2

A Simple View of Consciousness

Adam Pautz

I will argue for *primitivism* about sensory consciousness. On primitivism, sensory consciousness cannot be fully reductively explained in physical or functional terms. Others have defended primitivist views of color, personal identity, the intentionality of thought, semantic properties, and goodness.

My argument for primitivism will not be based on the usual a priori considerations, for instance the knowledge argument, the explanatory gap, or the thesis of revelation. Instead, the argument will be based on a philosophical claim about the structure of consciousness together with an empirical claim about its physical basis. The philosophical claim is that having an experience with a certain phenomenal character is a matter of bearing a 'consciousness relation' to a certain item external to the subject. For instance, intentionalism about sensory consciousness holds that having an experience with a certain phenomenal character is a matter of standing in an intentional relation to an intentional content into which external properties enter. The empirical claim is that phenomenology can vary due to internal differences. These two claims create a puzzle and I will argue that the only solution to the puzzle involves adopting the view that the consciousness relation is a simple relation—one that cannot be analyzed in terms of an individual's physical or functional relations to the external world.

Primitivism does not automatically lead to the rejection of physicalism—at least if physicalism is a mere thesis of supervenience. G. E. Moore held that goodness is primitive, yet supervenient on the natural as a matter of metaphysical necessity. Likewise, one could hold that the consciousness relation is primitive, yet supervenient on the physical as a matter of metaphysical necessity.

My plan is as follows. In sections 1 and 2 I introduce the two claims that will play a significant role in my argument. In sections 3–11 I develop the argument. Finally, in section 12 I briefly address the prospects for the view that the consciousness relation is primitive yet supervenient on the physical with metaphysical necessity.

1. THE RELATIONAL STRUCTURE OF SENSORY CONSCIOUSNESS

The first claim that will play a significant role in the argument is that a *relational view* of sensory consciousness is correct. Suppose you have a visual experience as of a tomato. A natural view is that having an experience with this phenomenal character is a matter of standing in a relation to an item that somehow involves the property of being red and the property of being round. Maybe the relevant item is a sense datum instantiating the properties, or the tomato instantiating the properties, or an intentional content that merely attributes the properties. In any case, the properties are not properties of your experience or your brain. Instead, they are properties of the object of your experience, if they are properties of anything at all. The relational view endorses this natural conception of experience. Say that a property is *external* iff it is not instantiated by an individual's experience or brain. Then the relational view holds that, for some types of experience, to have an experience with a certain phenomenal character is to stand in a certain relation to an item involving certain external properties; the phenomenal character of the experience is determined by the external properties that figure in the item. The argument I will be developing requires that the relational view applies to color experience, taste experience, and pain experience.

The relational view goes beyond the uncontroversial claim that in non-hallucinatory experience we are related to external items. On the relational view, phenomenal character is at least sometimes *constituted by* our relations to external properties, rather than by properties of our brains or experiences. For instance, on typical *sense datum theories*, having a visual experience with a certain phenomenal character is a matter of sensing mental objects whose properties determine the phenomenal character of the experience, for instance color and shape properties. These properties qualify as external in my sense, since they are not instantiated by the experience itself or by the brain. *Disjunctive theories* hold that the property of having an experience with a certain phenomenal character is the disjunctive property of standing in a certain relation to physical objects instantiating certain external properties *or* being in some other state. Disjunctive theories are akin to sense datum theories in holding that in some cases phenomenal character is determined by our relation to objects having external properties. *Intentionalist theories* are importantly different from sense datum and disjunctivist theories, but still count as relational in my sense. Whereas sense datum and disjunctivist theories hold that the determinants of phenomenology are *concreta* involving external properties, intentionalist theories hold that they are *abstracta* involving external properties. In particular, intentionalist theories have it that the determinants of phenomenology are *intentional contents* which involve external properties in the sense that the contents attribute them to

external objects. On most versions of intentionalism, the relevant contents are *propositions*. On another version of intentionalism, the *property-complex theory*, the contents are not propositions but *complex properties* or *property-structures* built up from external properties and spatial relations. In non-veridical cases the property-structures are not instantiated before one, but one is still related to them. I favor intentionalism and in this chapter I will be working with the property-complex version of intentionalism for convenience.[1]

There are also prominent theories which reject the relational view. The *identity theory* is one. On this theory, having an experience with a certain phenomenal character does not incorporate *any* external properties; it is necessarily identical with the property of being in a certain internal neural state. Phenomenal differences are always constituted by differences in non-relational neural properties.

One argument for the relational view of phenomenology is semantic: it provides the best explanation of why we use expressions for external properties, expressions such as *round, red*, or *in my foot*, to characterize phenomenology. For instance, we might truly say of two individuals undergoing hallucinations that one is conscious of every shape the other is conscious of; and the truth of such a report seems to supervene on the phenomenal characters of their experiences alone. We need a relation to serve as a semantic value of the expression *x is conscious of y* which occurs in this statement. Another argument is introspective: the relational view agrees with the transparency observation that when we try to focus on what our experiences are like we focus on external properties ostensibly instantiated by external objects or bodily regions. I think that the best argument is epistemic: the relational view is required to explain why merely having an experience with a certain phenomenal character necessarily grounds the capacity to have beliefs involving external properties, for instance shapes, colors, and properties ostensibly located in bodily regions. These are certainly not properties of our experiences or brains. I will not develop these arguments here. Suffice it to say that there are strong arguments for the relational view.[2]

As mentioned, I favor intentionalism and in this chapter I will be working with the property-complex version of intentionalism for convenience. I will call the relation we bear to the properties the *consciousness relation* and I will call the external properties the consciousness of which determines phenomenal character the *sensible properties*.

Some comments. First, I hold that the relational view is correct for all aspects of sensory phenomenology. But some disagree, holding for instance that the relational view is incorrect in the case of *blurriness*. And some hold that the

[1] For a defense of intentionalism, see Pautz (2007a) and Pautz (2008). For the property-complex theory in particular, see Johnston (2004).

[2] For a defense of the relational view, see Pautz (2007a) and Pautz (2008).

relational view fails for some types of non-sensory experiences, for instance moods and emotions. But, as we shall see, such exceptions would not matter to the argument. It is enough that the relational view is correct for color, taste, and pain experience. This is why above I equated the relational view with a restricted thesis only about these types of experiences. Second, some hold that the colors, tastes, and pains presented in experience are *response-dependent properties* in the sense that they are properties of objects or bodily regions concerning how they affect the nervous system. I will remain neutral on this view, but at one point my argument requires that this view cannot be extended more generally to all the sensible properties (see the discussion of the manifestation relation in section 7). The argument for this assumption is that having a series of visual experiences, even hallucinatory, is enough to give one the capacity to have beliefs involving *geometrical properties*, which evidently cannot also be identified with response-dependent properties of this form. So the epistemic argument for the relational view supports the additional claim that not all the sensible properties are such response-dependent properties.

2. THE PHYSICAL BASIS OF SENSORY CONSCIOUSNESS

There are obviously actual cases of perceptual variation, and they are much discussed by philosophers. The second claim that will play a large role in my argument for primitivism about sensory consciousness is that a certain type of perceptual variation is possible, but it is not one of the uncontroversial types of variation which philosophers typically discuss. Further clarification will be provided later on, but to a first approximation my second claim is that there are possible cases in which individuals bear the consciousness relation to different ostensible external properties of objects *even though their physical relations to external properties are the same*. In these cases the individuals involved are conscious of different external properties owing to internal differences between them. Now in the present section I only intend to introduce the claim; exactly how this claim will contribute to the case for primitivism will be revealed in the next section of the chapter, section 3.

I said that my second claim is that there are possible cases in which individuals bear the consciousness relation to different ostensible external properties of objects even though their physical relations to external properties are the same. In particular, I will argue that there are possible cases in which two individuals bear the consciousness relation to different ostensible external properties of objects even though they bear the *optimal cause relation* to the same properties of those objects. I choose to focus on the optimal cause relation because, as we will see in section 3, some philosophers have attempted to reduce the consciousness relation to this relation. The optimal cause relation may be defined as follows:

The optimal cause relation: x is in a state that plays the *e*-role and that would be caused by (for short, would track) the instantiation of external property *y* were optimal conditions to obtain.

The *e-role* is the functional role characteristic of brain states that realize experiences. On one view, the e-role is being poised to influence the formation of beliefs and desires. The notion of *optimal conditions* might be defined in different ways. Here I will equate them with conditions in which the sensory systems operate in accordance with design and result in adaptive behavior.

