

Draft of 12/2003. I plan to divide this paper into two.

Sensory Awareness as Irreducible: From Internalist Intentionalism to Primitivism¹

Adam Pautz
University of Texas at Austin
The Australian National University
apautz@mail.utexas.edu

ABSTRACT: This paper is a companion to “Sensory Awareness is not a Wide Physical Relation: An Empirical Argument against Externalist Intentionalism” (forthcoming, with a reply by Michael Tye, in *Nous*). I generalize the argument of that previous paper. I attempt to show that, given the mild form of Internalism argued for in the previous paper, it is not only the case that the awareness relation cannot be identified with a wide physical relation; it also cannot be identified with a narrow physical relation, for lack of any candidates with the right properties. This concludes my case for the Brentano/Meinong view that the awareness relation, or the sensory representation relation, is a primitive, non-physical relation. So, in my view, the case for Primitivism need not rest on the alleged conceivability of Zombies or the like. Primitivism about sensory consciousness may be proved on the basis of the Relational View of sensory consciousness together with a mild form of Internalism.

§1. Introduction

I am going to develop an argument against Physicalism concerning qualitative mental properties. Unlike most arguments against Physicalism, it is not based on the usual *a priori* considerations, such as what Mary learns when she comes out of her black and white room or the apparent conceivability of Zombies. Rather, it is based on two broadly *a posteriori* premises about the structure of experience and its physical basis.

Physicalism is usually defined in modal terms. Roughly, it is defined as the thesis that the pattern of instantiation of all properties supervenes with metaphysical necessity on the pattern of instantiation of the basic physical and topic-neutral properties. In other words, Physicalism is nowadays usually equated with **Physicalist Supervenience**.

We may distinguish between two forms of Physicalism about qualitative mental properties. **Identity Physicalism** holds that qualitative mental properties *just are* the complex physical or topic-neutral properties on which they are alleged to supervene: for instance, complex neurobiological properties of the brain or functional properties (narrow or wide) realized by such properties. It is a broadly reductive view which encompasses both Type-Type Identity Theory and the various forms of Functionalism. Identity Physicalism is the majority position. **Dualistic Physicalism**, by contrast, holds that qualitative mental properties are not identical with such properties. It is non-reductive in a very strong sense: on this view not only are qualitative mental properties not first-order neurobiological properties (a fact which most would grant on the grounds of multiple realizability), they are not even second-order functional properties realized by such first-order

¹ For helpful discussion or comments, I would like to thank Ned Block, David Chalmers, John Hawthorne, Christopher Peacocke, David Robb, Stephen Schiffer, and members of the NYU Friday Forum in 2002.

properties. But the Dualistic Physicalist nevertheless holds that the pattern of instantiation of qualitative mental properties supervenes with metaphysical necessity on the pattern of instantiation of such properties, thus securing the supervenience thesis which is usually taken to be definitive of Physicalism. Dualistic Physicalism is a minority view, but it has adherents (McGinn 1991, Nagel 2001).² Most Physicalists are Identity Physicalists of some kind or another.

The argument against Physicalism I will be developing proceeds in two stages. In the **first stage** I argue, on the basis of two broadly empirical premises about the structure of experience and its physical basis, that Identity Physicalism is false in the case of qualitative mental properties. In other words, I use these two premises to argue that there is a gap – an ontological gap – between a person’s qualitative mental properties and her physical and functional properties. The Physicalist might retreat to Dualistic Physicalism. But, in the **second stage**, I argue that we have no reason to accept Dualistic Physicalism over traditional Property Dualism. Indeed, it is hard to tell the two apart. Once we accept that Identity Physicalism is false, as I think we must, we also must accept that we have no reason to believe in Physicalist Supervenience.

So my principal aim is not to show that Physicalism is *false*, but that (given the failure of Identity Physicalism) we have no reason to accept it. My main purpose is to argue for a **skeptical position** with respect to Physicalism.

My plan is as follows. As noted, the first stage of the argument depends on two broadly empirical premises about the structure of experience and its physical basis. In §2 and §3, I say what those premises are and why we should accept them. In §4, I develop the first stage of the argument, using these two premises to rule out Identity Physicalism. In §5, I develop the second stage of the argument: I argue that Dualistic Physicalism is no better than Dualism. In §6, I offer some concluding remarks.

§2. The First Premise: Relationism about Qualitative Character

Let us say that R is an **experience relation** iff bearing R to O involves experiencing O or some part of O. Then my first premise can be stated as follows:

Relationism There is an experience relation R and a corresponding class of items Q such that, for any possible or actual individuals x and y, x and y have experiences with the same qualitative character (or more simply, have the same experience) iff x and y stand in R to the same items in Q or items in Q that have the same properties. The items in question might be sense data, the properties we perceive in the outside world and our own bodies, propositions about one’s environment and body, physical objects, events, Meinongian objects – in general, whatever there may be.

Put in ontological terms rather than in modal terms, what the Relationist claims is that qualitative mental properties are certain kinds of **relational** properties. For instance, let us say that a color experience is **reddish** iff what it is like to have it resembles what it is actually like to have a color experience produced under normal conditions by paradigmatically red things. The Relationist will say that the property of having a reddish experience is a relational property of the form *bearing R to O*, where R is an experience relation and

² Some people may disapprove of my labeling Dualistic Physicalism a brand of ‘Physicalism’ at all, reserving that label for Identity Physicalism only. I take it that there is no fact of the matter about what is and is not a form of Physicalism. ‘Physicalism’ is a term of art. Here I will use ‘Physicalism’ to mean Physicalist Supervenience. Dualistic Physicalism upholds Physicalist Supervenience. So it counts as a form of Physicalism, under this stipulation.

O is an item of some kind – as it might be, a red sense datum, the universal redness, a Russellian proposition into which redness enters as a constituent, or whatever.

Many theories of qualitative character endorse Relationism (for an inventory, see Table 1 below). Different forms of Relationism differ on what the items in the class Q are and how to characterize the relevant experiential relation R which we bear to them. For instance, the Sense Datum Theory holds that qualitative character is determined what *sense data* we *sense*, Intentionalism holds that it is determined by what *propositions* we *sensorily entertain*, and so on.

*Sense Datum Theory: A experiences a **red sense datum*** (Bertrand Russell, H. H. Price)

*Sensationalism: A experiences a **red’ region of his visual field*** (Christopher Peacocke)

*Multiple Relation Theory: **Something** looks **red** to A* (William Alston)

*Property Theory: A experiences **redness*** (or maybe a complex property which has redness as a constituent) (George Bealer, John Foster, Colin McGinn, Mark Johnston, possibly Fred Dretske)

*Intentionalism (‘Representationalism’, ‘Representationism’): A sensorily represents **redness**. Or: A sensorily entertains **that p***, where *sensorily entertaining* is a *sui generis* propositional attitude that is only involved in experience and **that p** is a proposition containing redness itself, or mode of presentation of redness, or the ‘phenomenal property’ corresponding to redness. (Gilbert Harman, Michael Tye, William Lycan, possibly Fred Dretske, Alex Byrne, David Hilbert, Sydney Shoemaker, and many others)

Table 1: Some Forms of Relationism, and their Accounts of Having a Reddish Experience

The most notorious form of Relationism is the Sense Datum Theory. Suppose Maxwell has a hallucination of a pink elephant. Here is one way to motivate the Sense Datum Theory. Maxwell is evidently aware of the color pink. So he must be aware of a pink object. It is not a physical object – there is no pink physical object in Maxwell’s vicinity. So it must be a mental object: a pink sense datum. Since veridical experience can be indistinguishable from hallucinatory experience, we must be aware of mental sense data in veridical experience as well. A veil of ideas is placed between us and the external world.

This argument relies on the principle that if one is aware of property *P* (in this case the color pink) then one must be aware of an object that has *P*. Call this the **Principle of Instantiation**. Contemporary Relationists resist the argument by rejecting the Principle of Instantiation. One popular way to do this is to view experience as a species of propositional attitude. On one version of the idea, to be directly aware of a property *P* is to “sensorily entertain” a proposition in which this property figures as a “constituent”. Here *sensorily entertaining* is a *sui generis* propositional attitude relation which is only involved in experience. Now one can believe the proposition *that something has the property of being omniscient* even if nothing has this property. Likewise, it is held that one can sensorily entertain the proposition *that something at viewer-relative place p has property F* even if there is nothing that has property *F* at *p*. And so on this kind of view one can be aware of the property *F* even if there is no *F* object present: the Principle of Instantiation fails. This is the method of the **Intentionalists** (Harman 1990, Dretske 1995, Tye 1995, Lycan 1996). Other forms of Relationism on which the Principle of Instantiation fails include the Multiple Relation Theory and the Property Theory. The availability of such theories shows that one may accept Relationism while rejecting the Sense Datum Theory and an indirect theory of perception according to which we perceiving external objects only by perceiving interposing mental objects. (Thus Relationism is broader than ‘act-object’ theory if that is taken to mean just the Sense Datum Theory.)

