

Can Intentionalists be Reductionists?

Adam Pautz

University of Texas at Austin

Australian National University

<https://webspaces.utexas.edu/arp424/www/index.html?uniq=tpfy1b>

I am indebted to William Lycan for his very helpful comments. I have profited from them very much. I am pleased to find that he agrees with my argument against input-based theories of sensory representation. Roughly, input-based theories say that the content of a sensory state is determined by what it causally covaries with in the external world or in one's body. By contrast, output-based theories appeal to how the state is used by an individual to guide behavior. We all agree that simple input-based theories fail: for instance, the theory that a sensory state S represents F iff S is caused by F under normal conditions, where 'normal' is understood statistically. The failure of such account is shown by Ned Block's Inverted Earth example, among other things. My examples were designed to refute more sophisticated input-based theories, for instance Michael Tye's Tracking Theory and Fred Dretske's Indicator Theory, which appeal to a robust notion of optimal conditions or to biological function. For, though the individuals in my examples are internally and behaviorally different owing to the fact that different selection pressures operated on their ancestors, they *evolved naturally* to track the same external features. There are no inverting lenses involved, for example. So even a sophisticated input-based theory delivers the verdict of Same Content and hence (given Intentionalism) Same Experience in these cases. But, in view the fact that the individuals also undergo considerably different internal processing and have different behavioral dispositions, the more plausible verdict is Different Experiences and Different Content. The cases also constitute a counterexample to other input-based theories, for instance Stalnaker's brand of the Indicator Theory (1998) and Fodor's Asymmetric Dependence Theory (1990).

The problem is not merely one of detail. By their very nature, input-based theories are unable to deliver the correct verdict of Different Contents and Different Experiences in these cases, for the individuals involved are stipulated to be exactly the same as regards input-factors.

Lycan and I agree on something else. He writes:

I think intentionality is a much greater difficulty for physicalism than is anything to do with consciousness, qualia, phenomenal character, subjectivity, etc. If intentionality itself can be naturalized, those other things are pretty easily explicated in terms of it. (foot-note 3)

The issue of naturalizing intentionality is no longer in vogue. For instance, Fodor (1998, 12), who once toiled away on his own particular input-based theory, now only expresses a vague hope that some such theory will be adequate, and then moves on to other issues. But many of us still want to know whether consciousness is an extra ingredient of the world, or whether it is all atoms and void. And many of us also endorse Intentionalist theories of consciousness. But, given Intentionalism, the problem of consciousness just reduces to the problem of intentionality (plus or minus a bit). So if we wish to know whether consciousness is primitive (as I think) or reducible (as Lycan thinks, with fingers crossed), we have to revisit this well-worn issue.

I am trying to revisit the issue from a different angle. I use empirical results to argue that in the cases of Maxwell and Twin Maxwell, Yuck and Yum and Mild and Severe the right verdict is Different Experiences and hence (given Intentionalism) Different Contents. As our most developed, input-based theories do not accommodate this verdict, they must be mistaken. That leaves output-based theories. But I think that these theories are non-starters when it comes to sensory representation. They are not capable of generating the fine-grained contents of experiences. So I think that consciousness cannot be naturalized, not because of Zombies or Mary, but just because sensory intentionality cannot be. Once we finally give up on the project of naturalization or reduction, we may solve the problems facing reductive views (see §4).

My plan is as follows. As I said, my argument about Maxwell and Twin Maxwell is designed to refute input-based theories. In response to my argument, Byrne and Tye argue that the Tracking Theory, at least, survives unscathed. Lycan suggests an interpretation of Byrne and Tye's response which is different from my own. Although he accepts my argument and rejects Byrne and Tye's response, in §1 I will consider his different interpretation. Lycan expresses pessimism about all existing

reductive theories of representation. Nevertheless, he thinks that it is reasonable to suppose that Millikan's output-oriented Consumer Theory delivers the right verdict in my cases. In §2 I will argue that this theory cannot be applied to sensory representation. In §3 I will explain why I think David Lewis's Narrow Interpretation Theory (which is a very different output-oriented theory) fails as well. In §4 I will argue for a kind of Primitivist Intentionalism which gives up the on the problem-ridden naturalization project.

1. Input-Based Theories: What is Byrne and Tye's Reply?

In "Tracking Intentionalism and Optimal Conditions", I worked with a certain interpretation of Byrne and Tye's response. Lycan suggests a different interpretation. In fact, from his comment (as well as personal communication), I can glean two interpretations besides my own. I have found Lycan's comments here helpful. I will discuss these interpretations in turn. It will help if we have the argument before us. Where **B** is the set of non-representational, non-experiential facts in one of my cases, the argument is as follows:

- [1] If Tracking Intentionalism is true, then **B** → Same Optimal Cause. (The arrow means metaphysical necessitation.)
- [2] If Tracking Intentionalism is true, then Same Optimal Cause → Same Content and hence Same Experience. (Def. of 'Tracking Intentionalism')
- [3] If Tracking Intentionalism is true, then **B** → Same Experience. (1, 2)
- [4] As against Tracking Intentionalism, **B** → Different Experiences.
- [5] Tracking Intentionalism is false. (3, 4)

Lycan suggests that there is a better way to state the argument. I agree, and I will offer a different formulation of the argument at the end of the present section.

First Interpretation

Byrne and Tye appear to argue from Different Experiences to Different Contents, then from Different Contents to One Wrong (at least one of the subjects is misperceiving), and then from One Wrong to Not-

Optimal. In terms of the above argument, it appears that they would reject [1].

I was not exactly sure what to make of this response. One reason is that it appears circular to explain representation in terms of optimal conditions and then to explain optimal conditions in terms of veridicality (which is a representational notion). Another reason is that my aim was to argue that the Tracking Intentionalism delivers the verdict of Same Experience in my examples, but that the right verdict is Different Experience. In their response, Tye and Byrne begin by *asserting* Different Experiences and One Wrong. But this is exactly what I argued that the Tracking Intentionalist *cannot* assert. Consider an analogy. Suppose a defender of the Justified True Belief of Knowledge began his criticism of Gettier's counterexamples by saying "Well, since the subject's in these cases don't have knowledge . . .". Of course, if Gettier was right, the Justified True Belief Theorist is not entitled to use this claim in an objection to the cases.