My argument that the relevant type of variation is possible will not be based on intuition. Indeed, because there are no a priori links between phenomenal and physical concepts, I do not think that issues concerning the physical basis of consciousness can be decided a priori. Rather my argument will based on the empirical finding that the phenomenology of our experiences is poorly correlated with the external properties we bear the optimal cause relation to when we have those experiences, and is much better correlated with the internal neural goings-on taking place in us then. I will express this by saying that there is *bad external correlation* and *good internal correlation*. I will provide examples involving color, pain, and taste experience. Then I will clarify the relevant type of variation, and argue that the empirical findings support its possibility.[3]

First, consider color experience. Some color experiences are of *unitary* colors. Some shades of red, green, yellow, and blue are unitary colors: they do not contain any hint of any other shades. All other color experiences are of *binary* colors: shades of orange, for instance, contain hints of red and yellow, and shades of purple contain hints of red and blue. In addition, color experiences resemble one another more or less closely, depending on the degree to which the colors presented in them resemble. But psychophysics has revealed that there is no simple relationship between the character of color experience and the reflectance properties we bear the optimal cause relation to when we have those color experiences. When we have unitary experiences there is nothing unitary about the reflectance properties that we then bear the optimal cause relation to, and when we have binary ones there is nothing binary about the reflectance properties we then bear the optimal cause relation to. And resemblances among color experiences are not matched by resemblances among the reflectance properties we bear the optimal cause relation to when we have those color experiences.

By contrast, neuroscience has revealed a very modest relationship between the activity of red-green (R-G) and yellow-blue (Y-B) neurons in the lateral geniculate nucleus (a kind of halfway house between the eyes and the visual cortex) and the character of color experience. Some models have it that in the

[3] The empirical results concerning color vision I will present come from Werner and Wooten (1979), Hunt (1982), Hardin (1988), De Valois and De Valois (1993); those concerning taste come from Stevens (1975), Borg *et al.* (1967), and Smith *et al.* (2000); and those concerning pain come from Stevens (1975) and Coghill (1999).

visual cortex there is a much better correlation. Granted, the details remain poorly understood. But given that the explanation of color structure is not to be found in the physical properties we bear the optimal cause relation to, the explanation must lie in the brain. When one has a unitary experience there is something special about the processing occurring in one then, and when one has a binary experience there is something binary about the processing occurring in one then. And resemblances among one's color experiences are matched by resemblances among the processing occurring in one then, even though they are not matched by the reflectance properties one then bears the optimal cause relation to.

In the case of pain, the situation is much the same. First, there is bad external correlation. Psychophysics has revealed that in the case of pain there is response expansion. There is a non-linear, exponential relationship between intensity of bodily disturbance and pain intensity. So if John's pain is twice as great as Jim's, then the bodily disturbance that John bears the optimal cause relation to might well be much *less than* twice as great as the one that Jim bears the optimal cause relation to. Why then is his pain twice as great? In the case of pain the evidence of good internal correlation is stronger than it is in the case of color vision. The neural response is amplified further downstream. So John's somatosensory neural discharge rates are twice as great as Jim's. It is only in the brain that we find a nice correlation between pain intensity and anything in the physical world. Indeed, there is a linear relationship between pain intensity and neuronal discharge frequency rates in many areas of the primary somatosensory cortex.

Likewise, in the case of taste, there is a non-linear correlation between the character of our taste experiences and the character of the chemical properties we then bear the optimal cause relation to. By contrast, there is a linear correlation between perceived sweetness and neural response, and resemblances among tastes are matched by resemblances among so-called *across-fiber patterns* in the brain.

In general, when we have experiences the external properties we bear the optimal cause relation to are a mess. The nervous system transforms the mess into something more manageable, and it is only in the brain that we find a nice correlation between experience and anything taking place in the physical world. I will now develop a two-stage argument from this to the second claim that will play a significant role in my argument for primitivism. This is the claim that there are possible cases in which individuals bear the consciousness relation to different ostensible external properties of objects *even though their physical relations to external properties are the same*.

In the first stage, I will argue for *the physical possibility of coincidence cases*. These are cases in which the following two physical conditions co-obtain. *First*, the properties two individuals bear optimal cause relation to (in a certain sense-modality) exactly coincide. *Second*, at the same time the individuals vary

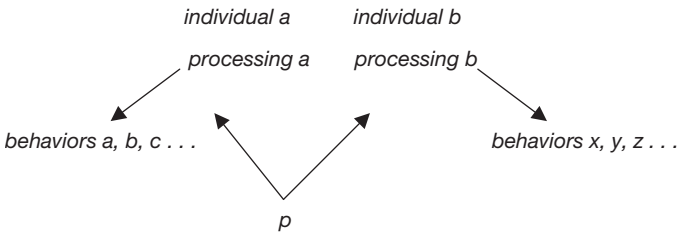

Fig. 2.1. The structure of a coincidence case.

profoundly in neural processing and functional organization. In particular, they are in quite different neural states, which play quite different output oriented functional roles with respect to behavior (see Figure 2.1). The possibility of such cases should be uncontroversial. Note that the first stage will be neutral on whether the individuals in such cases have the same experience or different experiences. This issue will be left open until the second stage.

In the second stage, I will use 'good internal correlation' and 'bad external correlation' to argue that, given that the individuals in coincidence cases differ profoundly in neural processing and functional organization, the most reasonable view concerning these cases is that in at least *some* of them an additional phenomenal condition obtains, namely, that the individuals also have different experiences. This is so despite the fact that they bear the optimal cause relation to the same external properties. Given good internal correlation and bad external correlation, the internal and functional differences are better evidence of phenomenal difference than the sameness of tracking is evidence of phenomenal sameness. This will provide an argument that does not rely on untutored intuition for the claim that experience can vary independently of optimal cause relations and other such relations to external properties. I will call this *coincidental variation*.

We begin, then, with the first stage. Unfortunately there are no obvious actual coincidence cases. As noted at the outset, the form of variation argued for here is importantly different from the forms of actual variation that philosophers typically discuss. To see this, consider *interspecies variation* first. Humans and pigeons differ profoundly in color processing and functional organization. But, since they have different receptor systems, they bear the optimal cause relation to *different* ranges of reflectances. So the second condition of coincidence cases, namely that the properties tracked are the same, is not met in this case. Consider *standard variation* next. On viewing a color chip with a certain reflectance property, Jack and Jill are put into different opponent processing states and differ functionally. So, in this one case, the neural states they are in are caused by

the same reflectance property. Since, we may suppose, the differences are within the range of normal, optimal conditions obtain, this looks like a coincidence case. But it might be argued that if we look at how their neural states respond to external properties under *all* optimal conditions, we find that under those conditions those neural states track *overlapping but distinct* ranges of reflectances. If so, then, on viewing the color chip, they might be in states that are *actually* caused by the same reflectance property, but they bear the *optimal* cause relation to distinct but overlapping reflectance properties. So these are not obviously coincidence cases.[4] Fortunately, it should be uncontroversial that there are *possible* coincidence cases, and this is all my argument will require. I will describe three. In the rest of the chapter, I will make essential use of all three of these cases in my argument for primitivism.

Mabel and Maxwell. Mabel and Maxwell occupy the same possible world but belong to different species that evolved on separate continents. By chance, Mabel and Maxwell evolved identical receptors systems. On viewing a fruit, they bear the optimal cause relation to exactly the same reflectance property, *r*. However, the fruit is an important food-source to Maxwell's species but not to Mabel's. So they evolved different postreceptoral wiring, with the result that *r* normally produces quite different color processing in Mabel and Maxwell. For instance, we might suppose that *r* normally produces 'unitary' opponent processing in Mabel that might underlie a vivid unitary color experience (for instance a unitary red experience), while it normally produces 'binary' opponent processing in Maxwell that might underlie a dull binary color experience (for instance, a desaturated red-yellow experience). We may also suppose that Mabel is easily able to pick out the fruit from the background foliage, while Maxwell has difficulty in this task. I will call the opponent channel state Mabel is in *u* and the different opponent channel state Maxwell is in *b*, because I will argue in the second stage of the argument that in at least some scenarios of this kind Mabel has a unitary color experience while Maxwell has a binary one.