Traditionally, Adverbialism is the main form of Anti-Relationism (Tye 1984). Defenders of Adverbialism agree that to have a reddish experience is to see red, but they analyze *seeing red* as *seeing redly* and deny that seeing redly involves standing in experience relation to the color red or to anything else.

It may be that some qualia-based theories of experience are also forms of Anti-Relationism. Qualia theorists say that the character of an experience is determined by its 'qualia'. They often define 'qualia' *via negativia* as non-intentional features of experience. This definition of 'qualia' makes the qualia theory compatible with *non-intentional* forms of Relationism, such as the Sense Datum Theory or Peacocke's Sensationalism. So it is not always entirely clear that qualia theorists mean to reject Relationism. But if qualia theorists hold that to have an experience with a certain quale is not a matter of standing in an experience relation to anything, then the view they defend is a form of Anti-Relationism.

The formulation of Relationism given above is very strong. One might wish to weaken it in various ways. (i) One may not wish to extend Relationism to all experiences. For instance, it may not be plausible to apply it to moods and emotions (but see Tye 1995, 2002). (ii) It may not be plausible to say that there is only one experience relation relevant to the determination of qualitative character. Maybe there is more than one, and what experience relation we bear to an item (a property, an intentional content, or whatever) can make a difference to qualitative character, as when we go from *seeing* a shape to *feeling* it or from *attending to* a sound to *not attending to* it. (iii) One may wish to restrict Relationism to veridical experiences. (This seems to be what the Disjunctivists do. See Hinton 1973.)

The argument I will be developing works on weakened forms of Relationism. But for the sake of simplicity I will work with the strong form of Relationism formulated above.

Why accept Relationism? Briefly, there are three kinds of arguments for Relationism. First, there is the phenomenology of experience. As it is sometimes put, experience is transparent: when we try to inspect the character of an experience we always end up focusing on what it is an experience of (Tye 2002). I can focus now on one feature of the content of my experience, now on another. The point applies equally to hallucinatory experience. Of course, the nature of the items we are related to might not be evident on the basis of the phenomenology. It is not evident whether they are physical objects, mental particulars, universals, tropes, possible or actual states of affairs, Meinongian objects, or whatever. But it seems evident on the basis of the phenomenology that we are related to items of *some kind or other*. Second, it seems that experience necessarily affords us with knowledge of items other than ourselves. For instance, necessarily, if you have an experience of red, orange and blue, whether it is hallucinatory or not, you are in a position to know that the first color resembles the second more than the third. How can this be unless in having the experience one stands in some kind of experience relation to the colors? Third, there is the fact that we use relational language to describe experience, even hallucinatory experience. (Jackson (1977) argues on semantic grounds for the Sense Datum Theory, but I believe that his arguments could be marshaled in favor of the other forms of Relationism.) Asked what Mary's color experience is like, we might say that she sees red. And from this we can infer that there is a color that Mary sees. The point applies equally to non-veridical experience. For instance, suppose that Jack and Jill are undergoing hallucinations. It might nevertheless be true that Jack sees every shade of color that Jill does. From this together with the premise that Jill sees a shade of red we can infer that Jack sees a shade of red. Given a Relational account of qualitative character, we can make good sense of these statements and inferences. By contrast, Anti-Relationism has notorious difficulties here. (I develop these arguments in greater detail in Pautz 2004b.)

For these reasons, and because it has become recognized that one may be a Relationist while avoiding mental objects, Relationism about qualitative character now enjoys widespread acceptance. In fact, I would go so far as to say that most philosophers of perception now accept some form of Relationism. Even once staunch Non-Relationists (notably, Michael Tye and Sydney Shoemaker) have converted to some version of

Relationism. So (if they also accept my second premise) they will face the problem for Physicalism which I develop here.

The purpose of the first stage of my argument is to show that, given Relationism and a second premise still to be introduced, Identity Physicalism is not true of qualitative mental properties. On some versions of Relationism Identity Physicalism can be ruled out immediately. Consider the Sense Datum Theory. On this view to have a reddish experience is to sense a red sense datum. Neither sense data, nor our sensing of them, can be identified with anything physical. For, on this view, when one has a hallucination of a pink elephant, one is aware of a pink object. But there need not be any pink *physical* object in one's environment or brain. So the pink object one is aware of must be a non-physical object.

But as I mentioned before, there are ways to be a Relationist while avoiding sense data. The basic idea is that the qualitative character of experience is determined by what properties our experiences (“non-conceptually”) represent things as having, or what properties things “phenomenally” look to have. And our experience can represent that something has a certain property, even though nothing – mental or physical – has that property. In this sense, experiences are a species of **intentional states**.

For the sake of argument, I will assume that this kind of view is correct. Following Fred Dretske, I will call the relation we bear to properties in experience **p-awareness** (Dretske 1999, p. 161). It must be distinguished from the awareness of objects (what Dretske calls *o-awareness*) and the awareness of facts (what Dretske calls *f-awareness*). In hallucination, you are p-aware of properties, but not o-aware of any objects that have those properties. Nor are you f-aware of any facts about any objects. I will call the properties the p-awareness of which determines qualitative character the **Q-properties**. I will be focusing on color experience throughout. I will assume that the qualitative character of color experiences is determined by what **colors** we are p-aware of.³

On this kind of view, experience does not involve sense data or any other strange non-physical objects. So there is no principled reason why Identity Physicalism must fail. The p-awareness relation is a kind of *intentional* relation in the sense that one can be p-aware of a property – one can have an experience that represents (that something has) a certain property, even if nothing in one's environment actually has that property. To reductively explain how it is that we are p-aware of properties – how it is that our experiences represent properties – in physical/functional terms, one might appeal to one of the many reductive approaches to intentionality now on offer: indicator semantics, teleological semantics, the asymmetric dependence account, or whatever. Relationism, then, does not by itself create a problem for Identity Physicalism. This is where my second premise comes in. My second premise means that the usual methods for finding a place for intentionality in the physical world do not work in the case of sensory intentionality.

³ Some comments. (i) We are typically aware of ostensible properties “at places”, but I will ignore this in what follows, and simply speak of the awareness of properties *simpliciter*. (ii) Some Relationists might say that “sensorily entertaining a proposition” is more basic than being p-aware of a property. They might say that A is p-aware of a property P iff A sensorily entertains a proposition of the form *there is something that has P at viewer-relative place p*. They might say that the most “basic” relation involved in the determination of qualitative character is a relation to propositions, and not a relation to properties (Pautz 2003b). The argument that I will be making carries over to this kind of view. (iii) Some philosophers think that modes of presentation, and not just properties, are somehow involved in the determination of qualitative character. Again, the argument I will be making is not significantly affected if this kind of view is right. (iv) Sydney Shoemaker (1994) calls the properties the awareness of which determines the qualitative character of color experience ‘phenomenal properties’ and refuses to call them ‘colors’. I think that this is a (verbal) mistake on Shoemaker's part (Pautz 2003b). He also thinks that these properties are certain kinds of dispositions to produce experiences. But again, the argument that I will be making works equally well if this kind of view is right. In general, the argument does not depend on what view is taken of the nature of the relation of sensory consciousness.

§3. The Second Premise: Internalism about Qualitative Character

My second premise may be stated as follows:

Internalism What properties we are p-aware of (what properties things look to have), and thereby the qualitative character of experience, is wholly determined by what goes on in the brain.⁴

So I combine the thesis that experience is relational with the claim that it is internally-determined. Experience is essentially relational, but the relation is fixed by the internal physical state of the subject. We might call the resulting view **Internalist Relationism**.

Some Relationists may argue for Internalism on **intuitive** grounds. They may argue as follows. Intuitively, qualitative character is internally-determined. This is “given to us by the light of reason.” By Relationism, qualitative character is a matter of what properties we are aware of. So what properties we are aware of must be internally-determined. (At one point Michael Tye was an Internalist Relationist and argued for Internalism by appealing to intuition. See Tye 1994, p. 168.)

By contrast, I argue for Internalism on **empirical** grounds. My reason for claiming that qualitative character is internally-determined is that the best explanation of the qualitative character of experience lies in our neurobiology.