Tye and Byrne's assertion (without argument) of Different Experience and One Wrong would be justified if they were stipulated features of the case. Then they would be common ground. So there is a kind of inference to the best explanation for this interpretation. The following passage provides further support for it:

So, as Pautz would describe the case, there is no obvious reason to suppose it is metaphysically possible. Allegedly, Twin Maxwell is a product of natural selection, someone operating under the same laws as Maxwell with a similar kind of visual system, whose experiences represent the same range of colors as Maxwell's, and who not only has no abnormalities whatsoever in his visual system *but also is subject to significant color illusions*. Pautz simply stipulates that *all these conditions* can be met together [my italics]. A defender of Tye's theory may reasonably deny it. Each condition is indeed metaphysically possible, but they are not all possible together. (Page 22 of the version on Alex Byrne's website).

On this interpretation, Byrne and Tye took Different Experiences and One Wrong as stipulated, and argued from them to Non-Optimality. This is the interpretation I have worked with. On this interpretation,

there are two problems with the response. First, these conditions are not stipulated, so the response rests on a misinterpretation of the case. Rather, the stipulated conditions of the case are the non-experiential, non-representational conditions (i)-(v) listed in “Tracking Intentionalism and Optimal Conditions”. And these conditions are evidently compossible. Surely, two species could naturally evolve to respond to the same distal properties with different neural processing, and different behavior, as a result of different selection pressures operating upon their ancestors. What could prevent it? Second, the response is circular. Given his theory, to provide a reason to think that optimal conditions do not obtain, the Tracking Intentionalist must find a *non-representational* fact X in **B**, and argue “X, therefore suboptimal” (for this point, see also Chalmers entry “Representationalism Showdown” on his weblog *Fragments of Consciousness*).

Second Interpretation

On a second interpretation, which is suggested by Lycan, Byrne and Tye were not *asserting* Different Experiences. Rather, they were *assuming for the sake of argument* that the right verdict is Different Experiences, that is, that I am right. Then they were using this to argue that “Pautz is not entitled to his own initial stipulation of optimality” (Lycan, 2).

Byrne and Tye did not discuss the presentation of the argument above. But if their response is successful, it would have to identify a mistake with this argument. The basic suggestion here is that the verdict of Different Experiences is incompatible with my optimality premise. Now [4] expresses my claim that the correct verdict is Different Experiences and [1] expresses my optimality premise. So, the idea would have to be that [4] and [1] are incompatible. But this is evidently not the case. (Note that [1] is conditional. Note also that the relevant notion of “optimal conditions” is defined in non-representational, non-experiential terms. Even if one of the subjects in my cases continually misperceives, that is not *logically* inconsistent with the obtaining of optimal conditions, in the relevant sense.) So, on this interpretation as well, the response fails.

Third Interpretation

An implicit premise of the argument is

[0] Possibly, **B**.

Without this premise, the argument is not valid. A third interpretation is that Byrne and Tye were offering an argument against [0]. On this interpretation, they were not arguing that the premises of the argument are inconsistent. Rather, they were arguing that the case itself is inconsistent, so that [0] is false. In his footnote 2, Lycan suggests on behalf of Byrne and Tye an argument against [0]. According this suggestion, Byrne and Tye were attempting “to turn that argument on its head”, by holding on to both Tracking Intentionalism and Different Experiences and using these to argue that the case is not possible. Here is a simpler argument which uses the same premises:

- [a] Tracking Intentionalism and **B** entail Same Experience (in effect granting my [3])
- [b] **B** entails Different Experiences (granting my [4])
- [c] Tracking Intentionalism is (necessarily) true (the Tracking Intentionalist sticking to his guns)
- [d] Therefore, it is not possible that **B** is true

In effect, the idea is follows. If he grants the rest of the premises, the Tracking Intentionalist faces a choice: reject Premise [0] or reject Tracking Intentionalism. The idea is that the Tracking Intentionalist can stubbornly hold on to his view, and use it as leverage to reject [0], that is, to reject the possibility of **B**.

Consider an analogy. Let **B** be the conditions of a Gettier case described in non-epistemic terms. The defender of Justified True Belief (JTB) analysis of knowledge might argue as follows. [a] JTB and **B** entail Knowledge. [b] But, intuitively, **B** entails No Knowledge (in effect granting that the case would constitute a counterexample). [c] But JTB is true! Therefore, [d] **B** is not possible. In short: if **B** were possible, JTB would be false, but JTB is true, so such a situation could not possibly arise.

But this kind of move is obviously a desperation move. Here I think it is much more reasonable to apply *modus tollens*. (The same point applies to Lycan’s version.) It would be absurd to stubbornly hold on to Tracking Intentionalism and use it to reject Premise [0], the claim that the case is possible. The conditions in **B** are just these:

- (i) Maxwell and Twin Maxwell belong to different species and view an object with reflectance property $F&H$. Owing to *naturally evolved* differences in their postreceptoral wiring Maxwell is put into “binary” opponent channel state P involving activity in two chromatic channels, while Twin Maxwell is put into “unitary” opponent channel state Q involving activity in only one channel
- (ii) in consequence of these hard-wired differences, their color-related behavioral dispositions differ considerably (they would differ in their performance on the psychophysical tests for color vision)
- (iii) but, because their receptor systems are the same, P and Q both track the same reflectance property, $F&H$
- (iv) further, since they naturally evolved their different wiring, both their visual systems are working in accordance with design
- (v) their behaviors, though different, are equally adaptive and appropriate, because they are adaptations to different selection pressures

These conditions are evidently compossible. Surely, two species could naturally evolve to respond to the same distal properties with different neural processing and different behavior, as a result of different selection pressures operating upon their ancestors. What could prevent it? That this case is possible is much more plausible than any philosophical theory. Such cases are possible; Tracking Intentionalism entails that in all such cases Maxwell and Twin Maxwell have the same experience; but this is just not plausible. So, on this interpretation too, the response fails. And Lycan agrees with me here (personal communication).