Likewise in general. On viewing the same objects, Mabel and Maxwell bear the optimal cause relation to exactly the same ranges of reflectances, but they are put into neural states which differ in two ways. First, they differ in whatever neural respect underlies the distinction between the experience of unitary colors like red and the experience of binary colors like red-yellow. Second, they fall into different internal resemblance-orderings. So, for instance, if both Mabel and Maxwell look at the same two objects consecutively, Mabel might be put into two radically different neural states, while Maxwell is put into two similar neural states. In consequence, they differ markedly in their sorting, discrimination, recognition and other color-related behavior with respect to the

[4] This is explained more fully in Pautz (MSb); see also section 4 of the present chapter. It follows that, contrary to Byrne and Tye (2006: 250), coincidence cases such as the one developed in Pautz (2006) cannot be assimilated to cases of standard variation.

same objects. But when they track the same properties by way of different internal processing, optimal conditions obtain. Their visual systems operate differently, but when they do so they are operating exactly as they were designed by evolution to operate. And their behavioral dispositions, although different, are adaptations to different selection pressures. Thus, Mabel and Maxwell constitute a coincidence case, because they bear the optimal cause relation to properties that exactly coincide, but they vary profoundly in neural processing and functional organization.

Yuck and Yum. Yuck and Yum belong to different species. If they taste the same foodstuff under optimal conditions, then their taste systems respond to the same chemical property of that foodstuff, *c*. So, they bear the optimal cause relation to the same property, *c*. However, the foodstuff is poisonous to Yuck but not poisonous and indeed an important food-source to Yum. In consequence, they so evolved as to respond to *c* with different across-fiber patterns (which, as we saw above, are well-correlated with taste experiences in the actual world) and different affective reactions. For instance, Yuck withdraws from it violently, while Yum is drawn to it. I will call the across-fiber pattern Yuck undergoes *d* and the one Yum undergoes *p*, because the second stage of the argument I will argue that in at least one scenario of this kind the patterns realize a displeasing and pleasing taste experience, respectively.

Likewise in general. When Yuck and Yum taste the same foodstuffs, they bear the optimal cause relation to the same properties of those foodstuffs, but they undergo quite different across-fiber patterns and exhibit different taste-related affective and sorting behaviors. The neural and behavioral differences do not impugn the assumption of optimality. These differences evolved naturally. Moreover, they are adaptive, since the same foodstuffs have different nutritional values for Yuck and Yum. I believe that there are actual cases of roughly this kind. But, to avoid controversy, I will continue with the hypothetical case.

It may be said that in this scenario Yuck and Yum do not bear the optimal cause relation to exactly the same properties, contrary to what I have said. In particular, on tasting the foodstuff, Yuck bears the optimal cause relation to the dispositional property of being poisonous for Yuck and Yum bears the optimal cause relation to the dispositional property of being healthy for Yum. But this is ruled out if we make an additional supposition. Suppose that the foodstuff has two chemical properties, *c* and *c*′. The property which is responsible for the foodstuff's being poisonous for Yuck and for its being healthy for Yum is *c*′. However, *c*′ has no causal effect on their taste systems. *A fortiori*, the foodstuff's being poisonous or healthy has no causal effect on their taste systems. Instead, only the other chemical property *c* has a causal effect on their taste systems. Since the optimal cause relation is defined in causal terms, it follows that Yuck and Yum do not bear the optimal cause relation to the foodstuff's being poisonous or healthy. Instead, they only bear the optimal cause relation to the causally relevant chemical property *c*, as originally stipulated.

Mild and Severe. Two communities of pain-perceivers evolve separately. Mild belongs to one community and Severe belongs to the other community. Both occasionally experience bodily disturbance d in the leg. In Mild's community, d is not very dangerous. So d normally puts his primary somatosensory cortex into state m involving a certain mild rate of firing of neurons. Recall that in our own case there is a linear correlation between the neural discharge frequencies of the relevant neurons and pain intensity. By contrast, in Severe's community, d is much more dangerous. For instance, maybe it is more susceptible to dangerous infections in this community because the community occupies an environment in which bacteria are more plentiful. In consequence, in Severe, d normally causes somatosensory state s, involving a rate of firing of somatosensory neurons which is twice as great as that which is involved in m. As a result, Severe attends to his leg with greater urgency than does Mild. But optimal conditions obtain in each case, because the different behaviors are completely adaptive given the noted difference in the significance of the damage to them. So, Mild and Severe bear the optimal cause relation to the same property, d, but they differ radically in pain processing and behavior.

Of course, there are indefinitely many such possible cases in which two individuals differ profoundly in neural processing and functional organization but bear the optimal cause relation to the same external properties. Everyone must accept the physical possibility of coincidence cases, for these two physical conditions are certainly compossible. The real question is not whether such cases are possible, but whether the individuals in some such cases have the same or different experiences.

Now for the second stage of the argument. I will argue that the best view is that in at least *one* such coincidence case an additional phenomenal condition obtains: the individuals involved have different experiences. This is so despite the fact that they bear the optimal cause relation to the same properties. This yields *coincidental variation*. Of course, I think that this is true in many such cases. But, as we will see in section 4, my argument only requires that it is true in one. I offer two arguments for this claim.

First, as we have seen, experiential properties are very well correlated with neural properties and very poorly correlated with the external properties we bear the optimal cause relation to. This suggests that, if two individuals stood in the optimal cause relation to the same external properties but differed in the relevant neural properties, then they would have different experiences. In other words, translating from counterfactual language into the language of possible worlds, in at least some nearby possible worlds in which coincidence cases actually obtain, the individuals have different experiences, even though they bear the optimal cause relation to the same properties. What is being invoked here is a general principle: if we know that magnitudes x and y are well correlated but x and z are not, then we have some reason to believe that, if two objects differed on y but were the same on z, they would still differ on x.

Second, the individuals in the cases exhibit robust and systematic differences in color-related, taste-related, and pain-related behavior. We may suppose that the differences are not learned but innate. And we may suppose that they are widespread in the relevant populations.[5] To explain these behavioral differences, the opponent of coincidental variation might say that the individuals involved have experiences with the same phenomenal characters, but have systematically different beliefs and desires about the same objects. But this is a poor explanation because the behavioral differences are supposed to be innate and widespread. Further, if the individuals involved do not have different experiences, there would be no explanation of why they have systematically different beliefs and desires about the same objects. The only reasonable explanation is that in at least some of the cases they have different experiences, in accordance with coincidental variation.

The alternative to accepting coincidental variation is holding that the individuals *in every possible coincidence case* have the same experiences in spite of the vast neural and behavioral differences between them (or else are Zombies who have no experiences at all, a possibility I will ignore). This is simply unbelievable. Imagine meeting Yuck and Yum, Mild and Severe, or Mabel and Maxwell. To say that they have the same experiences in spite of all the evidence against this would be unreasonable.

Coincidental variation says that, in *some* possible coincidence cases, internal and functional *differences* are accompanied by phenomenal *differences*. It would be a mistake to confuse coincidental variation with the much-discussed thesis of internalism. Internalism says that only internal factors are relevant to phenomenology, so that, in *every* possible case, internal *sameness* guarantees phenomenal *sameness*. As we will see at the end of section 7, some might say that there are functionalist reasons to doubt this pure internalism. Coincidental variation is quite consistent with the externalist view that sensory consciousness is somehow determined jointly by the properties tracked on the input side, internal factors, and behavioral dispositions on the output side. This would yield a form of externalism, but with internal as well as external factors playing a role. My argument for primitivism only requires coincidental variation. The issue of internalism is not relevant here, and I am neutral between pure internalism and some form of externalism.

Coincidental variation says that in at least one coincidence case the individuals involved have different color, taste, or pain experiences in spite of bearing the optimal cause relation the same external properties. On a non-relational view such as the identity theory, their having different experiences simply consists in their having different internal neural states. This is not so on a relational view. For instance, on the property-complex version of intentionalism assumed here, their having different experiences consists in their bearing the consciousness relation to

[5] My thanks to Fred Dretske for pointing out that the argument is stronger if it is supposed that the behavioral differences are innate and widespread.

different external color, taste, or pain properties. (Coincidental variation is neutral on the issue of whether these different properties are different response-dependent properties of objects and bodily regions, or different projected properties that the objects and bodily regions do not actually have.) So, on a relational view, coincidental variation means that in at least one coincidence case two individuals, *a* and *b*, bear the *consciousness relation* to *different* color, taste or pain properties *x* and *y*, despite bearing the *optimal cause relation* to the very *same* property *p*. Those who combine the relational view and variation will say that this is somehow *owing to* the internal or functional differences between them. Diagramatically:

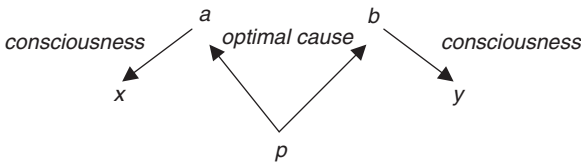

Fig. 2.2. The relational view and coincidental variation entail that the depicted situation obtains in some coincidence cases.