For instance, consider color vision. The colors that we experience vary along four chromatic dimensions: red, green, yellow and blue. *Unitary colors* (reds, for instance) have a value on one of these dimensions but not the others, while *binary colors* (oranges and purples, for instance) have a value on both. But some binary combinations are ruled out: we do not experience red-greens or yellow-blues. These facts about the phenomenology of color experience are best explained by our neurobiology. There are two “opponent channels” – the R-G channel and the Y-B channel – in the brain. Each can assume a positive or negative state of activation. When the R-G channel is in a positive state of activation we are aware of a reddish color; when it is in negative state of activation we are aware of a greenish color. Likewise for the Y-B channel. And there is a good correlation between the values that the opponent channels have on the neural parameters *positive/negative R-G activity* and *positive/negative Y-B activity* and the values the colors we experience have on the phenomenal parameters *red/green* and *yellow/blue*. Thus we are p-aware of a unitary color when one channel is active but the other is not. We are aware of a binary color when both channels are active. For instance, we are p-aware of a balanced orange (an orange that is equally reddish and yellowish) when the R-G and Y-B channels are put into equal positive states of activation. We are not p-aware of red-greens or yellow-blues because neither channel can assume a positive and negative state of activation at once. So:

Internalism_{ev} What colors we are p-aware of (what colors things look to have) at a time is completely determined by the levels of activity in our R-G and Y-B opponent channels at that time.

This is the so-called **opponent process theory** of color vision brought to the attention of philosophers by C. L. Hardin (1988). (The version I have presented is vastly oversimplified, but not in a way that will

⁴ In formulating Internalism as I do here I assume Relationism. One might formulate Internalism in a more theory neutral way, namely, as the thesis that qualitative character is internally-determined, leaving it open what qualitative character consists in. But as I believe that qualitative character is a matter of what Q-properties we are p-aware of, I will continue to formulate Internalism as the thesis that what Q-properties we are aware of is internally-determined.

affect the points I will be making.) It is one of the great success stories of vision science. It explains a bewildering variety of additional phenomena besides the structure of color appearance: the forms of color deficiency, complementary afterimages, wavelength discrimination, color naming, temporal resolution for flickering stimuli, so-called second site adaptation effects, and much else besides. And color vision is not an isolated case: in general, I would maintain, phenomenology is best explained by neurobiology (Goldstein 1999). What determines the character of color experience is not the character of the distal stimulus, but the character of the neural processing which it produces in us.

Two important clarifications must be entered at this point. Typically, Internalism about qualitative character is formulated as follows:

It is *metaphysically necessary* that neurobiological duplicates (living under the same laws) have experiences with the same qualitative character (that is, given Relationism, are p-aware of the same Q-properties). In short, neural *sameness* necessarily makes for phenomenal *sameness*.

The Internalism I defend is both stronger and weaker than this. It is *stronger* because it doesn't just imply that neural sameness makes for phenomenal sameness: it also implies that neural differences – at least neural differences “above certain level of grain” – make for phenomenal differences. For instance, suppose that Maxwell's R-G and Y-B channels are in equal positive states of activation, while Twin Maxwell's R-G channel is in a positive state of activation in the absence of Y-B activity (i. e., his Y-B channel is in neutral balance). Then Maxwell is p-aware of orange while Twin Maxwell is p-aware of unitary red. Or again, there is good reason to believe that humans and pigeons are aware of a different range of colors just because of differences in their internal opponent color processing (Thompson et al. 1992). Of course, it is not the case, according to the Internalism I defend, that *all* “internal” differences make for phenomenal differences. That is obviously too strong. It is only neural differences at the level of the processing that determines qualitative character which make for phenomenal differences. Neural differences “below” this level are compatible with phenomenal sameness. For instance, Maxwell and Twin Maxwell might have the same color experiences even if there are very low-level neural differences between them – for instance, in the positions of neurons. In fact, Maxwell's opponent channel states might be realized in carbon-based neurons while Twin Maxwell's is realized in silicon-based neurons. So the Internalism I defend is also compatible with some measure of **multiple realizability**.

The Internalism I defend is *weaker* than the usual kind of Internalism formulated above in the following respect. Internalism as it is usually formulated is expressed in terms of **metaphysical necessity**. By contrast, the kind of Internalism I defend may hold with only **nomological necessity**. For the science is neutral on modal strength of the connection between the neural and the phenomenal. Its explanations would go through even if the connection were merely nomologically necessary, as Dualists maintain. To determine whether the connection is metaphysically necessary or merely nomologically necessary reflection of a more philosophical kind is required of the sort we are soon to be engaged in.

Before I proceed to the first stage of my argument, let me answer some questions that are likely to arise at this point.

Question. Who denies Internalism? Who denies that qualitative character is determined by what brain states we are in? Whenever you are in a certain brain state – even if the brain state is artificially induced and there is nothing in the outside world that is causing that brain state – you have an orangish experience. Who denies this?

Answer. There are some externalists about qualitative character who *do* deny Internalism in the sense I intend (Tye 1995, Dretske 1995, Lycan 1996). To illustrate the difference between me and the externalist,

let us focus on a specific case. Let the ORANGE brain state be the brain state which is responsible for the experience of orange. (It consists in equal positive R-G and Y-B activity.) When normal conditions obtain, this brain state is produced by objects with a certain kind of reflectance, or reflectance-type.⁵ Call it the ORANGE reflectance. The ORANGE reflectance is not always what produces the ORANGE brain state. Sometimes the ORANGE brain state is produced by other means, for instance by the RED reflectance under strange lighting conditions or by the use of a drug. But, under normal conditions, the ORANGE brain state is caused by the ORANGE reflectance. Let us express this by saying that the ORANGE brain state **tracks** the ORANGE reflectance.

Here now is the crux of the difference between the externalist and me. On the externalist view, the ORANGE brain state results in the experience of orange only because it tracks the ORANGE reflectance. There are differences of detail, but this what the externalists say. What determines the character of color experience is not the character of the brain state we are in, but the character of the reflectance-type which that brain state tracks. I deny this. To illustrate, imagine a world, *W*, where the ORANGE brain state does not track the ORANGE reflectance because (through a fluke of evolution) we evolved slightly different wiring (Pautz 2003c). On the externalist view, in *W*, the ORANGE brain state does not result in the experience of orange. By contrast, on my internalist view, in *W* the ORANGE brain state *does* result in the experience of orange, even though it does not track the ORANGE reflectance. The connection between our brain states and our color experiences is modally robust in a way that the externalist denies.

Why accept Internalism over Externalism? As I have mentioned, our color experiences vary along four chromatic dimensions: red, green, yellow and blue. Our opponent channel states also vary along four dimensions: positive/negative R-G activity and positive/negative Y-B activity. And there appears to be a good correlation between the four phenomenal parameters, on the one hand, and the four neural parameters, on the other. By contrast, reflectances do not vary along four recognizable dimensions. In fact, if you look in the extradermal world, the situation is extremely messy: there are no correlations at all (at least not that you can write down) between the character of our color experiences and the character of the reflectances which our underlying brain states track. So it is clear that it is our internal neurobiological states which determine the character of color experience.

Question. Isn't Internalist Relationism impossible? Colors are items external to our brains. The same goes for the other ostensible properties of external things – shapes, orientations, and so on. How then can our awareness of these properties depend on what brain states we are in?

Answer. There is no a priori reason to think that Internalist Relationism is impossible. It is true that what *objects* we are o-aware of cannot depend just on what brain states we are in. Just go to a world where we are in the relevant brain state but the relevant object doesn't exist. But Internalism doesn't say that what objects we are o-aware of is internally-determined. It says that what *properties* we are p-aware of is internally-determined. And it is plausible to say that properties – like being red or being square – are abstract objects that exist whether or not they are instantiated in the external environment. So the kind of worry gestured at doesn't arise. There is no world where we are in the right brain states but the relevant properties do not exist. (Compare: what numbers objects bear the mass-in-grams relation to supervenes on their intrinsic mass-properties. Yet numbers are “external to” objects.)

Question. Internalist Relationism entails that the same object could (within a world or across worlds) look to have different colors to different creatures under optimal conditions due to evolved differences in their opponent processing. For instance, a square might look orange to a human but unitary red to a monkey, even though both are looking at it under optimal conditions and their visual systems are working

⁵ The *reflectance* of an object is its disposition to reflect various proportions of light at various wavelengths. Many different reflectances produce the same color experience in us. When they do, they are called *metamers*. A *reflectance-type* is simply a disjunction of metameric reflectances.

as they were designed to work. (Indeed, we might even suppose that their visual system is tracking the very same reflectance property of the square.) In such a case who is right?

Answer. The view that one is right and the other is wrong seems out of the question. So we must say either that both are right or that both are wrong.

But I do not see this as an objection against Internalist Relationism. There are at least a couple of good options here. (i) There is at least one well-developed view on the nature of colors which delivers the verdict that both are right. This is the non-reductive Dispositionalist view of Colin McGinn (1996). On this view, colors are simple properties that supervene on dispositions to produce experience of those very colors. (ii) And I do not think that the view that both the human and the monkey are wrong – a kind of Error Theory of color – is entirely implausible either. In fact, I think that it is the right view. On this view, creatures have evolved to see objects as having colors to help them negotiate the world, but nothing really has the colors they appear to have (I develop this approach in Pautz 2003e).