“Optimal conditions” as a red herring

I close my discussion of Byrne and Tye’s response with a general comment. I think that Tracking Intentionalism, as well as the argument, should be stutable without using the expression that is causing all the trouble: “optimal conditions”. To get a reductive theory of sensory representation, the Tracking Intentionalist must define optimal conditions in naturalistic terms. For instance, he might define optimal conditions as follows:

An individual X is in a state that tracks F *under optimal conditions* iff (i) X is an evolved creature and his sensory system is working as it evolved to work OR (ii) X is a non-evolved creature and X engages in behavior which is adapted to his environment.

This is rough, but let's work with it for the purposes of illustration. Now consider the following view:

Tracking Intentionalism*. X and Y have the same experience iff they track the same property and each is either an evolved creature and his sensory system is working as it evolved to work OR a non-evolved creature whose behavioral dispositions are adapted to his environment.

Tye does not offer a full dress definition of optimal conditions, so he does not strictly speaking defend Tracking Intentionalism*; but he does offer an informal gloss, and his informal gloss comes close to something like the above.

Now, the formulation of Tracking Intentionalism* does not contain the term of dispute, 'optimal conditions'. This term is "defined out" of the formulation. And we can formulate an argument against it which also does not use this term. Let *Same Tracking* be the claim that the right hand of the biconditional above holds of two individuals from different species. Let *Different Neurobiology and Behavior* be the claim that, owing to naturally evolved differences, they undergo different postreceptoral processing, and have different behavioral dispositions; yet their sensory systems are working just as they were designed to work, and, although their behavioral dispositions are different, they are equally adaptive, because they are adaptations to the different selection pressures operating on them. Now let **B** be the conjunction of the following conditions:

- [i] Different Neurobiology and Behavior
- [ii] Same Tracking

Evidently, [i] and [ii] are compossible. Now consider the following argument against Tracking Intentionalism*:

- [1] There are worlds at which Same Tracking and Different Neurobiology and Behavior are true (what I just said)
- [2] If Tracking Intentionalism* is true, then at every such world, Same Experience is true.
- [3] Against this, at such worlds (or at least some of them), Different Experiences is true. This is reasonable in view of what we know about the neural basis of experience, and in view of the fact that Different Neurobiology and Behavior is also true at these worlds.

Now, any charge that optimal conditions don't obtain at the relevant worlds is irrelevant. [1] is obviously true; [2] follows from the definition of 'Tracking Intentionalism*'; and [3] is much more plausible than Tracking Intentionalism*.

This argument shows that Tracking Intentionalism*, at least, is false. One could offer a different naturalistic definition of "optimal conditions", and so endorse a different version of Tracking Intentionalism, Tracking Intentionalism**. But it would be vulnerable to the same style of argument. If, on the other hand, the theorist leaves "optimal conditions" completely undefined, then he does not really have a theory but only a theory schema. One would have to make it plausible that there is some naturalistic definition of optimal conditions which would not result in a theory which is open to this style of counterexamples. But this is not plausible, for Maxwell and Twin Maxwell's situations are symmetrical in respect of all the naturalistic candidates.

Now I think that the only reply to the above argument which is open to the Tracking Intentionalist* (or to the defender of Dretske's Indicator Intentionalism) is to accept that there are worlds at Same Tracking and Different Neurobiology and Behavior are both true, but to reject [3], claiming that in all such worlds Same Experience is indeed true. In other words, he might stick to his guns and say it is not possible to have Same Tracking and Different Experiences, even where Same Tracking is accompanied by Different Neurobiology and Behavior. In response, I cannot show apriori that this reply is wrong, for [3] is not apriori. I can

only appeal to general reasonableness. I can only say how ludicrous I think this would be. Imagine meeting Maxwell and Twin Maxwell, or Yuck and Yum, or Mild and Severe. For instance, Yuck would withdraw from a foodstuff violently, throw-up, make a face; while Yum would be drawn to it, devouring as much as possible. And our best theories of taste processing would predict that they have different taste experiences. Imagine how implausible it would be to say that they have the same experience in the name of a philosophical theory. Given what we know about the biological basis of experience, the proposition that at least some of the Same Tracking/Different Neurobiology&Behavior worlds are also Different Experiences worlds, so that Same Tracking and Different Experiences are compossible (contrary Tracking Intentionalism*), is far more plausible than Tracking Intentionalism*.

So, in at least some of these worlds, Different Experiences is true. Given Intentionalism, that is to say that Maxwell and Twin Maxwell's color experiences represent the object as having different colors (presumably colors from the same color space), even though Same Tracking holds. Now, intuitively, the object cannot have both colors, so at least one gets it wrong. Call this Error. Thus, we get the result that Error is compatible with Same Tracking without malfunction. But if this is correct, then it is just something we have to live with. I have argued that Different Experiences is compatible with Same Tracking without malfunction; and if Different Experiences entails Error, then Error is compatible with Same Tracking without malfunction. There is no logical inconsistency between Error and Same Tracking without malfunction. And this result is not so implausible that we should apply modus tollens and reject the claim that Different Experiences holds in at least some of the Same Tracking/Different Neurobiology&Behavior worlds. For, as we have seen, that claim is very plausible, and more plausible than any philosophical claim to the effect that Error and Same Tracking without malfunction are incompatible.

I also note that we do not have to accept Error. For instance, on Shoemaker's non-standard brand of Intentionalism, both Maxwell and Twin Maxwell can get it right. And if you combine standard Intentionalism with a response-dependent view of color, you can avoid Error. And, in the case of pain and taste, we can accept Different Experiences without having to accept Error. For instance, owing to neural differences between them, the same bodily disturbance might

produce different pains in Mild and Severe, and the same chemical property might produce different taste experiences in Yuck and Yum, without either party getting it wrong. But if Different Experiences means that we must accept Error in the case of color vision, so be it.

I conclude that the cases of Maxwell and Twin Maxwell, Yuck and Yum and Mild and Severe pose a decisive problem for input-based theories.

2. Output-Based Theories: Papineau and Millikan

My focus in Pautz (2006) was on input-based theories. There I simply assert that Papineau's *Success Theory* and Millikan's *Consumer Theory*, which are output-based, also deliver the mistaken verdict of Same Content and Same Experience. But Lycan is right that I did not argue for this claim. The reason was that such theories were not my focus. In a companion paper (2005) I argue against output-based theories.

Lycan says that he is "pessimistic about all materialist psychosemantics" (3). This suggests that he would reject Millikan's theory in the end. But he notes that it was designed to avoid some of the problems with input-based theories. And he says, "So it is reasonable to suspect that it will also give the right judgment in the Maxwell case" (3).