On a non-relational view such as the identity theory, coincidental variation is not puzzling. On such a view, the individuals' having different experience simply consists in their being in different neural states. By contrast, the combination of a relational view and coincidental variation creates a puzzle. On this combination of views, two individuals can be conscious of different *external* properties owing to *internal* or *functional* differences between them. In other words, the consciousness relation is at once externally directed and internally sensitive. Now I do not say that this is incoherent. On the contrary, since we have good reasons to accept both the relational view and coincidental variation, I believe that it is true. That it is not incoherent may be brought out with an analogy. The relation *x has mass-in-grams y* is a relation between objects and numbers which are ''external'' to objects, but what numbers objects bear this relation to is sensitive to the ''internal'' mass properties of those objects. Still, I admit that the combination of the relational view and coincidental variation is puzzling. I will argue that the only solution to the puzzle involves adopting a primitivist view of sensory consciousness according to which the consciousness relation is a primitive relation that cannot be analyzed in terms of an individual's physical or functional relations to the external world.

The argument applies to any version of the relational view. It may seem that the argument does not apply to disjunctivism because the disjunctivist has a radically externalist view of consciousness that is inconsistent with coincidental

variation. Elsewhere I attempt to show this is not the case: the argument applies to disjunctivism as well.[6] The disjunctivist can and indeed must accommodate coincidental variation, and the only way they can do so is by adopting a primitivist view of consciousness. However, as noted in the previous section, here I will focus on how the argument plays out on the type of intentionalist view I favor.

3. THE STRUCTURE OF THE ARGUMENT

By *primitivism* about a property (or a relation, that is, a polyadic property) I just mean the denial of reductionism. *Reductionism* about a property holds that it is a complex property constructible from the fundamental physical and functional properties of the world. Here I will use 'physical property' to mean all and only such complex properties. So I understand reductionism broadly to include the various forms of functionalism, even though others would consider them to be non-reductionist views. And I understand primitivism about sensory consciousness to be the strong claim that some properties or relations involved in sensory consciousness are properties or relations over and above all those constructible from the fundamental physical and functional properties of the world. How do the relational view and coincidental variation create an argument for primitivism about sensory consciousness? In the present section, I will indicate the structure of the argument that I will be developing.

On a relational view, every episode of sensory consciousness has two components: the consciousness relation and the complex of sensible properties to which we bear this relation. In the history of philosophy perceptual variation has often been used to draw conclusions about the nature of the sensible properties. By contrast, I will use a unique type of perceptual variation, coincidental variation, to draw a conclusion about the nature of the *consciousness relation*, namely that it is primitive. I will set aside the second component of sensory consciousness, the sensible properties that are *relata* of the consciousness relation. I will give to the reductionist about sensory consciousness any view of the sensible properties they wish: they might identify them with response-independent physical properties, response-dependent physical properties, or primitive properties. My argument will be entirely neutral on this issue.

The argument for primitivism about the consciousness relation will take the form of a dilemma. We may divide all physical relations into two categories. Our most promising reductive theories of the consciousness relation identify it with a physical relation that the individuals in coincidence cases bear to the *same* properties. For example, one such theory identifies the consciousness relation with the optimal cause relation. I will call such physical relations *A-type relations*.

[6] See section 12 of Pautz (2007b).

The idea is that the mind's capacity to be conscious of the external items can be explained in terms of a causal process from those items to minds. Indeed, it is very difficult to see *how else* we might reductively explain the consciousness relation. But, as we will see, there are also physical relations that the individuals in coincidence cases bear to *different* properties. I will call such physical relations *B-type relations*.

I will argue that there is principled reason to believe that the consciousness relation cannot be an A-type or B-type relation. Since these exhaust all physical relations, this will be an argument against reductionism and for primitivism. The argument will unfold as follows. Previously, I argued for a relational view of sensory consciousness and for coincidental variation. These two claims entail that there is a consciousness relation with the following two properties:

Relationality In at least some cases, the consciousness relation holds between individuals and external properties, for instance shapes, colors, pains felt in bodily regions, and tastes felt in the tongue.

Variation The consciousness relation is such that *some* pairs of individuals in coincidence cases bear it to different external properties.

These properties yield constraints on the reduction of the consciousness relation. Evidently, they immediately entail that the consciousness relation is not an A-type relation, thereby ruling out our most promising reductive theories of this relation. Such relations satisfy the relationality constraint: they are relations between individuals and external properties. But, by definition, they do not satisfy the variation constraint. For instance, in at least some coincidence cases two individuals bear the *consciousness relation* to *different* sensible properties, but they bear the *optimal cause relation* to the very *same* property (see Figure 2.2). So far, I have focused on the optimal cause relation. But I will generalize the argument to other A-type relations. This will be the easy part of the argument.

The larger and more difficult part of the argument will involve showing that the consciousness relation cannot be identified with a B-type relation. To rule out B-type relations, the relationality constraint and the variation constraint will be insufficient. By definition, B-type relations satisfy the variation constraint. And, as we will see, some satisfy the relationality constraint as well. So we will have to rely on considerations that have not yet been introduced. I will argue that these relations are ruled out by two other properties of the consciousness relation:

Scrutability The consciousness relation is the subject of our talk and thought about consciousness.

Extensionality The consciousness relation has a certain actual-world extension—individuals bear it to countless shapes, colors, and so on.

As we will see, B-type relations may be subdivided into two categories: those defined in *internal terms* and those defined in *functional terms*. I will argue that there is principled reason to think that B-type relations defined in internal terms

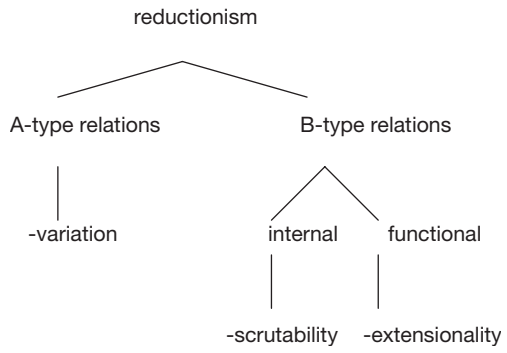

Fig. 2.3. The structure of the variation argument against reductionism and for primitivism.

fail to satisfy the scrutability constraint and B-type relations defined in functional terms fail to satisfy the extensionality constraint. So the complete structure of the argument is represented in Figure 2.3.

Coincidental variation plays a key role in this argument. It rules out our otherwise most promising theories of the consciousness relation, namely A-type theories. For this reason I will call it the *variation argument* for primitivism.

4. A-TYPE RELATIONS FAIL TO SATISFY THE VARIATION CONSTRAINT

The most common reductions of the consciousness relation are A-type. The idea is that sensible properties like colors, tastes, and pains are physical properties that external objects and bodily regions actually possess. And the consciousness relation is some A-type relation between individuals and such properties. Here are several A-type relations that the consciousness relation might be identified with:

The optimal cause relation: x is in an internal state that plays the *e*-role and that would be caused by the instantiation of external property *y* were optimal conditions to obtain.

The indication relation: x is in an internal state plays the *e*-role and that has the biological function of indicating external property *y*.

The asymmetric relation: x is in an internal state that plays the *e*-role and whose tokening asymmetrically depends on the instantiation of *y*.

The input–output relation: x is in an internal state that plays the *e*-role and that under optimal conditions tracks the instantiation of *y* and that in turn enables *x* to distinguish objects that have *y* from objects that do not.[7]

[7] For these relations, see, respectively, Tye (2000), Dretske (1995), Fodor (1990), and Armstrong (1968).

Actual forms of variation are not a problem for these theories. Consider a case of standard variation. On viewing a chip, John and Jane bear the consciousness relation to different color properties, namely unitary blue and green-blue, owing to internal differences. But, as we saw in section 2, it might be said that John and Jane also bear the optimal cause relation to *different* but overlapping reflectance properties r and r' of the chip. Likewise for the other A-type relations on the list. This would mean that this case is not a coincidence case in my sense. And it would mean that the case is not a problem for the view that the consciousness relation is the optimal cause relation. The optimal cause theorist could say that the consciousness relation is the optimal cause relation, that r is unitary blue, and that r' is green-blue. This entails that the chip is unitary blue and green-blue, and that John is conscious of the first color while Jane is conscious of the second color. In general, the combination of the relational view and *actual* cases of standard variation is not problematic because it can be said that what is going on is that every object has a set of multiple colors, and on viewing the objects different individuals bear the optimal cause relation to colors in the set.[8] The same strategy applies to interspecies variation and indeed all actual forms of variation in color experience. On this view, colors are response-independent properties, and objects have many of them. So this is a kind of *color pluralism*.[9] One could imagine similar views of taste and pain.