§4. The First Stage: From Relationism and Internalism to the Failure of Identity Physicalism

The purpose of the first stage of my argument for skepticism about Physicalism is to show that these two premises – Relationism and Internalism – rule out Identity Physicalism in the case of qualitative mental properties.

On Relationism, qualitative mental properties are composite, relational properties of the form *standing in the p-awareness relation to Q-property P*. Here I will set to the side the issue of whether the Q-properties – which include colors, smells, sounds, and so on – can be reduced (I address this issue in Pautz 2003d, 2003e). I will give to the Identity Physicalist whatever view on the Q-properties she wants. Instead, I will focus on the p-awareness relation. Since I will be focusing exclusively on the awareness of properties (which determines qualitative character) as opposed to the awareness of objects and facts (which does not), I will drop the distracting prefix.

I will argue from Relationism and Internalism that the awareness relation cannot be identified with any physical relation, where ‘physical relation’ is to be understood broadly to include topic neutral relations (causal relations, dependence relations) and functional relations.

The structure of the argument is as follows. By Relationism, the awareness relation, *individual x is aware of Q-property y*, is a two-place relation between individuals⁶ and Q-properties: colors, smells, tastes, and so on. By Internalism, what we bear this relation to is determined by what brain states we are in. So by Leibniz’s Law if the awareness relation is identical with some physical relation R, then R must also have these two properties. It must be a two-place relation between individuals and Q-properties. And what we bear it to must be determined by what brain states we are in. For short: it must satisfy the **Relationality Constraint** and the **Dependence Constraint**. But I will argue that no physical relation satisfies both constraints at once. No physical relation is both a relation between individuals and Q-properties, and yet internally-determined in the way that the awareness relation is. It follows that the awareness relation is not a physical relation.

To show this, I will divide all physical relations into two categories, *wide* and *narrow*. Say that physical relation R is **narrow** iff neurobiological duplicates living under the same physical laws necessarily bear R to the very same things. Say that a physical relation R is **wide** iff it is not narrow. Wide physical relations do not satisfy the Dependence Constraint. Narrow physical relations do not satisfy the Relationality Constraint or may be ruled out for other reasons.

⁶ I use ‘individuals’ broadly to include non-human sentient creatures.

★ ★ ★

Our most plausible accounts of intentionality are wide. What makes COW-thoughts *about* cows? Many Physicalists would answer (very roughly) along the following lines: the fact COW-thoughts are causally correlated with the presence of cows in the environment under optimal conditions. Many Relationists favor a similar account of sensory intentionality (Dretske 1995, Tye 1995, Lycan 1996). They in effect claim that the awareness relation can be identified with a *wide* physical relation. So let's consider wide physical relations first.

The basic idea is as follows. First, it is claimed that the Q-properties – including the colors and other secondary qualities – we are aware of are identical with the physical properties that our brain states track. For instance, colors are reflectance properties, sounds are sound-waves, and so on. (Recall that a brain state B **tracks** a property P iff, if optimal conditions were to obtain, things with P would cause the tokening of B.) Call this **Physicalism** about the Q-properties. Second, it is claimed that what it is to be *aware of* – what it is to “sensorily represent” – a certain Q-property P (that is, on this view, a certain physical property) is just to be in a brain state that plays the “sensory consciousness role” and that tracks the instantiation of P in the environment or, in the case of bodily experience, one's own body.⁷

What is the “sensory consciousness role”? Not all states that track external properties – or that “represent” external properties – realize states of awareness of those properties. (Consider the states of the retina.) The “sensory consciousness role” is the magical ingredient which is supposed to turn *mere* representation into *conscious* representation. It is supposed to turn mere tracking into the kind of tracking that makes for awareness of what is tracked. Typically defenders of a wide reductive view of the awareness relation claim that it can be spelled out in terms of a state's being poised to influence the formation of beliefs and desire or something along these lines (Tye 1995, Dretske 1995). But as this component of the view will not play a role in my argument, I will ignore it in what follows.

So, in effect, the proposal is that the awareness relation, **x** is aware of Q-property **y**, is identical with:

The Tracking Relation: **x** is in a brain state B that plays the “sensory consciousness-role” (for short, a *C-brain state*) and that tracks the instantiation of property **y**

Of course, there are different ways of carrying out the details. Instead of tracking, one might pin one's hopes on biological function (Dretske 1995), or asymmetric dependence (Fodor 1990), or whatever. The objection I will raise against wide physical relations here does not depend on the details (Pautz 2003c).

How does the Tracking Relation fare on our two constraints?

The Tracking Relation is a relation between individuals and the physical properties their brain states track. So if Physicalism about the Q-properties is correct, so that the Q-properties *just are* these physical properties (colors are reflectances, sounds are sound-waves, and so on), then the Tracking Relation, like the awareness relation, is a relation between people and Q-properties: it satisfies the Relativity Constraint.

⁷ If correct, the Tracking account of the awareness relation provides a nice explanation of the mind's somewhat mysterious ability to be *en rapport* with a property even if nothing in one's environment has that property. When one hallucinates a pink elephant (for instance) one bears the Tracking Relation to the color pink: one is in a brain state that plays the “sensory consciousness role” and that, if optimal conditions *were* to obtain (if one's nervous system were working right, and so on), *would be* caused by the color pink (which on this view = a physical property of the surfaces of objects). On the Tracking account, the awareness relation *just is* this relation. So one is aware of the color pink, despite the fact that there is nothing pink in one's environment.

I happen to think that Physicalism is false (Pautz 2003d, 2003e). The Q-properties we are aware of are distinct from the physical properties our brain states track. If this is right, then the Tracking Relation is *not* a relation between people and Q-properties: it does not satisfy the Relationality Constraint. Since the awareness relation is a relation between people and Q-properties, it would follow that the awareness relation cannot be identified with the Tracking Relation. But I wish to set this objection aside here. As I said, for the sake of argument, I will give to the Identity Physicalist whatever view of the Q-properties he wants.

The real problem with the Tracking Relation is that it doesn't satisfy the Dependence Constraint. So even aside from problems with Physicalism about the Q-properties it can be ruled out immediately as a candidate to be the awareness relation. What Q-properties we are aware of is internally-determined. By contrast, what properties our brain states track is externally-determined. So the awareness relation cannot be identified with the Tracking Relation.

Case 1: Maxwell and Twin Maxwell. In the actual world, Maxwell views a square with the ORANGE reflectance under optimal conditions. As a result, he is put into the ORANGE brain state (equal positive R-G and Y-B activity). In general, in the actual world, the ORANGE brain state tracks the ORANGE reflectance. In a nearby counterfactual situation, *W*, Twin Maxwell also views a square with the ORANGE reflectance under optimal conditions. But in *W* humans evolved somewhat different postreceptoral wiring than that which they evolved in the actual world. As a result, when his visual system is working as it was designed to work and conditions are otherwise normal, the presence of an object with the ORANGE reflectance puts Twin Maxwell into the UNITARY RED brain state (positive R-G activity in the absence of Y-B activity) as opposed to the ORANGE brain state. In general, in *W*, it is the UNITARY red brain state that tracks the ORANGE reflectance.

By Internalism, when they view the square, Maxwell and Twin Maxwell bear the awareness relation to different colors. But they bear the Tracking Relation to the very same property (viz. the ORANGE reflectance). So the awareness relation cannot be identified with the Tracking Relation. (In Pautz 2003c, I develop this case in detail.)

Case 2: Pigeons. Pigeons appear to possess *three* or maybe even *four* opponent channels. We have good reason, both on the strength of the opponent process theory for humans and on the basis of psychophysical investigation, to think that they are consequently aware of colors that do not belong to our color-space: *alien colors* (Thompson et al 1992). As a matter of fact, their receptor systems are different as well. So they also bear the Tracking Relation to different physical properties. But we can imagine a nearby counterfactual situation where this is not so. In such a situation, by Internalism, they are still aware of different colors, yet they bear the Tracking Relation to the same external properties that we do. So the awareness relation is not identical with the Tracking Relation.

Case 3: The Brain in the Vat. Arguably, given the biological basis of color vision, if a brain in a vat undergoes the right internal color processing, it is aware of certain colors, even though its color experiences are hallucinatory and nothing in its environment actually has those colors. But its brain states don't track anything.⁸ Again, the awareness relation cannot be identified with the Tracking Relation.