Two preliminary points (i) Lycan *seems* to think that, Maxwell and Twin Maxwell aside, the Consumer Theory is superior to the Tracking Theory and other input-based theories. I am not sure about this. True, input-based theories face well-known problems, for instance the distance problem and the disjunction problem. But the Consumer Theory faces problems of its own. For instance, it faces the indeterminacy problem (Fodor 1990), problems about concepts (Peacocke 1992), and problems about adaptive misrepresentation (Peacocke 1993, 224-225). One might think that input-based theories and Millikan's theory simply trade problems. So, if we put Maxwell aside, I am not sure that there is a strong case for the Consumer Theory over input-based theories. (ii) Suppose that the Consumer Theory solves some of the problems with input-based theories. I am not sure why this means we should be confident that it solves the problem of Maxwell and Twin Maxwell as well, which seems to be distinct from the usual problems (the disjunction problem, etc.). To show this, one would, as Lycan says, engage in "an exercise in Millikan studies".

I propose, in a rough and preliminary way, to do that now. I will also consider Papineau's theory. I will argue that Papineau's Success Theory and Millikan's Consumer Theory are inapplicable to sensory representation. There are, I think, two main problems with reductive theories of sensory representation:

The Problem of Different Contents: They fail to deliver the correct verdict of Different Contents in cases like Maxwell and Twin Maxwell, Yuck and Yum and Mild and Severe.

The Underdetermination Problem: Or else they fail to generate assignments of fine-grained contents

I believe that input-based theories, such as those defended by Dretske and Tye, fall victim to the second problem. On this point, Lycan and I are agreed. But I believe that output-based theories fall victim to the second problem. If anything, Millikan's Consumer Theory is worse off than input-based theories when it comes to explaining sensory representation, in my opinion.

(In fact, I believe that both input-based and output-based theories fall victim to both problems. But I think that the most *obvious* problem for input-based theories is the Problem of Different Contents, and the most *obvious* problem for output-based theories is the Underdetermination Problem.)

I begin by providing a very rough and oversimplified exposition of the theories. Then I turn to evaluation.

The theories

First, Papineau's Success Theory (1993). The Success Theory as Papineau develops it is a theory of belief content; he does not talk about experiential content, as far as I know. On the Success Theory, the content of desire is primary and the content of belief is explained in terms of the content of desire. In particular:

B has content *C* iff *B*, in combination with the subjects desires, will produce successful action if (iff?) *C* obtains.

An action is *successful* iff it satisfies the subject's desires. The idea here goes back at least to Frank Ramsey. To take a simple example: if B and the desire for food yield to produce successful action (an action that satisfies the desire for food) if there is food in the fridge, then B has the content *there is food in the fridge*.

Since this theory explains belief content in terms of desire content, one might wonder how Papineau explains desire content. He says that a desire has content *C* iff *D* has the function of making it the case that *C* is true. For Papineau, one source of function is evolution; another is learning. This is the teleological element of his theory.

Although Papineau does not apply his Success Theory to sensory representation, it might be so applied. For instance, one might say:

E has content *C* iff *E*, in conjunction with the subject's desires, will produce successful action if (iff?) *C* obtains.

To take a simple example: if *E* and the desire for a red ball produce successful action if a red ball is in front of the subject, then *E* has the content there is a red ball in front of the subject.

Millikan's Consumer Theory (1989) is as follows:

B has content *C* iff *C* is Normal condition for the proper function of the consumer devices that use *B* in their historically Normal way.

This needs to be unpacked. Consumer devices are often devices that use B to produce behavior. (Although not always. For instance, in the frog example below, the digestive system counts as a consumer device.) Suppose that B often leads to behavior X in the relevant species. This will have an evolutionary explanation. What explains the proliferation of consumer devices which use B to produce behavior X? The explanation may essentially mention a salient external condition C. Sometime in the past, when B lead to behavior X and condition C was present, some good consequence for the organism followed. In that case, C is *now* the content of B. (This is so even if condition C never obtains anymore.)

For example, take the well-worn example of the frog and the fly. (In my opinion, this is a *very* strange example to use as a test case, for of course we have no pretheoretical intuition about the contents of frog

representations.) Under many different conditions, an internal state B of frogs cause them to dart out their tongues: when there is a black thingy moving in front of it, when there is frog food moving in front of it, and when there is a fly moving in front of it. What of these conditions does B represent? According to Millikan, B represents frog food, because the most salient external condition which explained why such a consumer device proliferated in frogs is the presence of frog *food*. The presence of this condition is what explains why tongue-darting had good consequences, and hence what explains the proliferation of the consumer device that produces this behavior.

The Consumer Theory may also be applied to sensory representation. In fact, Millikan herself seems to do this:

If the position of the chair does not correspond, so, to my visual representation of its position, that will hinder me equally in my attempts to avoid the chair when passing through the room, to move the chair, to sit in it, to remove the cat from it, to make judgments, etc. (1989, 290)

Applied to sensory representation, the Consumer Theory goes something like this:

E has content *C* iff *C* is Normal condition for the proper function of the consumer devices that use *E* in their historically Normal way.

The above passage suggests that Millikan's theory is very similar to Papineau's. But the theories remain importantly different. Papineau explains one difference in the following way:

This analysis is just a version of Millikan's idea that we should identify the content of an informational state with that condition which will ensure that the behaviour this state generates will have biologically successful results. But it goes beyond Millikan in focusing specifically on results which are successful in the specific sense of satisfying desires, not on any results which are biologically successful in any sense. (Papineau, 1998, 7)

He says that this means that his theory is better as regards Fodor's indeterminacy problem. Of course, there are many other differences.¹ And both theories face many problems of detail. One problem I see with Papineau's theory is as follows. He says that the content of a belief is the condition which guarantees that the belief, in combination with different desires, will lead to successful action. But there is no such condition, or if there is, it will contain a lot of irrelevant "background" conditions which intuitively are not part of the content of most beliefs (for instance, that the earth does not blow up). (I seem to remember that Papineau addresses this problem in *Philosophical Naturalism*, but I don't have that book handy.) But do these theories provide a solution to the problem of different contents? Do they deliver the desired verdict of Different Contents in these cases? And are they otherwise satisfactory?

