other things being equal. Then we may present the argument as follows.

- (3) If you are rational you abide by rules such as the no-contradiction rule.
- (4) If you abide by rules such as the no-contradiction rule, you must know (and so believe) what belief states you are in.

From these premises it is reasonable to infer that (1) is true of any state that falls within the purview of the no-contradiction rule, at least if the premises themselves are necessary.

But the weakness in this argument is that (4) is unconvincing. If one abides by the no-contradiction rule, it is true that what belief states one is in need to be changed over time, revised, regulated and so on. However, it seems perfectly possible that such revision goes on, as Richard Moran puts it in an important critique of this line of thought, at a sub-personal level: "the purposes of belief-regulation do not require that the person get involved here at all" (Moran 2001, p. 110).

Now, in response to Moran, Shoemaker has suggested that, while this might be true in some cases, it is not true in all:

But I think that in an important class of cases the revision and updating does require that that there be second-order beliefs about what the contents of the belief system are. These are cases in which the revision of the belief system requires an investigation on the part of the subject, one that involves conducting experiments, collecting data relevant to certain issues, or initiating reasoning aimed at answering certain questions. Such an investigation will be an intentional activity on the part of the subject, and one motivated in part by beliefs about the current contents of the belief system. These will include the belief that there are certain apparent inconsistencies or incoherencies in the system, the belief that there are gaps in how the system represents the world, and the realization that the system represents the existence of certain states of affairs for which it provides no explanation. (2009, p. 244).

However, a version of Moran's objection applies here too. It is true that if one embarks on the sort of investigation Shoemaker describes it is irrational to believe the relevant proposition. But that implies only that a rational agent who undergoes this sort of investigation must somehow or other fail to believe that proposition—either they never did believe it at all, or else they will suspend belief in it for the course of the investigation. But this failure to believe is something that on the face of it could be brought about in the way Moran suggests, i.e. sub-personally. If so, there is no reason to accept Shoemaker's claim that this sort of investigation is motivated by beliefs about the current content of the belief system.

No doubt there is more to say about the dispute between Shoemaker and Moran on that point. I will not discuss it further here, except to note that my sympathies are with Moran. My main point is to notice that Shoemaker's argument here is strikingly similar to Armstrong's and yet is used in the service of a non-perceptual theory rather than a perceptual one; further evidence, then, that the lines between these two are starting to blur. Likewise, the problem with Shoemaker's argument is extremely similar to the problem we isolated for Armstrong's. Moran himself says that the 'person need not get into the picture'. That's one way to put it. Another is that abiding by the no-contradiction rule requires first-order but not second-order informational sensitivity. If so, both Shoemaker's and Armstrong's position are mistaken and for the same reason.

## **7. Conclusion**

Our overall discussion may now be summarized as follows. (i) Armstrong's just-so story about the function of consciousness is a revealing and under-noticed part of MTM. (ii) Its second premise, P2, undermines the distinction between perceptual and non-perceptual theories of introspection. (iii) P2 is false since it neglects the possibility of a first-order approach to information sensitivity. Finally, (iv) further support for the previous two points can be found by noticing that Armstrong's discussion is strikingly similar to Shoemaker's discussion.

## References

- Armstrong, D.M. 1968. *A Materialist Theory of the Mind*, London: Routledge.
- Armstrong, D.M. 1973. *Belief, Truth and Knowledge*, London: Routledge.
- Armstrong, D.M. (1980) *The Nature of Mind and Other Essays*, Ithaca, NY: Cornell University Press.
- Armstrong, David and Malcolm, Norman 1984: *Consciousness and Causality: A Debate on the Nature of Mind*, Oxford: Blackwell.
- Barron Andrew B. and Klein Colin 2016 "What insects can tell us about the origins of consciousness" *PNAS* May 3, 2016 113 (18) 4900–4908; first published April 18, 2016
- Block, Ned. 1997. "On a Confusion About a Function of Consciousness." In *The Nature of Consciousness*, edited by N. Block, O. Flanagan, and G. Guzeldere. Cambridge, MA: MIT Press.
- Boeckx, Cedric 2000 'A Note on Contraction', *Linguistic Inquiry* 31: 357–366
- Godfrey-Smith, P 2017 *Other Minds: The Octopus, the Sea and the Deep Origins of Consciousness*. Farrar, Straus and Giroux, 2016, and Harper-Collins UK, 2017.
- Godfrey-Smith, P 2019 'Evolving Across the Explanatory Gap' *Philosophy, Theory, and Practice in Biology* (2019) 11:1.
- Gould, S. J., and R. C. Lewontin, 1979. "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme," *Proceedings of the Royal Society London, Series B*, 205: 581–598.
- Lasnik, H. and M. Saito 1984 'On the nature of proper government', *Linguistic Inquiry* 15, pp. 235–289
- LeDoux, Joseph 2019 *The Deep History of Ourselves* Penguin
- Lewis, D. 1972. 'Psychophysical and Theoretical Identifications' *Australasian Journal of Philosophy*, 50: 249–258, repr. in his *Papers in Metaphysics and Epistemology*

- (Cambridge: Cambridge University Press, 1999), 248–261. All references are to the reprinted version.
- Lycan, William G. 1996: *Consciousness and Experience* Cambridge MA: MIT Press.
- Lycan, William G and Ryder, Zena 2003 ‘The Loneliness of the Long-distance Truck Driver’ *Analysis* Vol. 63, No. 2 (Apr., 2003), pp. 132–136 .
- McGinn, Colin. 1982. *The Character of Mind* Oxford: Oxford University Press
- Mercier H and Sperber D 2017 *The Enigma of Reason: A New Theory of Human Understanding* Cambridge, MA: Harvard UP.
- Moran, Richard. 2001 *Authority and Estrangement: An Essay on Self-Knowledge* Cambridge, MA: Harvard University Press.
- Orzack, S.H. and Forber, P, 2017 ‘Adaptationism’, *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2017/entries/adaptationism/>
- Pautz, A. and Stoljar, D. (2019) *Blockheads! Essays on Ned Block’s Philosophy of Mind and Consciousness* Cambridge MA: MIT Press.
- Polger, Thomas 2017 ‘Rethinking the Evolution of Consciousness’, in Susan Schneider and Max Velmans (eds) *The Blackwell Companion to Consciousness*, 2nd edn., Oxford: Blackwell, 77–92.
- Polger, Thomas W., and Owen J. Flanagan. (2002) ‘Consciousness, Adaptation and Epiphenomenalism’. In James H. Fetzer (ed.) *Consciousness Evolving* (Amsterdam: John Benjamins), 21–42.
- Robinson, Zach, Maley, Corey.J and Piccinni, Gualtiero (2015) ‘Is Consciousness a Spandrel?’, *Journal of the American Philosophical Association*, 1: 365–383.
- Rosenthal, David. (2008) ‘Consciousness and its function’ *Neuropsychologia* 46: 829–40
- Siegel, Susanna, "The Contents of Perception", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/win2016/entries/perception-contents/>](https://plato.stanford.edu/archives/win2016/entries/perception-contents/).
- Shoemaker, Sydney (2009) ‘Self Intimation and Second Order Belief’, *Erkenntnis* 71: 35–51. Reprinted in Smithies, D. and Stoljar, D. (eds) 2012 *Consciousness and Introspection* New York: Oxford University Press pp 239–257. References are to the reprinted version.
- Shoemaker, Sydney. 1996. *The First Person Perspective and Other Essays*. Cambridge: Cambridge University Press.
- Stoljar, D. 2018 ‘Introspection and Necessity’, *Nous*, 52: 389–410