The same goes for any wide physical relation: biological function, asymmetric dependence, or whatever. These relations don't satisfy the Dependence Constraint. What properties we bear the awareness to is internally-determined. What properties we bear wide physical relations to is externally-determined. So the

⁸ It might be replied that the brain states do track certain properties. For tracking is defined in counterfactual terms: a brain state tracks what *would* cause it, if optimal conditions *were* to obtain. Now suppose that the brain in a vat has sense organs or receptor systems (e. g. eye balls). Then there are certain properties that *would* produce its brain states, if conditions *were* optimal – that is, if it were not a brain in a vat, but a brain in a skull, in a normal environment. These are just the properties that would stimulate its receptor systems. But to get around this response, we can simply stipulate that the brain in the vat is entirely without receptor systems.

awareness relation cannot be a wide physical relation. Our most plausible approaches to reductively explaining intentional relations in general can be ruled out immediately in the case of the awareness relation because they run afoul of the Dependence Constraint.

* * *

Given the Dependence Constraint, the only hope for the Identity Physicalist seems to lie with narrow physical relations. But such relations either fail to satisfy the Relationality Constraint or can be immediately ruled out for other reasons. Once wide physical relations are ruled out, there simply remain no physical relations that are suitable candidates to be the awareness relation.

Recall that R is a narrow relation iff, necessarily, internal neurobiological duplicates bear R to the same things, regardless of how things stand in the outside world. When we try to come up with examples of narrow relations we think of relations to things in our own skulls. Consider, for instance,

The Brain State Relation: x is in brain state y

The Brain State Relation is a relation between people and their own brain states. A person bears it to a brain state just in case they are in that brain state. It satisfies the Dependence Constraint: for trivially, what we bear it to – what brain states we bear it to – is determined by what brain states we are in. But it does not satisfy the Relationality Constraint. The awareness relation is a relation between individuals and Q-properties. The Brain State is a relation between individuals and their own brain states. Clearly, the awareness relation cannot be the Brain State relation.

So while wide physical relations can be ruled out as candidates to be the awareness relation on the grounds that they do not satisfy the Dependence Constraint, it seems that narrow physical relations can be immediately ruled out on the grounds that they do not satisfy the Relationality Constraint. How could there be a narrow physical relation that satisfies the Dependence Constraint, and yet reaches outside the skull and so satisfies the Relationality Constraint?⁹

It may seem that a solution is not far to seek: just take a wide relation and “rigidify” it. For instance, consider:

The Rigidified Tracking Relation: x is in a C-brain state that *actually* (in the actual world) tracks y

If Physicalism about the Q-properties is correct, so that the Q-properties we are aware of are among the physical properties our brain states track, then the Rigidified Tracking Relation, like the Tracking Relation, satisfies the Relationality Constraint: it is a relation between people and Q-properties. As I mentioned before, I happen to think that Physicalism about the Q-properties is false, so that the Rigidified Tracking Relation does not satisfy the Relationality Constraint. But again, I will set this objection aside here.

The Rigidified Tracking Relation satisfies the Dependence Constraint. To see this, take a certain opponent-channel state, say the UNITARY RED brain state. In the actual world, the UNITARY RED brain state tracks the UNITARY RED reflectance (i. e., the reflectance-type of objects that normally look

⁹ There is an artificial trick whereby one might obtain such a relation. Let B1 be the brain state which results in the awareness of P1, let B2 be the brain state which results in the awareness of P2, let B3 be the brain state which results in the awareness of P3, and so on. Now take the disjunctive relation: x is in B1 and y = P1 or x is in B2 and y = P2 or x is in B3 and y = P3 or . . . x is in Bn and y = Pn. But this relation is far too unnatural and disjunctive. It is impossible to see how it could be the relation we are talking about when we say that some one is aware of the color red, for instance.

unitary red to us). So, no matter what world you go to, the UNITARY RED brain state tracks the UNITARY RED reflectance in the actual world. So, no matter what world you are in, if you are in the UNITARY RED brain state, you bear the Rigidified Tracking Relation to the UNITARY RED reflectance (and so, given Physicalism about Q-properties, the color *unitary red*). In general, the connection between being in a sensory state and bearing the Rigidified Tracking Relation to the property it actually tracks is modally robust: indeed, metaphysically necessary. In consequence, the Rigidified Tracking Relation satisfies the Dependence Constraint.

So it seems that (at least given Physicalism about the Q-properties) the Rigidified Tracking Relation satisfies both the Relationality Constraint and the Dependence Constraint. Nevertheless, the proposal that the awareness relation is identical with the Rigidified Tracking Relation is a non-starter, for three reasons.

First, the view that the awareness relation is the Rigidified Tracking Relation involves an objectionable kind of actual-world chauvinism. Consider the case of Maxwell and Twin Maxwell. Maxwell (in the actual world, @) and Twin Maxwell (in *W*) look at a square with the ORANGE reflectance under optimal conditions. As a result, Maxwell is put into the ORANGE brain state, while Twin Maxwell is put into the UNITARY RED brain state. On the Rigidified Tracking account of the awareness relation, Maxwell is aware of the ORANGE reflectance (that is, on Physicalism about the Q-properties, the color orange). For the ORANGE brain state tracks the ORANGE reflectance in the actual world. By contrast, on the Rigidified Tracking account, Twin Maxwell is aware of the UNITARY RED reflectance (that is, on Physicalism about the Q-properties, the color unitary red). For, on this view, to determine what property Twin Maxwell is aware of in *W*, we must look to what property his brain state tracks in the *actual* world. And, in the actual world, the UNITARY RED brain state tracks the UNITARY RED reflectance. But the square before Maxwell does not have the UNITARY RED reflectance, but the ORANGE reflectance. So the Rigidified Tracking account implies that Twin Maxwell's color experience is non-veridical, even though he is viewing the square under optimal viewing conditions and his visual system is working properly. In general, our counterparts in *W* are subject to a pervasive illusion with respect to the colors of things. The same goes for our counterparts in any world where, as in *W*, the actual connection between reflectances and opponent channels states is broken. We in the actual world, on the other hand, generally see things in their true colors.

But why suppose that *we* are the lucky ones? Why not, so to speak, pin down the awareness relation to *W* rather than to the actual world? That is, why not say that the awareness relation is the relation: **x** is in a C-brain state that is causally correlated with **y** in *W*. Then Twin Maxwell, and more generally our counterparts in *W*, are the ones who get it right, and we are the ones who are subject to the pervasive and systematic illusion, our opponent channel states representing not the colors they track in the actual world, but the different colors they track in *W*. I can't imagine any good reason to favor us over our counterparts in *W*.¹⁰ It seems arbitrary to pin down the awareness relation to @ rather than *W*. Call this the **Argument from Arbitrariness** against the Rigidified Tracking account of the awareness relation.

My second objection to the identification of the awareness relation with the Rigidified Tracking Relation is as follows. Let us say that a brain state B is **alien** iff it does not track any property in the actual world. It seems possible that there is a creature which has only alien brain states, yet which is aware of properties. But on the present proposal there could be no such creature, for the brain states of such a creature do not track any properties in the *actual* world. Thus, the Rigidified Tracking account is actual-

¹⁰ My objection here is not simply that according to the rigidification proposal the inhabitants of *w* are subject to a pervasive color illusion. I am an Eliminativist about color, so I think that in our own world, and any world like our own, color experiences are illusory (Pautz 2003e). I am therefore happy with color illusion. My objection is rather to the asymmetry. Why think that *our* color experiences are veridical and those of the *inhabitants in W* are illusory, as the Rigidified Tracking account implies?

world chauvinistic in yet another way. It implausibly restricts the range of creatures that can enjoy experiences to the species of the actual world. There are giraffes, pigeons, humans, bats, and so on; but there couldn't have been any other sorts of sentient creatures. I take this to be a *reductio ad absurdum* of this account. Call this the **Problem of Alien Species**.

My third objection is as follows. It seems that people in other possible worlds, such as W , could believe exactly what we do, when we believe that people are aware of colors. But the Rigidified Tracking proposal, it seems, cannot accommodate this datum. On this view, what we believe, when we believe that people are aware of colors, is a proposition that is true (with respect to any world w) iff people are in brain states that track colors in $@$. But it is implausible that those in W believe a proposition that is true iff people are in brain states that track colors in $@$. For them, $@$ is another possible world. They have no perceptual or epistemic contact with $@$. So it seems that they do not, and cannot, have beliefs involving $@$. Thus, the Rigidified Tracking Proposal seems to imply that people in other worlds cannot believe what we do, when we believe that people are aware of colors. Call this the **Semantic Objection**.¹¹

In view of these problems, the proposal that the awareness relation is identical with the Rigidified Tracking Relation cannot be sustained.

The only other proposal that I can think of concerning what narrow physical relation the awareness relation might be which has any promise at all is based on David Lewis's theory of content. The proposal is that the sensory awareness relation is the following relation:

The Narrow Interpretation Relation: 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.