Evaluation

As I said, my view is that these accounts do not apply to sensory representation at all. Experiences have fine-grained contents. But these theories are unable to generate these contents. In short, while input-based theories fall victim to the Problem of Different Contents, output-based theories fall victim to the Underdetermination Problem.

Let us take Papineau's Success Theory first. Consider Twin Maxwell. I have only stipulated the *physical facts* about Twin Maxwell. He is in unitary opponent channel state Q. His sensory system is working as it evolved to work. Q tracks reflectance property F&H. And so on. (For

¹ For instance, Millikan's theory is supposed to be compositional in some sense, while Papineau's theory is not, at least as he typically describes it. She also speaks of mapping, and this is supposed to be central to her account [although others, such as Peter Godfrey-Smith (*Complexity and the Function of Mind in Nature*, pp. 184-187), have argued otherwise], but I cannot see that it will help with the problems I will raise. Likewise for the distinction between direct and derived proper function. In addition, Millikan accounts for the contents of learned concepts by way of the biological functions of parts of the cognitive system that are designed to adapt the organism to new conditions. By contrast, Papineau holds that learning can be a kind of selection process itself, and so a source of new functions. Again, I do not think that this is relevant to what I will say.

more, see Pautz 2006, and “Tracking Intentionalism and Optimal Conditions.”) The case is one of radical interpretation. Given the physical facts as I have described them, what verdict does the Success Theory deliver about this case?

To begin with, let us state a condition of adequacy. Twin Maxwell has color experiences. And his color experiences have contents. And they don’t just have coarse grained contents like there is a round, orange object there. They also have fine-grained contents like there is a round, orange₁₇ object there. So, if the Success Theory is adequate, it must generate assignments of such fine-grained contents. In short, it must avoid the Underdetermination Problem.

But it cannot do this. It attempts to explain experiential content in terms of desire content. But desire contents are typically very coarse-grained. There is more information in our “sensation boxes” than in our “desires boxes”. For instance, Twin Maxwell might have a desire for an orange (or F&H) ball. (The Success Theorist would have to explain how this desire itself got its content naturalistically; but let us just suppose that has been done.) But whether the ball is orange is some other color is a matter of indifference to him. So, this theory will not deliver the verdict that his color experience represents anything like there is a round, orange₁₇ object there. For that is irrelevant to the satisfaction of his desires. Since there is more information in our sensation boxes than in our desire boxes, one cannot explain sensory content in terms of desire content.

Of course, Twin Maxwell might have a desire for a ball with *that very color*, where “that very color” refers to the very color that he visually represents, that is, orange₁₇. And maybe his experience and this desire will lead to action that satisfies this desire if there is an orange₁₇ ball there. Could the Success Theorist say that this makes it the case that his experience represents that there is an orange₁₇ ball there? Could he say that the fine-grained content of this desire is primary, and then pin down the fine-grained content of his experience in terms of it? Obviously not. This procedure is circular. For what makes it the case that John’s desire concerns *that color*, orange₁₇? Intuitively, part of the answer is that *his experience represents* orange₁₇. The idea was to explain the fine-grained representation of colors in terms of the satisfaction of the content of such hypothetical, super specific desires. But the visual representation of those

colors is implicated in the explanation of how those desires acquire their super specific contents. Thus, the account ends up being circular.

In short, when it comes to experiential content and the content of such demonstrative desires, experiential content is primary. So, one cannot reductively explain experiential content in terms of such demonstrative desires. The problem applies also to pain and taste experience. Obviously, the contents of desires about pain and taste are constitutively as well as causally derivative on the contents of our pain and taste experiences. So, one cannot hope to explain the contents of these experiences in terms of desires about pain and taste. In the order of explanation, experience content comes first. In the case of sensory representation, the Success Theory is a non-starter. Hence the Success Theory cannot solve the Underdetermination Problem. The only solution I can imagine makes it circular.

This problem, in my view, is decisive. But it is worth mentioning that, at least in the case of Maxwell and Twin Maxwell, the Success Theory delivers even the wrong assignments of *coarse-grained* contents. The content vehicle of Maxwell's experience is binary opponent channel state *P* and the content vehicle of Twin Maxwell's experience is unitary opponent channel state *Q*. Suppose that Maxwell also has state *D1* and Twin Maxwell has state *D2*. Further, suppose that when conditions are optimal, *D1* brings it about that Maxwell grasps an F&H object; and likewise, when conditions are optimal, *D2* brings it about that Twin Maxwell grasps an F&H object. Then both *D1* and *D2* are, on Papineau's theory, desires with the content *that one grasp an F&H object*. Further, *P* and *D1* lead to successful action if there is an F&H object before Maxwell. And *Q* and *D2* lead to successful action if there is an F&H object before Twin Maxwell. So, it would seem that Papineau's Success Theory delivers the verdict that both *P* and *Q* have the coarse-grained content *there is an F&H object there*. Hence it delivers the mistaken verdict of Same Content and (given Intentionalism) Same Experience in this case. Likewise, it seems to me, in the case of Yuck and Yum and Mild and Severe. But this, I think, is not the main problem with this view. The main problem is the Underdetermination Problem.

Now let us consider Millikan's Consumer Theory. As far as I can tell, this theory does not generate *any unique predictions at all* about the content of Maxwell's binary opponent channel state *P* or the content of Twin Maxwell's unitary opponent channel state *Q*, not even coarse-

grained ones. Consider Twin Maxwell. I guess the relevant consumer systems here are the belief-desire system and the motor-system. Typically, Twin Maxwell's "consumer devices" do not really use Q to do much of anything. Maybe our consumer devices use Q to engage in generic behaviors – the same kind of behaviors that our other opponent channel states get us to engage in. For instance, they might use Q to lead Twin Maxwell to reach out. On another occasion they might use Q to allow Twin Maxwell to distinguish some F&H object from some object with another reflectance property. What is the *unique* Normal condition required for these consumer devices to use Q to produce these behaviors in their historically Normal way? (By behaviors here, I mean behavior-types construed abstractly without reference to any particular currently existing objects.) What is the *unique* external condition C such that there consumer devices that use Q to grasp and distinguish proliferated because, in the past, when C obtained and the consumer devices used Q to grasp and distinguish, then this behavior had good consequences? There perhaps is a general condition which obtained when such behavior had good consequences: *that there is an object in front of the subject, which is different from other objects*. But this condition is not unique to Q. What we want is a theory that generates a *different* content for each different state of the color vision system. I do not see how the Consumer Theory can do that. In short, it faces the Underdetermination Problem.²

To bring the point home, consider the following example. Suppose Twin Maxwell is looking at two objects. Suppose that his visual experience of the first has the content orange₁₇ object present and his visual experience of the second has the content orange₁₉ object present; and that his the content-vehicle of the former experience is Q₁₇ and that of his latter experience is Q₁₉. If the Consumer Theory is correct, then it

² And I do not see how appeal to derived proper functions or mapping rules would help here.