By contrast, the relational view and hypothetical coincidence cases create a decisive argument against A-type theories. This argument is just the two stage argument for coincidental variation presented in section 2. The first stage established *the physical possibility of coincidence cases*: there are possible coincidence cases in which two individuals bear the optimal cause relation to *exactly the same* properties, but vary profoundly in internal neural processing and behavior. The idea is that, even though objects and bodily regions have multiple properties, the individuals in these cases bear the optimal cause relation to the very same properties of those objects or bodily regions. A moment's reflection will reveal

[8] See Pautz (MSb). This is how the intentionalist who accepts the optimal cause theory can solve Johnston's (MS, chapter 5) selection problem.

[9] For color pluralism about interspecies variation cases, see Byrne and Hilbert (2003) and Tye and Bradley (2001). For color pluralism about standard variation cases such as John and Jane, see Byrne and Hilbert (1997: 273) and Tye (2000: 91). It should be noted that, while these philosophers continue to accept color pluralism in cases of interspecies variation, they now reject it in cases of standard variation: they now maintain that different minimal colors within human color space are incompatible, so that in these case at least one individual must get it wrong. Pautz (MSb) argues these philosophers would do better to accept color pluralism in both cases, as they once did. For, as we have seen, in both cases color pluralism follows from the optimal cause theory; indeed, it follows from all available reductive theories of our consciousness of colors. But, as I am about to explain in the text, color pluralism does not help the reductionist about the consciousness relation when it comes to hypothetical *coincidence cases*. For in these cases, even if objects have many colors, the individuals involved bear A-type physical relations to exactly the same colors of objects, yet it is reasonable to suppose that they bear the consciousness relation to different ostensible colors of those objects.

that the individuals also bear the other A-type relations listed above to the same properties. The first stage of the argument should be uncontroversial. The second stage argued for *coincidental variation:* in view of the profound neural and behavioral differences, the most reasonable view is that in at least *one* such case the individuals involved have different experiences and so bear the consciousness relation to *different* properties. The most reasonable view, then, is that the consciousness relation is not identical with the optimal cause relation or any of the other A-type relations on the list (see Figure 2.2). This is an argument against A-type theories that does not rely on the mere *intuition* that phenomenology could vary independently of A-type relations.[10]

It might be objected that the optimal cause relation and the other relations on the list are vaguely specified. Maybe, then, there is some precisification of the optimal cause relation or one of the other relations on the list that is not vulnerable to the argument. In response, what we have here is a general

[10] Kalderon (2007) has taken up the view that Byrne and Hilbert (1997, 2003) and Tye (2000) once accepted (see the previous note): color pluralism in cases of both interspecies variation and standard variation. He also accepts *selectionism*, which is a view concerning what determines what colors of objects we are conscious of. He writes that 'the relation between object, perceiver, and circumstances of perception . . . determines the perceptual availability of [one of the many colors of an object]' (2007: 577). Later he says that the determination proceeds by way of something about color similarity: 'given the nature of Norm's visual system, Norm's visual system selects certain relations as relations in color similarity and, hence, which colors are perceptually available to Norm' (2007: 593). The selectionist component of Kalderon's view is difficult to understand, but I think that coincidence cases may create a problem for it. What is it to select a relation as a relation of color similarity? And how precisely does the visual system determine what colors are perceptually available *by* determining what relations are relations of color similarity? In the first quote, Kalderon speaks of a relation between the object and the perceiver as determining what color of the object the perceiver is conscious of, but does not specify what this relation is. On one natural interpretation of selectionism, the relevant relation is a *causal relation:* the mechanism of selection is causation. Then selectionism is very similar to Tye's optimal cause theory. Humans and pigeons are conscious of different colors of the same objects because, owing to their different receptor systems, their visual systems are causally sensitive to different colors of those objects. The pigeons but not the humans are causally sensitive to ultraviolet light (Tye and Bradley 2001). The optimal cause theory entails a similar story about cases of standard variation such as John and Jane, as I explained in the text. (In the previous note, I explained that Tye previously accepted such a parallel pluralist view of standard variation, but that he now rejects it in favor of an inegalitarian view, even though this seems inconsistent with his optimal cause theory.) But if selectionism is explained in terms of causation (and it is hard to see how else it might be explained), so that it is like the optimal cause theory, then it is also refuted by coincidence cases, in which two individuals are relevantly causally related to exactly the same color properties of objects and relations among objects, but it is nevertheless reasonable to hold that they bear the consciousness relation to different ostensible color properties of those objects. Of course, the two views endorsed by Kalderon are separable. One could accept a pluralist response-independent view of color and reject a selectionist view of color-consciousness (if such an account is indeed inconsistent with coincidental variation). Instead, one could combine color pluralism with a broadly internalist view of color consciousness. By a *broadly internalist* view, I mean one that entails that, on viewing the same objects, Maxwell and Mabel bear the consciousness relation to different colors owing to internal differences, even though their visual systems respond to the same chromatic properties of those objects. (As noted in section 2, such an account is not committed to pure internalism.) But I believe that, once we accept a more internalist view of color-consciousness, an epistemic problem arises for the pluralist response-independent view (see note 21).

recipe for refuting the view that the consciousness relation can be identified with *any* relation of this kind. Let *r* be any physical relation within the general ballpark of the relations listed above. There will always be a class of possible cases in which two individuals bear *r* to the *same* external properties, but differ profoundly in neural processing and functional organization. The claim that the individuals in at least one of these cases have different experiences and so bear the consciousness relation to *different* properties will always be more plausible than the philosophical theory that the consciousness relation is identical with *r*.[11]

5. ARE THE CASES POSSIBLE?

Responding to an earlier presentation of this argument directed specifically against the optimal cause theory, some commentators have objected that coincidence cases are *impossible*.[12] Presumably, they do not mean to reject *the physical possibility of coincidence cases*: that there are possible cases in which two individuals track the same properties under optimal conditions while differing profoundly in internal processing and functional organization. As we saw in the first stage of the argument of §2, this claim ought to be uncontroversial. On the only reasonable

[11] Lycan (2000), who says he tends to accept the present argument against the optimal cause theory, helpfully made the suggestion of presenting the argument in this way: the argument is that the claim of coincidental variation that in at least some coincidence cases the individuals involved have different experiences is more plausible than any philosophical theory, such as the optimal cause theory, which delivers the contrary verdict.

[12] The earlier presentation is Pautz (2006; see also Pautz (2003) and the commentators are Byrne and Tye (2006). Byrne and Tye raise four further objections to the earlier presentation of the argument, the first two of which rely on misunderstandings. First, in the earlier presentation, I introduced the thesis of *Dependence* and said that it has the consequence that in coincidence cases the individuals involved have different experiences. Byrne and Tye consider two interpretations of Dependence and argue that on neither does it have this consequence. My reply is that neither of these interpretations is correct. On the correct interpretation, Dependence is *equivalent to* the thesis that in coincidence cases the individuals involved have different experiences, so that the entailment is trivial (Pautz 2006: 207). Here I have used the more appropriate title of *coincidental variation* for this thesis, and I have offered a different, two-stage formulation of my argument for this thesis and against A-type theories such as the optimal cause theory. Second, Byrne and Tye object that the failure of existing A-type theories would not show that externalism about phenomenology is false and internalism is true because externalism is not committed to any particular reductive theory (2006: 251). This objection, too, relies on a misunderstanding, because in the earlier presentation of the argument I did not take myself to have shown that externalism is false and internalism is true, but only that all of the versions of externalism I considered in the paper are false (2006: 228). Given my language in the earlier presentation, the misunderstanding was natural. I now call these theories *A-type theories* to remove the impression that my target is externalism in general. In fact, in my (2006) and in the present chapter, I take no stand on the issue of externalism *versus* internalism (see section 2 of the present chapter). Third, Byrne and Tye (2006: 252) point out that A-type theories are often vague, which makes it unclear whether they are refuted by coincidental variation. I addressed this objection at the end of section 4 in the present chapter. Fourth, Byrne and Tye argue that the failure of every existing reductive theory of the consciousness relation would not show that reductionism fails (2006: 252). In other words, we could take the view that the correct reductive theory is unknown. I call this view *mysterian reductionism* and argue against it in section 11.

interpretation, they are rejecting *coincidental variation.* In other words, they are rejecting the further claim argued for in the second stage of the argument that in some of these cases an additional phenomenal condition obtains: the individuals involved bear the consciousness relation to *different* sensible properties. If they are right in rejecting this claim, then of course my argument fails. Their rejection of coincidental variation requires their acceptance of the radical view that in every coincidence case the individuals involved bear the consciousness relation to the same sensible properties in spite of the profound neural and functional differences between them. (I ignore the view that the individuals are not conscious of any sensible properties at all.)