This needs to be unpacked. What are "Narrow Functional Roles"? They are second-order properties of brain states to do with their interactions with other inner states and their interactions with behavior (narrowly and non-intentionally construed). They are "narrow" in that they supervene on neurobiology together with the laws of physics. Necessarily, if two individuals have the same neurobiology (and live under the same physical laws), then they will have the same behavioral dispositions (non-intentionally and narrowly construed).

What are the "Narrow Constraints on Interpretation"? They are principles taken from our common sense theory of persons about how mental states interact with one another (Internal Constraints), and how they interact with behavior (Output Constraints). They are "narrow" in that they don't include principles about how the *environment* affects our mental states.

What is the "Best Narrow Interpretation"? It is the assignment of mental states – beliefs, desires, and experiences – to the brain states of members of K which best satisfies the Narrow Constraints on Interpretation, given the Narrow Functional Roles of those states.

The basic idea is that our mental states, including our sensory states, have contents that are independent of how things stand in the outside world. We can determine, in a rough and ready way, what those contents

¹¹ This objection to the Rigidified Tracking account is similar to an objection Scott Soames (1998, pp. 15-17) raises against the Rigidified Description analysis of proper names.

It might be replied that the defender of the Rigidified Tracking account can accept that those in W do not believe what we do, when we believe that people are aware of colors. When they say 'people are aware of colors', they express belief in a proposition that is true iff people are in brain states that track colors in W . But this response has several absurd results: for instance, that we can truly say 'people in W are aware of colors, but do not believe that they are aware of colors'.

are, given only the narrow functional roles of a person's brain state. For instance, if brain states D and E cause someone to reach out and attempt to grasp something one foot in front of him, then we might interpret E as the experience as of something two feet in front of one, and D as the desire to grasp something two feet in front of one. The principles that we use in determining these "narrow contents" are simply the Narrow Constraints on Interpretation. Given the narrow functional roles of a person's inner states, we can (theoretically anyhow) use these constraints to determine a Best Narrow Interpretation.

The Tracking account of the awareness relation is **input-based**. By contrast, the Narrow Interpretation account is **output-based**, because it accords outputs (behavioral dispositions) the principal role in pinning down sensory content.

It might be argued that the Narrow Interpretation Relation satisfies both of our constraints. Take the Relationality Constraint. If the Best Narrow Interpretation assigns to our brain states experiences of Q-properties, as seems likely, then the Narrow Interpretation Relation satisfies the Relationality Constraint: it is a relation that holds between individuals and Q-properties.

Recall that tracking relations (and other wide physical relations) satisfy the Relationality Constraint only if the Q-properties are among the physical properties that our brain states track – in other words, only if Physicalism about the Q-properties is correct. It is worth noting that this is not true of the Narrow Interpretation Relation. Suppose that Q-properties are simple, non-physical properties which external objects appear to have but do not (Pautz 2003e). Then our brain states do not track Q-properties. It might nevertheless be that the Narrow Interpretation Relation satisfies the Relationality Constraint. Consider the case of colors. Even though (on this view) colors are not instantiated in our environment, it might nevertheless be that the assignment of experiences of **colors** to our brain states makes best overall sense of our behavior. Thus, on this view, we can be aware of colors, even if they are not properties of objects in our environment. Compare numbers. On some views, numbers are *sui generis* entities that play no causal role in the generation of our number thoughts. But, on an Interpretation account of content, it might nevertheless be that we have thoughts about numbers, because it might be that the Best Narrow Interpretation assigns thoughts about numbers to some of our brain states.

It might also be argued that the Narrow Interpretation Relation satisfies the Dependence Constraint. For instance, it belongs to the narrow roles of our opponent channel states that they are apt to cause certain discriminatory behavior and perhaps certain affective responses (for instance, the opponent channel state underlying the experience of red might cause excitement). Take Maxwell and Twin Maxwell (or a human and a pigeon, or whatever). Because they are in different opponent channel states, they are in states with subtly different narrow functional roles. For instance, their discriminatory behavior might differ in certain subtle ways. So – it might be thought – the Best Narrow Interpretations will assign to their different opponent channel states experiences of different colors. In particular, it will assign to Maxwell's ORANGE brain state the experience of orange, and it will assign to Twin Maxwell's UNITARY RED brain state the experience of unitary red. Hence they will bear the Narrow Interpretation Relation to different colors. And, in this way, the Narrow Interpretation account of content will deliver the right result, namely, that they are aware of different colors in the same objective circumstances.

Input-based relations don't satisfy the Dependence Constraint because inputs are not internally-determined. Our brain states and their optimal distal causes can come apart, as in the Maxwell-Twin Maxwell case. The suggestion here is that output-based relations, such as the Narrow Interpretation relation, satisfy the Dependence Constraint because outputs are internally-determined.

Yet the proposal that the awareness relation is identical with the Narrow Interpretation Relation, like the proposal that it is identical with the Rigidified Tracking Relation, is a complete non-starter.

First, the content of experience is extremely rich – much richer than that of belief and desire. For instance, I am right now aware of thousands of different shades of colors and shapes and so on. So, as regards

experience, the Unique Best Interpretation will have to be correspondingly rich. I think that it is clear that the Narrow Functional Roles plus the Narrow Constraints on Interpretation are not enough deliver the goods.

Consider the Narrow Roles first. Some brain states have quite rich narrow functional roles. In particular, the brain states underlying **primary quality perception** have fairly rich narrow functional roles. For instance, the brain states underlying shape perception control our behavior (narrowly construed) in an extremely detailed way. By contrast, the brain states underlying **secondary quality perception** have very meager narrow functional roles. Consider for instance the brain states underlying color vision. They don't much affect how we interact with the world. They only affect how we are disposed to sort things and help us in re-identifying objects. (This is not surprising, because the properties they track – surface spectral reflectances – don't much affect how things interact with us and with other things.)

Now consider the Narrow Constraints on Interpretation. The Narrow Constraints on Interpretation are supposed to come from our common sense theory of persons. But then they are few and far between and very rough and ready. For instance, we might think that if A has an experience of two similar colors in different regions of his visual field, and has the desire to sort objects by color, then *ceteris paribus* A will attempt to sort the objects in those regions together. Or we might think that if A desires p and believes that p only if q, then *ceteris paribus* A desires q. But I don't think that they ever get much more detailed than this. As it is sometimes put, you cannot codify human rationality.

The problem is now clear. The Narrow Functional Roles plus the Narrow Constraints on Interpretation are far too meager to determine a Unique Best Interpretation. To illustrate, suppose that you are given all the Narrow Functional Roles of a person's inner states and the Narrow Constraints on Interpretation. Suppose, for instance, you are given how Twin Maxwell's opponent channel states shape his sorting behaviors plus a few common sense beliefs about how experiences of colors influences how people behave. You are just given his neurobiology and his trajectory through space and time. But you are given no information about his external environment. Could you – or for that matter, God – ascertain with these meager resources exactly what shade of color and shape and so on Twin Maxwell experiences at every region in their visual field? No. *There is no backward road from behavior to content* – at least, not in the case of sensory content. So there is no such thing as the Unique Best Narrow Interpretation. The Narrow Roles plus the Narrow Constraints on Interpretation just aren't enough determine one. Call this the **Underdetermination Problem**.

Now x bears the Narrow Interpretation Relation to 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 . But we have just seen that there is no such thing as the Best Narrow Interpretation. It follows that no one ever bears the Narrow Interpretation Relation to anything. The relation has no application. The Narrow Interpretation Relation trivially fails to satisfy the Relationality Constraint. It follows that the awareness relation cannot be identified with the Narrow Interpretation Relation.

I think that this objection is decisive, but it is worth mentioning a second objection. Imagine a creature C which is aware of a range of Q -properties P which we are not aware of. In fact, to make the case even more extreme, imagine that no creatures in the actual world are aware of the Q -properties in P . Suppose, moreover, that C 's behavior, consequent upon its experiencing a Q -property in P , is highly peculiar. It doesn't wince, or sort things together, or whatever. It engages in behaviors which we have never confronted in any actual creature. For instance, maybe it turns purple and expands. Surely such a creature is possible.

Now *ex hypothesi* C is aware of the alien Q -properties in P . So if the Narrow Interpretation account is correct, and C the awareness relation just is the Narrow Interpretation Relation, then C must bear the Narrow Interpretation Relation to the alien Q -properties in P . But this requires that the very strange Narrow Roles of C 's brain states plus the Narrow Constraint on Interpretation taken from *our* common sense

theory of persons determine a Unique Best Narrow Interpretation which assigns to C's "brain" states experiences of the alien Q-properties in P. The problem is clear. The Narrow Constraints on Interpretation taken from *our* theory of persons do not include principles connecting experiences of the alien Q-properties in Q with the very unusual behaviors in which C is apt to engage. So they cannot yield such a Unique Best Narrow Interpretation of C.