I see another problem with the Consumer Theory of sensory representation. It seems that the condition required for the consumer devices to use Q perform their function in guiding behavior in the historically normal way is not merely an *external* condition: the subject's *background beliefs* must also be true. So the Consumer Theory seems to entail that the content of an experience will be something like: *that there is an F&H object there, and one's background beliefs are true*. Of course, this is not right.

must be able to generate these fine-grained assignments of content. But this is so iff (i) Q17 and Q19 are used by the consumer devices in different ways, and (ii) there are distinct external conditions which enter essentially into the historical explanation of the proliferation of consumer devices that use Q17 and Q19 in these different ways. But this is just not plausible.³ So, the Consumer Theory succumbs to the Underdetermination Problem.

The problem is made worse by the fact that many consumer activities may have no adaptationist explanation at all, and *a fortiori* no explanation in terms of an external condition. Many features of our neural wiring, and hence many features of our behavior, are not the result of selection pressures, but of chance mutations, genetic drift, and so on. Suppose that objects a and b produce in us opponent channel state Oa and Ob, and b and c produce in us opponent channel states Ob and Oc; and that Oa and Ob are more similar to one another than Ob and Oc are similar to one another. Then we might be more disposed to sort together a and b than b and c. What is the salient external condition that explains why this consumer activity evolved? There might not be one. This behavior may have been adaptively neutral, so that the mutations that lead to it were not selected against. Only a fanatical adaptationist would

³ Part of the explanation for the falsity of (ii) is that chromatic properties are fairly acausal. They don't really affect how objects interact with one another or with humans. Even if Realism is correct and objects had colors prior to the evolution of color vision, it is not the case that a particular chromatic property of objects will enter essentially into the explanation of why we evolved a consumer system that uses Q17 to produce certain behaviors under conditions of well-functioning. (By contrast, consider shape. A certain shape demands a certain behavioral response. So, for instance, *the presence of a round object* might enter essentially into the explanation of evolution of consumer devices that use a certain internal state to guide circle-appropriate behavior.) Instead, what explains why we evolved the color vision system we did evolve (to the extent that it has an adaptationist explanation) are general features of our ecology: our particular habits, dietary needs, predators, and environment. There is not a unique explanation for the evolution of every opponent channel state which essentially concern particular chromatic property. Hence the Consumer Theory does not generate the assignment of *any* color contents to our opponent channel states.

claim every feature of our behavior has some adaptive explanation in terms of some external circumstance.

So, Millikan's Consumer Theory does not deliver a unique content-assignment at all – even an assignment of coarse-grained content – in the case of Maxwell and Twin Maxwell as originally described. I venture to say that the same is true in the case of Yuck and Yum and Mild and Severe. Like Papineau's Success Theory, Millikan's Consumer Theory cannot be applied to sensory representation. Like Papineau's theory, it succumbs to the Underdetermination Problem.

But I can imagine a twist on the case of Maxwell and Twin Maxwell in which it does deliver a verdict – only a verdict which is wrong twice-over. In the new version of the case I have in mind, there is a salient external condition which explains a certain consumer activity; but it is intuitively not the content (or anyway, not the only content) of any color experience. Suppose that F&H objects are almost always a nutritious fruit. Thus, Maxwell's community so evolved that when P is tokened as the result of an F&H object the consumer devices use P to produce the obtaining and consumption of the F&H object because such objects are nutritious. Likewise, Twin Maxwell's community so evolved that when Q is tokened as the result of an F&H object the consumer devices use Q to produce the obtaining and consumption of the F&H object because such objects are nutritious. Now, the salient condition which explains proliferation of consumer devices which use P to produce such behavior in Maxwell's community is the presence under these circumstances of a *nutritious fruit*, not the presence of an F&H object. Likewise, the explanatorily salient condition which explains proliferation of consumer devices which use Q to produce such behavior in Twin Maxwell's community is the presence under these circumstances of a *nutritious fruit*, not the presence of an F&H object. Thus, Millikan's theory entails that P and Q both represent *the presence of nutritious fruit*. Hence it delivers the verdict of Same Content and given Intentionalism) Same Experience in this case.

Of course, this verdict is wrong twice over. First, Maxwell and Twin Maxwell's experiences do not represent the presence of nutritious fruit. Or at least, they do not *only* represent this condition. They represent many fine-grained colors. The Consumer Theory does not have the resources to generate fine-grained content concerning colors. This is the Underdetermination Problem. But the problem is not so much that Millikan's Consumer Theory does not assign *enough* contents; rather it is

that it assigns contents not contents at all (as in the original Maxwell/Twin Maxwell case) or contents of *completely the wrong kind* (as in case of Maxwell/Twin Maxwell with a twist). This is a problem that arises in other sense modalities. Second, given the neural and behavioral differences between them, the verdict of Same Content and Same Experience is not plausible. Counterexamples of this general kind could be multiplied.

If it is denied that the Consumer Theory generates the verdict that P and Q represent *nutritious fruit*, just consider the frog case again. Here Millikan holds that her theory generates the assignment of the biologically relevant condition *frog food*, not the assignment of any biologically irrelevant sensory property such as *moving black dot*. By analogy, one would expect that in the case of Maxwell and Twin Maxwell with a twist it generates the assignment of the content *nutritious fruit present*, not *F&H object present*.