Of course this radical view follows from the philosophical theory that the consciousness relation is the optimal cause relation, but one would like a non-question-begging argument for it. The argument seems to be as follows. If, for instance, Mabel and Maxwell bear the consciousness relation to different color properties on viewing the fruit, then at least one of them must be conscious of a color that the fruit does not have, because such color properties are mutually exclusive. But this is incompatible with the condition that they track the same external properties *under optimal conditions*. According to the objection, contrary to coincidental variation, the only verdict compatible with this condition is that they bear the consciousness relation to the very same color and so have the same color experience. This is so despite the fact that there are profound differences between them in opponent processing and color-related behavior. So this is the verdict we should accept. Call this the *argument from error* against phenomenal variation in coincidence cases.

Now, since coincidental variation (and hence my argument against the optimal cause theory) only requires phenomenal variation in *one* coincidence case, the argument from error against coincidental variation is successful only if it is general. For instance, it must also be assumed that, if Yuck and Yum bear the consciousness relation to different tastes on tasting the same food, one must be wrong about the food's true taste, so that this verdict is inconsistent with the optimality condition. In that case, as against coincidental variation, we must accept the implausible verdict that they are conscious of the same taste, in spite of the radical neural and behavioral differences between them.

One problem with the argument from error against coincidental variation is the implausibility of its key assumption that phenomenal variation in these cases requires error. Those who provide pluralistic theories of color would deny this in the case of color vision. Indeed, as I explained in section 4, I think that the optimal cause theorists themselves should deny that variation requires error in cases such as John and Jane. For my part, I hold that phenomenal variation requires error in the case of color vision because I accept a general color exclusion principle. Indeed I accept a projectivist theory of color according to which all color experience involves error. But I reject the assumption in the cases of taste and pain. Here the assumption is very implausible. Why couldn't individuals

from different communities have different taste or pain experiences in response to the same stimulus, and yet both have true beliefs about the tastes of things in their communities or about the pains they feel in their bodies? So the argument from error against coincidental variation cannot succeed because the assumption that phenomenal variation entails error does not hold in general.

But I think that there is a more basic problem with the argument from error that applies even if the assumption is granted. The problem is that the optimality condition and error are incompatible only if optimal conditions are defined as conditions in which there is no error, that is, in which individuals are conscious of properties that objects actually have. But, since the defender of the optimal cause theory is attempting a reduction of this intentional relation, they cannot define optimal conditions in terms of notions such as error which are explained in terms of that very relation. Instead, they must define optimal conditions in terms notions such as adaptive fitness, design, and so on.[13] So even if we grant the implausible assumption that phenomenal variation would in every case entail error on the part of one of the parties involved, there is no reason to think that it is inconsistent with their tracking the same properties under optimal conditions. Indeed, we can imagine many possible cases of adaptive error. So why cannot coincidence cases be cases of this kind? Of course, if the optimal cause theory were correct, there could not be cases of this kind, for it is a kind of verificationist theory of sensory content according to which tracking under optimal conditions is inconsistent with error. But I am offering an argument against this theory. In view of the arguments offered in section 2, the claim that there is phenomenal variation in at least some coincidence cases is the most reasonable one to make, even if it means that at least one of the individuals is in error. It is much more reasonable than the alternative view that in all such cases the individuals involved have the very same experiences in spite of the radical neural and functional differences between them.

So the original argument succeeds. The individuals in coincidence cases bear the optimal cause relation and the other A-type relations to the *same* properties. But, in view of the profound internal and functional differences between them, the most reasonable view is that some such individuals have different experiences and so bear the consciousness relation to *different* properties. So the consciousness relation cannot be identical with the optimal cause relation or any other A-type relation.

6. NO B-TYPE RELATION SATISFIES THE OTHER CONSTRAINTS

This brings us to B-type relations. The individuals in coincidence cases bear such relations to different properties. So such relations satisfy the variation constraint

[13] For this point, see also Chalmers (2005).

on the reduction of the consciousness relation. This means we need another argument to rule out B-type relations.

As I already mentioned, my argument is that B-type relations either fail to satisfy the scrutability constraint or the extensionality constraint. The individuals in coincidence cases differ only in two respects. First, they differ internally, in particular in neural processing. Second, they differ functionally, in particular in how their internal states guide their behaviors. So B-type relations fall into two categories. The first category contains B-type relations defined in internal terms and the second category contains B-type relations defined in functional terms. I will provide general reasons to think that B-type relations belonging to the first category fail to satisfy the scrutability constraint and those belonging to the second fail to satisfy the extensionality constraint (see again Figure 2.3).

7. RELATIONS DEFINED INTERNALLY

The view that the consciousness relation can be identified with such a relation is very unpromising and to my knowledge no one has advocated such a view. But we must get the view out of the way before considering the view that the consciousness relation can be defined functionally.

An initial hurdle is to see how a relation defined in internal terms might satisfy the relationality constraint. Consider:

The brain state relation: x is in a brain state that plays the *e*-role and that has internal neural property *y*.

The problem with this relation is the reverse of the problem with A-type relations. It satisfies the variation constraint and hence is a B-type relation. But it fails to satisfy the relationality constraint. The semantic, introspective and epistemic arguments mentioned in section 1 show that to have an experience with a certain phenomenal character is to stand in a relation to shapes, colors, pains felt in bodily regions, tastes felt in the tongue, and so on. These properties are not all neural properties instantiated in the brain. If they are instantiated at all, they are instantiated by external objects or bodily regions, not by parts of the brain. So the consciousness relation at least *sometimes* holds between individuals and external properties. By contrast, the brain state relation *never* holds between individuals and external properties; it always holds between individuals and neural properties of their own brains. So the consciousness relation is distinct from the brain state relation.[14]

[14] Alex Byrne proposed in discussion that the optimal cause theorist could handle the cases of Yuck and Yum and Mild and Severe by claiming that these individuals bear the optimal cause relation and hence the consciousness relation to different neural properties instantiated by their own brains. The trouble is that, like the brain state view, this proposal violates the relationality constraint, because it entails that the phenomenology of experience is always constituted by the

To obtain a relation that satisfies both the relationality constraint and the variation constraint, the reductionist needs an algorithm that goes from internal states to *external* properties and that is sensitive to the internal or functional properties of those states. We are now after an algorithm that is sensitive to the internal properties of those states. I do not suppose that the algorithm must be specifiable in some relatively compact way. But I do require that it satisfies all four constraints on the reduction of the consciousness relation.

The reductionist might rely on the analogy with the mass-in-grams relation mentioned at the end of section 2. It is a relation between objects and numbers which are 'external' to objects, and yet it is sensitive to the 'internal' mass properties of objects. And it can be defined in terms of a kind of structural isomorphism between masses and numbers. Likewise, the reductionist might claim that there are structural isomorphisms between our internal states and the external properties we are conscious of, and claim that the consciousness relation is definable in terms of these isomorphisms. As noted in section 2, neuroscience has revealed a very modest relationship between the activity of R-G and Y-B neurons in the lateral geniculate nucleus and the degree to which the colors we are conscious of are reddish, greenish, yellowish and bluish. And there appears to be a linear relationship between neural discharge frequencies of neurons in the primary somatosensory cortex and the intensity of the pains we are conscious of. Now, presumably, there are infinitely many or at least very many possible but non-actual sense modalities, and corresponding to each of them there might be a different algorithm of this kind. We could truly say of any creature possessing such an alien sense-modality that it is conscious of properties that we are not conscious of. This leads to the idea that the consciousness relation might be identified with:

The infinitely disjunctive relation: x is in a color state c and $f(c) = y$ or x in is in a pain state p and $g(p) = y$ or x is in some alien state a and $h(a) = y$ or x is in some alien brain state a' and $i(a') = y$ or . . . and so on for every possible sense-modality.