So, even if we set aside the Underdetermination Problem, the Narrow Interpretation account must be rejected, because it is unable to accommodate the result that alien creatures can be aware of alien Q-properties. Like the Rigidified Tracking proposal, it is far too chauvinistic. Call this the **Alien Species Problem for the Narrow Interpretation account**.¹²

To solve this problem, one might be tempted to simply add to the "Narrow Constraints on Interpretation" constraints besides those in our common sense theory of persons in order to accommodate (all actual and possible!) non-human creatures. But what constraints should be added? There are many to choose from. For instance, take our creature C and his expansion-behavior. One constraint is that the experience of red leads the alien creature to expand, another is that the experience of pain does, another is that the experience of some (to us) alien Q-property does, and so on. All of these different constraints yield different Best Narrow Interpretations. Which ones should be used? What are the criteria for deciding what the "right" constraints are, and hence what the Best Narrow Interpretation is, for an alien creature? We might just say: they are the ones that assign to the states of the alien creature *the experiences that it actually has*. But this procedure is circular. We wanted to know: how do the (narrow) physical facts determine what Q-properties the creature is aware of. This response appeals to facts about what Q-properties the creature is aware of in its answer.

I conclude that the awareness relation cannot be identified with a narrow physical relation. Narrow physical relations either fail to satisfy the Relationality Constraint (the Brain State Relation, the Narrow Interpretation Relation), or can be decisively ruled out for other reasons (the Rigidified Tracking Relation).

* * *

The awareness relation cannot be identified with any *wide* physical relation: wide physical relations fail to satisfy the Dependence Constraint. And it cannot be identified with any *narrow* physical relation: narrow physical relations either fail to satisfy the Relationality Constraint or can be ruled out for other reasons. It follows that the awareness relation is irreducible.

It might be objected that I have not exhausted all of the options. But I think that this kind of "wait and see" response is not plausible in this case.

Sometimes such a response *is* appropriate. Take belief. There are several "externalist" or wide physicalist accounts of belief on offer – those due to Stalnaker, Fodor, Dretske, Millikan, and so on. There are problems with all of the proposals. So one might argue – and people have argued – that there is no correct physicalist reduction of belief. Here the objection that we haven't hit on the right reductive account seems entirely appropriate. The accounts on offer seem to be on the right track. Mostly the problems are just problems of detail – the disjunction problem, or whatever. Maybe there is some physicalist reduction of belief, but it is too complicated for us to hit upon.

But the situation with the awareness relation is totally different. Given the Relationality Constraint and the Dependence Constraint, there is no proposal out there which is even half-way plausible. All the proposals one can think of are just clear non-starters, given the Relationality Constraint and the Dependence

¹² Another problem is worth mentioning. According to the present proposal, the causal roles of our experiences with respect to other inner states and behavior determine their contents. But, intuitively, it is the other way around: the contents of our experiences determine their causal roles. There is some tension between these two ideas.

Constraint. The “tracking” proposals favored Tye and Dretske and Lycan – though otherwise very nice – can be immediately ruled out because they don’t satisfy the Dependence Constraint. And all the narrow physical relations that we can think of either fail to satisfy the Relationality Constraint or are absolute non-starters. In the absence of anything even *approaching* an initially plausible reductive view of the awareness relation, it is reasonable to conclude that there is no good reductive account.

So far, I have focused on the awareness relation, and I have argued that, given our twin constraints, it is irreducible. I have set the issue of the reducibility of the Q-properties – and in particular, the secondary qualities – to the side. But it is worth noting that the view of experience developed here goes naturally with a non-reductive view of the secondary qualities also. Consider the case of color. There is very good reason to think that colors are not literally identical with dispositions to produce color experiences (Boghossian and Velleman 1989, McGinn 1996). For one thing, on the present view of experience anyway, this reduction can be immediately ruled out by the fact that color experiences themselves are relations to colors. The only other viable reductionist view of color is reductive Physicalism, according to which colors are observer-independent physical properties of objects, for instance, reflectance properties. But, on the Internalist form of Relationism I have argued for here, the identification of a color with any particular reflectance property is entirely unmotivated. For the very same reflectance property can (within a world or across worlds) produce under optimal conditions experiences of quite different colors in different creatures due to differences in neurobiology. Think of Maxwell and Twin Maxwell. So there can be no warrant for identifying any given SSR-property with one color rather than another. That would be completely arbitrary. We might call this the **argument from arbitrariness** against color physicalism (I develop this argument in Pautz 2003d).¹³ For this reason, Internalist Relationism goes most naturally with a non-reductive or **Primitivist** view of colors and the other secondary qualities.¹⁴

§6. The Second Stage: From the Failure of Identity Physicalism to Skepticism about Physicalism

In any case, I have argued that the *awareness relation* is in a very strong sense irreducible. It is not identical with any physical, topic-neutral, or functional relation. So the state of standing in the awareness relation to certain property is a state of a person over and above her physical and functional properties, both narrow and wide. This shows that **Identity Physicalism** fails in the case of qualitative mental properties.

Sometimes philosophers accept that some property or relation of the manifest world cannot be identified with any physical or functional property or relation, yet nevertheless maintain that its pattern of instantiation supervenes with metaphysical necessity on the pattern of instantiation of physical and functional properties. In other words, they still accept **Physicalist Supervenience**.

For instance, Colin McGinn (1991) and Thomas Nagel (2001) hold that Identity fails in the case of qualitative mental properties (= they hold that qualitative mental properties are not identical with physical properties or functional properties realized by physical properties), yet uphold Physicalist Supervenience for such properties. (They also claim that it is beyond our powers to understand the supervenience.) Likewise, McGinn (1996), Yablo (1995) and Campbell (1993) hold that colors are simple, irreducible, non-physical properties. Yet they hold that they are supervenient on physical properties. Paul Horwich (1998) holds that the *meaning* relation is irreducible, yet he says that what a word means supervenes on use properties. Similarly, Stephen Schiffer (2003) holds that the *belief* relation is irreducible, but he says that what

¹³ Another argument is the argument from color structure.

¹⁴ This doesn’t by itself entail an error theory. For instance, one might claim that colors supervene on certain properties of objects (e. g. their dispositions to produce experiences of those colors in us), without being identical with those properties. See McGinn 1996. But I myself favor an error theory (Pautz 2003e).

propositions we believe supervene on our functional properties. And Moore seems to have held such a view of *goodness*.

These are forms of what in the Introduction I called **Dualistic Physicalism**. It might be wondered why a similar view cannot be taken of the awareness relation. In other words, why can't the Physicalist simply respond to the first stage of the argument as follows? –

True, the awareness relation is a primitive, irreducible relation. But I can still uphold Physicalist Supervenience (=Physicalism). I will just say that it is metaphysically necessary that if someone is in a certain internal brain state (or computational state, or narrow functional state), then *viola!* he or she bears the primitive awareness (or sensory representation) relation to a certain property *P*. True, on this view, the state of being aware of *P* is a relation to *P*, but what it supervenes on is not a relation to *P*: it supervenes on a brain state. But who ever said that properties/relations must supervene on properties/relations with the same “structure”?

Briefly, the problem with the retreat to Dualistic Physicalism is as follows (I say more in Pautz 2003a). Once we give up Identity, we have (at least) two choices. One is to accept Dualistic Physicalism. This view claims that people are physical things, but it claims that properties of people of the form *being aware of P* are not identical with any of their physical or functional properties. Yet it is *metaphysically* necessary that if you have a certain neurobiological property (or whatever), then you have one of these properties.¹⁵

But, of course, there is another option available. It agrees with the first option that people are physical things, and that properties of people of the form *being aware of P* are not identical with any of their physical or functional properties. But it holds that it is only *nomologically* necessary that if you have a certain neurobiological property (or whatever), then you have one of these properties. This, of course, is just traditional **Property Dualism**.¹⁶

Ontologically, Dualistic Physicalism and Property Dualism are identical. They both uphold a kind of Dualism about awareness properties. They differ only modally: whereas Property Dualism holds that the irreducible awareness properties supervene only as a matter of nomological necessity, Dualistic Physicalism holds that they supervene as a matter of metaphysical necessity. Thus, where Dualism requires “nomological danglers” (a swarm of fundamental psychophysical laws), Dualistic Physicalism requires “modal danglers” (a swarm of unexplained metaphysically necessary connections between wholly distinct properties). So they both require brute psychophysical laws. It is just that Property Dualism holds that they are nomologically necessary, while Dualistic Physicalism holds that they are metaphysically necessary.

Given that they are so similar, why should we accept Dualistic Physicalism over Property Dualism? Once we give up Identity, why should we accept metaphysically necessary supervenience instead of merely nomologically necessary supervenience? What advantages accrue?