To some extent, it is not so obvious that input-based theories face the Underdetermination Problem, for they can attempt to pin down the fine-grained, biologically irrelevant contents of sensory experiences in terms of their distal causal antecedents. An object's having a certain fine-grained sensible quality might not be relevant to the satisfaction of our desires, or to the evolution of our consumer devices. Nevertheless, on an input-based theory, our visual systems might represent it by *tracking it under conditions of well-functioning*. However, I believe that the phenomenon of standard variation means that input-based theories do ultimately fall victim to the Underdetermination Problem as well.⁴ And, of course, input-based theories fall victim to the Problem of Different Contents. They deliver the mistaken verdict of Same Contents in my cases.

I conclude that the Underdetermination Problem means that the Success Theory and Consumer Theory cannot be applied to sensory representation. Further, the problem is not merely one of detail. By their very nature, output-oriented theories are bound to fall victim to this problem, for just by going by output factors a theory cannot assign to our experiences the fine-grained contents they actually have. Such fine-

⁴ "Reductive Intentionalism and the Problem of Standard Variation", MS <https://webpace.utexas.edu/arp424/www/index.html?uniq=tpfy1b>

grained conditions are often irrelevant to the satisfaction of our desires, and often irrelevant to the evolution of our behavior. What is relevant to the satisfaction of our desires and to the evolution of our behavior are coarse-grained contents like *nutritious fruit*. This, I think, is the *main* problem with these views.

But there is another problem. Insofar as these theories deliver verdicts, they deliver the mistaken verdict of Same Content, at least the case of Maxwell and Twin Maxwell. More generally, there is no reason to expect that these theories would deliver the verdict of Different Contents in my cases. For, even though the individuals evolved to have different behavioral dispositions, those different behaviors may be *successful* under the *same* external conditions. Then the Success Theory will deliver the mistaken verdict of Same Content. Further, the evolution of those different behavioral patterns (the evolution of consumer devices which use our sensory states to produce such behavior) might be best explained by citing the *same* external conditions (for instance, the presence of an object with a certain reflectance property, a bodily disturbance in the leg, or the presence of an object with a certain chemical property). Then the Consumer Theory will deliver the mistaken verdict of Same Content in the cases of Maxwell and Twin Maxwell, Mild and Severe and Yuck and Yum. But, to repeat, the main problem with these theories is the Underdetermination Problem; this problem, I think, means that these theories are non-starters when it comes to sensory representation.

Therefore, these theories, far from being superior to input-based theories of sensory representation, seem to me to be actually inferior.

3. Output-Based Theories: Lewis

A quite different output-based theory, which in my view is superior to those of Papineau's and Millikan's, is defended by David Lewis (1974). He writes (1983, note 2):

The right assignment of content to Macbeth's [sensory] states will be the one given by the best general rule of assignment. The best rule will be the one that does best at assigning contents that rationalize behavior, according to the principles of common sense psychology.

Note that in this formulation of the view Lewis does not mention input-oriented factors, only output-oriented factors. Here is a more detailed statement of a theory inspired by Lewis. On this theory, x sensory represents property y , that is, x is conscious of property y , iff:

x is in brain state S and the Best Narrow Interpretation of the members of K (where K is the kind or species to which x belongs) assigns to S the experience of y , where the Best Narrow Interpretation of the members of K is the one that best satisfies the Narrow Constraints on Interpretation, given the Narrow Functional Roles of their brain states.

We might call this the *Narrow Interpretation Theory*. I will not unpack this here, as I have discussed this theory elsewhere.⁵ Following Lewis, Stephen White (1994) and Frank Jackson and David Braddon-Mitchell (1996, 181) defend theories of this type.

This theory strikes me as superior to Success Theory and the Consumer Theory. For instance, the assignment of a content to an experience might make some behavior intelligible even if its truth does not enter into the explanation of the *evolution* of that type of behavior (thus solving a problem with Millikan's Consumer Theory), and even if it is not relevant to the satisfaction of the subject's desires (thus solving a problem with Papineau's Success Theory). Further, one might think that it generates the right verdict of Different Content and hence Different Experience in my cases, for the individuals in those cases differ considerably in their behavioral dispositions. Thus, the assignment of different sensory contents is required to make intelligible this different behavior.

But elsewhere (see note 3) I argue that this theory runs afoul of the same kind of problem which afflicts the Success Theory and the Consumer Theory: the Underdetermination Problem. For there is no backward road from behavior to phenomenal content. The common sense view is that phenomenal content controls behavior. But there is no algorithm for going from behavior back to phenomenal content. This problem, I think, is endemic to output-based theories. Output-oriented

⁵ "Sensory Consciousness Awareness as Irreducible",
<http://philrsss.anu.edu.au/people-defaults/adamp/index.php3>

facts alone are insufficient to pin down the fine-grained content of our experiences.

4. Intentionalists Cannot be Reductionists: A Sketch of an Alternative Picture of Sensory Consciousness

Stephen Schiffer's *Remnants of Meaning* (1987) is, among other things, an attempt to demolish every reductive theory of intentionality. He concluded that there is no such thing. Jerry Fodor (1989) responded by saying, "the trouble with this way of arguing is that it's convincing only if you have a cat for every mouse hole . . ." One can always reply that there are paths through the thicket that have not been shown to be dead ends, according to Fodor.

I am trying to do for theories sensory intentionality what Schiffer did for theories of intentionality in general. Intentionalists about consciousness might respond to my argument in one of two ways:

Optimism: Ok, so we don't have a correct reductive theory of sensory intentionality. But there is one. This is Fodor's attitude transposed to sensory intentionality.

Pessimism: There is no correct reductive theory. Sensory intentionality is primitive. This Brentano's attitude transposed to sensory intentionality.

We might call the first option *Optimistic Reductive Intentionalism* and the second option *Primitivist Intentionalism*. Optimistic Reductive Intentionalism is defended by Lycan in his comment. And it is suggested by some remarks of Byrne and Tye. What is my argument for Pessimism over Optimism? What is my argument for Primitivist Intentionalism over Optimistic Reductive Intentionalism?

I must admit I see the attraction of Optimistic Reductive Intentionalism. I don't want a weird view. I don't want to believe in emergence, or anything ugly like that. There are general arguments for Reductionism. So maybe we should be Reductionists even if we cannot specify any correct reductive theory. But I think that there are two countervailing arguments for Primitivist Intentionalism over Optimistic Reductive Intentionalism. The first is a negative argument and the second is a positive argument. I find these arguments convincing. Fodor says someplace "if intentionality is real, it must be something else". At

least when it comes to this issue, I would recommend a different slogan: “everything is what it is and not another thing.”