However, there are a few reasons to doubt that there are any such modality-specific algorithms as *f, g, h, i*, . . . First, neuroscience has only revealed a very imperfect relationship between the activity of R-G and Y-B neurons in the lateral geniculate nucleus and the degree to which the colors we are conscious of are reddish, greenish, yellowish and bluish. Some hold that the discrepancies are corrected further downstream, but there is no evidence of this. If anything, it is more confusing as we move to the cortex. Second, sensible properties are not

consciousness of internal rather than external properties. So it is inconsistent with the arguments for accepting a relational view such as intentionalism mentioned in section 1: the semantic argument, the phenomenological argument, and the epistemic argument. The proposal is especially implausible in other cases. For instance, it would be very implausible to suggest that colors are neural properties and Mabel and Maxwell have different color experiences because they bear the optimal cause relation and hence the consciousness relation to different neural properties of their own brains.

easily quantifiable, so it is difficult to believe that there might be algorithms concerning them. Think of tastes and sounds, for instance. Third, we are not merely conscious of sensible properties; we are conscious of *property-structures* in which properties are presented at various locations. So the relevant algorithms would have to go from intrinsic neural properties to spatially arrayed property-structures. But it is very hard to imagine such algorithms. In humans, there is topographical mapping, but it is much too rough to provide such an algorithm.

The absence of algorithms would obviously not be inconsistent with my claim in section 2 that internal factors explain some aspects of phenomenology, so that internal differences, when accompanied by functional differences, are evidence of phenomenal differences. Here I take no stand on whether or not there are such algorithms.

If it should turn out that there are no algorithms, then the reductionist who hankers after an internal definition of the consciousness relation has no choice but to define the conscious relation in terms of an infinitely long list:

The infinitely disjunctive relation II: x is in total internal state b_1 and $y =$ property structure s_1 or x is in total internal state b_2 and $y =$ property structure s_2 or . . ., and so on for every possible property-structure.

So now we have before us two infinitely disjunctive relations. Evidently there are relations of this kind which satisfy the relationality constraint, the variation constraint and the extensionality constraint. When individuals are in different internal states b_1 and b_2, they bear this relation to different property structures s_1 and s_2, which might involve different external color, taste, or pain properties.

Could the consciousness relation be identical with either of these relations? The problem is that these relations do not satisfy the scrutability constraint. There are many infinitely disjunctive relations $r_1, r_2, r_3, \ldots$ with different extensions. For instance, consider the brain state b' of a creature, Blurg, we have never before encountered. These different infinitely disjunctive relations $r_1, r_2, r_3, \ldots$ map b' onto different sensible properties. The problem is that none of the infinitely disjunctive relations $r_1, r_2, r_3, \ldots$ could be the semantic value of our predicate *x is conscious of y*. This follows from a theory that the semantic value of a predicate is the most natural property or relation that fits our use of the predicate.[15] For all of these relations $r_1, r_2, r_3, \ldots$ fit actual use, and they are equally natural because they share the same very low degree of naturalness. What could make it the case that the semantic value of *x is conscious of y* is one of these relations to the exclusion of the others? Indeed, none of these relations is a relation we *could* think about. By contrast, the consciousness relation is evidently the semantic value of *x is conscious of y*. And it is a relation we can think about. So the consciousness relation cannot be identical with any one of these infinitely disjunctive relations. Nor would it do to say that *x is conscious of y* indeterminately refers to all of these

[15] This is an oversimplified version of Lewis (1983a).

relations, so that there are no determinate truths about what properties Blurg is conscious of. There are such truths, even if we do not know what they are.

At this point, the reductionist seeking a definition of the consciousness relation in internal terms might appeal to a trick in order to avoid infinitely disjunctive relations. In particular, he or she might adopt a kind of global response-dependent theory according to which the sensible properties one and all are relational, dispositional properties of the form *normally causes internal state s.* If the external properties we are conscious of are relational properties whose *relata* are internal states, then it is easy to specify an algorithm going from internal states to such properties. In particular, the consciousness relation might simply be identified with:

The manifestation relation: x is in internal state *s* that plays the e-role and *y* = the property of normally causing *s*.

In other words, the idea is that we can simply say that a person is conscious of such a sensible disposition just in case they are in the internal state which is the manifestation of that sensible disposition. On this view, when Mabel is in *u* and Maxwell in *b*, Mabel bears the manifestation relation to *normally causing u* and Maxwell bears the manifestation relation to *normally causing b*. The colors they are conscious of are identical with these relational, dispositional properties. Likewise for Yuck and Yum and Mild and Severe. So this is a B-type relation that satisfies the variation constraint. Further, one might think that it could be the semantic value of *x is conscious of y* and hence could satisfy the scrutability constraint, on the grounds that it is the most natural relation that fits our use of this expression.

The problem is that this relation does not satisfy the extensionality constraint. As noted at the end of section 1, it is a fact about the actual extension of the consciousness relation that not all the properties we are conscious of are response-dependent properties of the form *normally causes internal state s*. For instance, having an experience with a certain character is enough to ground the capacity to have thoughts involving shapes. And shapes are not properties of this form, since it is obvious that objects might have shapes while entirely lacking such properties. For instance, objects might have had shapes, even if creatures with internal states had never evolved. And, in the actual world, objects that are too small to have an effect on perceivers have shapes but lack properties of this form. So this trick fails and the original conclusion stands.

Therefore there is a *principled* reason to think that relations defined in internal terms are bound to fail to satisfy the scrutability constraint. The reason is that there is an abundance of equally natural infinitely disjunctive algorithms going from the internal properties of an individual to external properties.

There is another potential problem with the view that the conscious relation is definable in internal terms. Those with functionalist intuitions will say that the idea that our internal neural properties *alone* determine what properties we are conscious of is somewhat implausible. Consider a twist on the case of

Mild and Severe. Two species are hardwired so that the *same* rate of firing among somatosensory neurons produces in them radically different pain-related behavior, and in general plays quite different functional roles in them. Theories according to which the consciousness relation is definable in purely internal terms entail that they are conscious of the same pain. Or consider a brain in a vat belonging to no actual species with no sense organs or motor-output system, and so without even the potential to act on the world. Theories according to which the consciousness relation is definable in purely internal terms entail that the brain in a vat has a vivid inner life. Or imagine some somatosensory neurons firing in a Petri dish, completely functionally isolated. Some (but not all) theories according to which the consciousness relation is definable in internal terms might entail that the Petri dish is conscious of pain properties! Reductionists with functionalist tendencies will reject such theories and adopt a theory according to which the consciousness relation is defined at least in part in functional terms.

Whereas the problem with relations defined in internal terms was that they fail to satisfy the scrutability constraint, the chief problem with relations defined in functional terms will be that they fail to satisfy the extensionality constraint. I will consider the consumer relation and the interpretation relation.

8. RELATIONS DEFINED FUNCTIONALLY: THE CONSUMER RELATION

On the *consumer theory*, the consciousness relation is identical with:

The consumer relation: x is in an inner state *s* that plays the e-role and that represents property *y*, where *s represents* a property *y* iff, in the past, when an object was present with property *y*, and the consumer devices used *s* to perform output behaviors *a*, *b*, *c*, . . . these behaviors frequently had advantageous results *because* an object with property *y* was present, so that now individuals have consumer devices that might use *s* to perform *a*, *b*, *c*,. . .[16]

This relation is defined in functional terms inasmuch as it appeals to how our internal states are used to guide behavior. Unlike many of the A-type theories we have considered, it is not an entirely input-based theory.

The stock illustration of this theory is the frog. A frog state *b* is caused by flies and causes tongue-darting. On the consumer theory, it does not represent the property of being a black dot. The instantiation of this property does not enter

[16] See Millikan (1989) for a consumer theory of intentional relations in general. Lycan (2006) says that the consumer theory is likely to deliver the correct verdict that in coincidence cases the individuals are conscious of different properties. Against this, in what follows I argue that the consumer theory does not deliver correct verdicts in coincidence cases. I should mention that Lycan does not go so far as to defend the consumer theory. Instead, he appears to endorse what I will call in section 12 *mysterian reductionism*.

into an explanation of why tongue-darting was beneficial in the past. Rather, it represents the response-dependent, biologically significant property of being frog-food. It is this property which enters into the explanation.

The evaluation of the consumer theory is complicated by the fact it is very difficult to see what types of properties individuals bear the consumer relation to in more complex cases, in particular in coincidence cases. There are two interpretations. On the first, the consumer relation fails to satisfy the variation constraint. On the second, it fails to satisfy both the variation constraint and the extensionality constraint.