As far as I can see, there is no reason. Dualistic Physicalism is simply *too close* to Dualism to have any advantages over Dualism. The traditional arguments for accepting Physicalism over Dualism do not carry over to Dualistic Physicalism.

¹⁵ You might say that this view can be ruled out immediately on the grounds it violates the Humean dictum that there cannot be strictly (or metaphysically) necessary connections between wholly distinct existences. For this view holds that it is metaphysically necessary that if you have the property of being in a certain brain state, then you have a further, wholly distinct property of standing in the irreducible awareness relation to a certain Q-property. I am sympathetic (Pautz 2003a), but I will not press the point here.

¹⁶ There is yet another option: Substance Dualism. But I will ignore Substance Dualism in what follows.

Induction. One traditional argument for Physicalism is this (Smart 1959). All macro properties besides qualitative mental properties are identical to physical or functional properties realized by physical properties. That is, Identity Physicalism is true of all macro properties besides qualitative mental properties. So it is likely that it is true of qualitative mental properties. As Smart writes, “There does seem to be, as far as science is concerned, *nothing but* increasingly complex arrangements of physical constituents. That everything should be explicable in terms of physics except the occurrence of sensations seems to me to be frankly unbelievable” (p. 142).

But this is only an argument for accepting *Identity Physicalism* over Dualism. There is no analogous “inductive” argument for accepting Dualistic Physicalism over Dualism. Such an argument would have to go like this. All macro properties besides qualitative mental properties are irreducible to physical/functional properties, but nevertheless supervenient on such properties with metaphysical necessity. That is, Dualistic Physicalism is true of all macro properties besides qualitative mental properties (or at least a great many of them). So it is likely that it is true of qualitative mental properties too. The first premise of this argument is false. It is not the case that Dualistic Physicalism is true of all macro properties besides qualitative mental properties (or even a great many of them). In general, macro properties seem to reduce to physical properties or, if not physical properties, then functional properties realized by physical properties.

Economy (Occam’s Razor). A second traditional argument for Physicalism is this (Smart 1959). Identity Physicalism is more economic than Property Dualism in its stock of basic principles. Property Dualism requires fundamental psychophysical laws – “nomological danglers” – connecting a person’s qualitative mental properties with her physical and functional properties. Such laws would be nomological *danglers*, in the sense that they would dangle off of the rest of the body of fundamental laws (presumably the laws of physics). By contrast, Identity Physicalism avoids fundamental psychophysical laws by simply *identifying* a person’s qualitative mental properties with a subset of her physical or functional properties.

But, again, this is only an argument for preferring Identity Physicalism to Property Dualism. There is no analogous argument for preferring Dualistic Physicalism to Property Dualism. Where Property Dualism requires a swarm of fundamental nomologically necessary connections between distinct mental and physical/functional properties, Dualistic Physicalism requires a swarm of fundamental *metaphysically necessary* connections between distinct mental and physical/functional properties: for instance, that it is metaphysically necessary that, if someone is in the ORANGE brain state, then they bear the irreducible awareness relation to the color orange. We might call these ‘modal danglers’, because they would dangle from the rest of the body of metaphysically necessary truths (the truths of logic and math, maybe the necessary truths governing parts and wholes, and so on). So Dualistic Physicalism and Dualism are equally uneconomic. Both require unexplained necessary connections between distinct mental and physical/functional properties – they just differ on the modal strength of these unexplained connections. Considerations of economy, then, do not favor Dualistic Physicalism over Property Dualism.

Mental Causation. A third argument for Physicalism can put as follows. Our brain states cause our behaviors. On Property Dualism, states of awareness (states of bearing the irreducible awareness relation to a certain Q-property) are distinct from our brain states. So the Property Dualist faces the following dilemma. If he says that states of awareness are also causes of our behaviors, then he gets rampant overdetermination. If he says that states of awareness are not causes of our behaviors, then he gets epiphenomenalism. The Identity Physicalist avoids this dilemma by identifying states of awareness with the brain states that cause our behaviors. So Identity Physicalism is to be preferred over Property Dualism.

But, once again, this argument does not carry over to Dualistic Physicalism. There is much to say on this subject (I say more in Pautz 2003a), but the Dualistic Physicalist, it seems to me, faces the very same dilemma as the Property Dualist. Suppose that a brain state ORANGE brain state causes some behavior – is “causally efficacious” in the production some behavior. The Dualistic Physicalist says that it is *metaphysically*

necessary that if one is in the ORANGE brain state, then one bears the irreducible awareness relation to the color orange. The Property Dualist says that it is only nomologically necessary. But the Dualistic Physicalist still faces the same dilemma that the Property Dualist faces. The stronger modal claim does not make the problem go away. If he says that bearing the irreducible awareness relation to the color orange is *also* causally efficacious with respect to the behavior, then he gets causal overdetermination. If he says that it is causally irrelevant, then he gets epiphenomenalism. In short, the dilemma between Rampant Overdetermination and Epiphenomenalism is generated by the following two claims. (i) States of awareness are not identical with any brain states (or even with functional states realized by neurobiological states). (ii) Brain states are the causes of our behaviors. Defenders of both Dualistic Physicalism and Property Dualism accept (i) and (ii). So both face the dilemma between Rampant Overdetermination and Epiphenomenalism.¹⁷

In sum, there is no reason to accept Dualistic Physicalism over Property Dualism. Both views are uneconomic, and both views have the same trouble with mental causation. Those who would advocate Dualistic Physicalism, I think, have lost sight of the traditional motivations for Physicalism. Once we give up Identity, the stronger modal thesis gets us nothing. We simply have no reason to accept it. This leads to the

SKEPTICAL CONCLUSION: Once we accept that qualitative mental properties are properties of people over and above their physical and functional properties, as I argued in the first stage of the argument, then we have no reason to Physicalist Supervenience (=Physicalism).

§7 Conclusion

In the first stage of the argument, I argued from Relationism and Internalism to Non-Identity. In the second stage, I argued that, once we accept Non-Identity, we have no reason to accept Physicalist Supervenience. I conclude that the two seemingly innocuous premises – Relationism and Internalism – create a serious problem for Physicalism.

REFERENCES

- Boghossian, Velleman 1989 Color as a Secondary Quality, *Mind*
Campbell, John 1993 A Simple View of Color Readings on Color, v 1, ed. Byrne and Hilbert
Fodor, J. 1990 A Theory of Content and Other Essays, Cambridge: MIT Press.
Hardin, C. L. 1988 Color for Philosophers
Hinton, J. 1973 Experiences. Oxford: Clarendon
Harman. G. 1990 The Intrinsic Quality of Experience
Horwich, Paul 1998 Meaning Oxford
Dretske, F. 1995 Naturalizing the Mind MIT Press

¹⁷ Of course, this is a vast subject, which I cannot hope to do justice to here. One might attempt to dodge the choice between Rampant Overdetermination and Epiphenomenalism. I argue elsewhere that this cannot be done (Pautz 2003a).

Here is one mistake that must be guarded against. The mistake is this. On non-reductive Physicalism, a mental property might be a functional property and so be causally efficacious. This is mistake, because the Dualist says that it is not even identical with a physical property. It is not just non-reductive, it is super non-reductive.

- Lycan, W. 1996 *Consciousness and Experience* MIT Press
- McGinn, C. 1996 Another Look at Color *Journal of Philosophy*
- Nagel, T. 2001. The Psychophysical Nexus. In *New Readings on the A Priori*, ed. Boghossian and Peacocke.
- Pautz, A. 2003a Physicalism and Complex Properties
----- 2003b Why be an Intentionalist?
----- 2003c Intentionalism and Neurobiology: An Empirical Problem
----- 2003d Physicalism about Color and the Unitary-Binary Distinction
----- 2003e Color as Illusion
- Shoemaker, S. 1994 Phenomenal Character *Nous*, 28: 21-38.
- Schiffer, S 2003 *The Things We Mean*. Oxford.
- Smart, J. C. 1959 Sensations and Brain Processes. *Philosophical Review*
- Soames, S 1998: The Modal Argument: Wide Scope and Rigidified Descriptions, *Nous*
- Thompson et al 1992. Ways of Coloring: Comparative Color Vision as a Case Study for Cognitive Science. *Behavioral and Brain Sciences*, 15: 1-74.
- Tye 1984 The Adverbial Approach to Visual Experience *The Philosophical Review*
- Tye 1994. Qualia, Content, and the Inverted Spectrum. *Nous*
- Tye, M. 1995 *Ten Problems of Consciousness*, MIT Press
- Tye, M. 2002 Representationalism and the Transparency of Experience, *Nous*
- Yablo 1995 Singling out Properties, *Philosophical Perspectives*.