First, the negative argument. Consider the following schema:

The relation x *sensorily represents (is sensorily conscious of) property* y = the relation x y

If Reductive Intentionalism is true, then there is some true completion of this schema in which the dots are replaced by broadly physical and function stuff. The negative argument points out that we cannot come close to providing such a completion.

But there is more to the argument than this. The argument is not merely an argument from elimination. Then Fodor’s style of response might be appropriate. Rather, I think that there is *principled* reason to think that the above has no true completion. We know, in broad outlines, the physical facts about human beings. (Maybe we don’t know weird micro stuff, but *that* is not going to help us with *this* problem.) Solutions to the problem come in two sorts: input-based and output-based. By their very nature, input-based theories fall victim to the Problem of Different Contents. By their very nature, output-based theories fall victim to the Underdetermination Problem. Since every reductive theory is bound to fit into one of these categories (plus or minus a bit), every reductive theory is bound to fail.

So much for the negative argument. Now for the positive argument. The positive argument extols the virtues of being a Primitivist. If you are a Primitivist Intentionalist about sensory consciousness, you can easily solve both the Problem of Different Contents and the Underdetermination Problem.

On Primitivist Intentionalism about sensory consciousness, the relation x *is conscious of property* y is primitive. So if we accept this view, we are free from the shackles of reductive theories. We do not have to say it supervenes on input-oriented facts: on what external properties our brains states track. We are free to say that it supervenes, at least in part, on internal facts, such as opponent processing, S1 activity, and across-fiber patterns. So, we can say that Maxwell and Twin Maxwell bear this relation to different color properties in part because of the internal differences between them. Likewise for the other cases. Thus, we solve the Problem of Different Contents.

Now consider the Underdetermination Problem. Once we accept Primitivism, we can solve this problem too. We do not have to accept an output-oriented account either. Such accounts face the Underdetermination Problem. Instead, we can say that the fine-grained content of experience is determined by our fine-grained neural processing. Further, from this point of view, we see why output views are bound to fail. True, neural processing determines behavior. But, in this process, there is a “loss of information”. Not all of the neural detail (and hence not all the phenomenological detail) is exhibited in our behavior. Further, there is no algorithm from going from behavior back to the phenomenological detail. Why should there be?

In a way, this view combines both Externalism and Internalism. It is Externalist because it is a version of Intentionalism, and there is a trivial sense in which Intentionalism is externalist. It views experience as a relation to propositions built up from external-world properties. As abstract objects, propositions aren't anywhere, and hence are not “in the head”. Thus, one might say that it is a version of *Ontological Externalism* about phenomenology. In another way, the view I describe here is Internalist, because it says that what propositions we sensorily represent is internally-determined. So, it is a version of *Modal Internalism*. Experience reaches out into the world. But what it gloms onto depends on internal facts.

But the kind of Modal Internalism I advocate is moderate. I would not say that a lone opponent channel state in a pitre dish, removed from a working brain, issues in sensory consciousness. And I am not sure about brains in vats. It may be that, for the brain to result in sensory consciousness, it must be embodied, and issue in behavior, and so on. How could we know one way or another? What I do insist on is the verdict of Different Experiences in cases like Maxwell and Twin Maxwell. Their brains are embodied; they undergo quite different processing; and they result in different actions. I say that they have different experiences. This is as far as my Internalism goes. It is not the same as what “Internalism” usually means in the literature.

On Primitivist Intentionalism, there is no derivation (not even an a posteriori one) of facts about sensory consciousness from the physical facts, together with some reductive theory. In this sense, we cannot account reductively for the presence of consciousness in the world. But I

think this is just something we have to live with. I still think that the balance of consideration tilts in favor of Primitivist Intentionalism.

One option at this point would be to reject Intentionalism and go back to the old reductive project: reducing experiences to brain states, or psychofunctional states, or whatever. But I think that Intentionalism, or something very much like it, has to be true. So, I opt for Primitivist Intentionalism.

But, again, while Lycan and I disagree about Reductionism vs. Primitivism, we agree about other things: that the problem of consciousness reduces (more or less) to the problem of naturalizing sensory intentionality, so that this is where the debate over the nature of consciousness should focus; and that input-based theories fail. I have profited from his comments on these issues.

References

Fodor, Jerry A. (1990). *A Theory of Content and Other Essays*. The MIT Press

Fodor (1998) *Concepts*, Oxford.

Fodor, Jerry. 1989 "Review Essay: *Remnants of Meaning* by Stephen Schiffer."

Jackson and David Braddon-Mitchell, 1996 *Philosophy of Mind and Cognition*, Blackwell.

Lewis, David 1974 "Radical Interpretation", *Synthese* 23

Lewis, David 1983. "Individuation by acquaintance and by stipulation". *The Philosophical Review* 92:3-32

Millikan, Ruth. 1989, "Biosemantics", in *Journal of Philosophy*, 86: 281-97.

Papineau, David. 1993, *Philosophical Naturalism* Blackwell, Oxford.

Papineau, David. 1998, "Teleosemantics and Indeterminacy", in *Australasian Journal of Philosophy*, 76: 1-14.

Pautz (2006) "Sensory Awareness is not a Wide Physical Relation", *Nous*.

Pautz (MS) "Reductive Intentionalism and the Problem of Standard Variation", <https://webspace.utexas.edu/arp424/www/index.html?uniq=tpfy1b>

Pautz (2005) "Sensory Awareness as Irreducible"

<http://philrsss.anu.edu.au/people-defaults/adamp/index.php3>

Peacocke, C., 1992, *A Study of Concepts*, Cambridge, MA: MIT Press.

Peacocke, C. 1993. "Externalist Explanation." *Proceedings of the Aristotelian Society* XCIII: 203-230

Schiffer, Stephen 1987 *Remnants of Meaning*, MIT

Stalnaker, Robert. 1998 "What Might Nonconceptual Content Be?"
Concepts, ed. by Enrique Villanueva. Atascadero, CA, Ridgeview
Publishing Co., 339-52.

White, Stephen. 1994. Color and notional content. *Philosophical Topics*
22, 471-503