On the first interpretation, the individuals in coincidence cases bear the consumer relation to the same response-independent physical properties, that is, the same reflectance properties, chemical properties, and bodily disturbance properties. The idea is that, although they exhibit different behavioral patterns, the same external properties enter into the explanation of the evolution of those behavioral patterns.

This interpretation is arguably incorrect. For instance, Mabel is in *u* and Maxwell is in *b*. By contrast to the frog's internal state, in the past, these states were used to perform a great variety of behaviors, such as picking out a fruit, finding a mate, avoiding a predator, and so on. Intuitively, on none of these occasions was the behavior advantageous *because* the presented food or individual had reflectance property *r*. This is simply not a true because-statement. Maybe *r* is *correlated with* properties that enter into true because-statements, such as being healthy food, being a potential mate, being a predator. But *r* itself does not enter into such a true because-statement. The same problem applies in the cases of Yuck and Yum and Mild and Severe.

Nevertheless, suppose this interpretation is correct for the sake of argument. Then the consumer relation is just another A-type relation which fails to satisfy the variation constraint. On this interpretation the individuals in coincidence cases bear the consumer relation to the same response-independent physical properties, just as they bear the optimal cause relation to the same response-independent physical properties. But the most reasonable view is that in at least some coincidence cases the individuals involved bear the consciousness relation to different properties. So on this interpretation the consciousness relation is not identical with the consumer relation.

On a second interpretation, on tasting the foodstuff, Yuck bears the consumer relation to the response-dependent property *being poisonous to his kind* and Yum bears it to the response-dependent property *being healthy to his kind*. The consumer theorist might say that these properties *just are* the different taste properties they are conscious of. Then, in this case at least, the consumer relation satisfies the variation constraint.[17]

[17] The suggestion that the different properties Yuck and Yum are conscious of are properties of this sort was made by Andy Egan in discussion.

This interpretation is more plausible than the first. Intuitively, it is these biologically significant properties which explain why in the past avoidance-behavior was advantageous to members of Yuck's species and why pro-behavior was advantageous to members of Yum's species. And this interpretation is in better accord than the first with what consumer theorists say about the frog. The frog does not represent the response-dependent property *being a black dot* that his visual system tracks, but the untracked response-dependent property *being food for frogs* which enters into the explanation of the relevant behavioral pattern.

But under the second interpretation the consumer relation fails to satisfy the variation constraint in cases other than Yuck and Yum. For instance, what different coarse-grained response-dependent properties do Mabel and Maxwell bear the consumer relation to as they view the fruit? Indeed, they arguably do not bear the consumer relation to *any properties at all.* As noted above, in the past, *u* and *b* were used to produce a variety of behaviors. And a variety of different properties explained why these different behaviors were advantageous. There is no *single* property which frequently explained why these behaviors were advantageous. So in this case there is *no* property which satisfies the consumer theorist's formula.

Consider next Mild and Severe. What different properties do they bear the consumer relation to? *Being a dangerous case of bodily damage* and *being an even more dangerous case of bodily damage?* These do not seem to be very good candidates for the different properties they are conscious of.

Finally, consider *adaptively neutral* coincidence cases. It is obvious that a creature's internal processing and overall functional organization are not rigidly determined by selection pressures. So we can imagine cases in which two creatures respond to the same external properties, and are under roughly the same selection pressures, but in which they evolved different internal processing and functional organizations by chance. In these cases, the same objects have the same coarse-grained, biologically significant properties with respect to the two creatures. On the second interpretation, then, the individuals involved cannot bear the consumer relation to different such properties. But the most plausible view is that they bear the consciousness relation to different properties.[18]

Another problem with the consumer relation under the second interpretation is that it fails to satisfy the extensionality constraint. To see this, suppose as before that Yuck tastes a foodstuff with chemical property *c* and undergoes across-fiber pattern *d*. But now suppose that Yuck takes another taste of the foodstuff, this time from a different part of the same foodstuff which has a slightly different chemical property *c′*. We might suppose that *c* is the property of having a certain concentration of a certain type of molecule and that *c′* is the property of having a slightly greater concentration of the same type of

[18] For an *adaptively neutral* coincidence case involving color vision, see Pautz (2003) and Pautz (2006).

molecule. As a result, Yuck now undergoes slightly different across-fiber pattern d', and withdraws from the foodstuff even more violently than before. On the second interpretation, in each case Yuck bears the consumer relation to the same coarse-grained response-dependent property, namely *being poisonous to his kind*. But, in view of his different internal processing and behavior, the best view is that on the different occasions he bears the consciousness relation to slightly different taste properties, the second more displeasing than the first. There are actual cases of this kind. For instance, on the second interpretation, when a bird takes two bites of a poisonous dart frog and gets a greater concentration of batrachotoxin on taking the second bite, the bird bears the consumer relation to the same property *being poisonous to its kind* on both occasions, but arguably bears the consciousness relation to slightly different tastes.

It may be said that there is a third interpretation of the consumer theory. On this interpretation, individuals bear the consumer relation to *fine-grained* response-dependent properties. For instance, in the case just mentioned, on the different occasions, Yuck bears the consumer relation to the property of normally causing across-fiber pattern d and then the property of normally causing across fiber pattern d'. So on this interpretation the consumer relation might satisfy the extensionality constraint. But this interpretation cannot be correct. Consider the case of color vision. Previously, I argued that *no* external properties satisfy the consumer theorist's formula in the case of color vision, because there are no properties that frequently explained the advantageousness of the many behaviors we used the color vision system to perform in our evolutionary past. If so, then on viewing objects individuals bear the consciousness relation to a myriad of colors, but there are no properties at all that they bear the consumer relation to. In particular, they do not bear the consumer relation to such fine-grained response-dependent properties as the property of normally causing opponent channel state u or the property of normally causing opponent channel state b. Or consider the case of Yuck again. Here again the interpretation fails. In the past, d was used by members of Yuck's kind to avoid the relevant foodstuff. But this behavior was never advantageous *because* the foodstuff had the property of normally causing that very brain state, d'. There is simply no sense in which this is a true because-statement. If anything, it was advantageous because it had the property of being poisonous to Yuck's kind. Therefore, when he undergoes across-fiber pattern d Yuck certainly does not bear the consumer relation to the property of normally causing that very brain state, d. If Yuck bears the consumer relation to any property, it is the property of being poisonous to Yuck's kind, as on the second interpretation. But, as we have seen, on this interpretation, the consciousness relation cannot be identified with the consumer relation, for the reasons explained previously.[19]

[19] The *success theory* of Papineau (1993) has some similarities to the consumer theory. In the case of the belief relation, it holds that *A has a belief according to which state s obtains* just in case *s* is

9. RELATIONS DEFINED FUNCTIONALLY: THE INTERPRETATION RELATION

On the interpretation theory, the consciousness relation is identical with:

The interpretation relation: x is in internal state *s* and the best 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 interpretation of the members of *k* is the one that best satisfies the constraints on interpretation, given the functional roles of their internal states.[20]

This needs to be unpacked. Functional roles are second-order properties of internal states to do with their interactions with the external world, their interactions with other inner states, and their interactions with behavior. Constraints on interpretation are principles taken from our common sense theory of persons about how mental states change as a result of evidence from the external world, and how they combine to produce behavior. They include the following general principles. *Rationalization:* assign beliefs, desires and other mental states so as to rationalize behavior. *Humanity:* assign beliefs that are reasonable on the evidence, and desires that reflect sane values. *Eligibility:* all else equal, assignments are to be preferred that assign contents involving natural properties. Roughly, the best interpretation is the assignment of mental states—beliefs, desires, and experiences—to the internal states of members of *k* which best satisfies the constraints on interpretation, given the functional roles of those states.

I have formulated the interpretation theory in a way that requires a unique best interpretation. The interpretationist might instead claim that there might be several interpretations that are tied for best. But I will continue to assume a formulation in terms of a unique best interpretation. Afterwards we will see that the problems I will raise apply even if we allow for multiple best interpretations.

In coincidence cases, the individuals involved exhibit radically different color-related, taste-related and pain-related behaviors on the output side. So they differ functionally on the output side. This will mean that, *if* there are such things as best interpretations of them, they will assign to their brain states experiences of quite different external color properties, taste properties, and pain properties, despite the fact that they track the same properties on the input-side. Whether

the state of the world which would guarantee that *A*'s belief, in combination with different desires, would lead to actions that satisfy those desires. But the theory does not apply to the consciousness relation. When a person is conscious of a shade of orange, he or she has no desires that are so specific that their satisfaction requires that the viewed object instantiate that very shade of orange. Even if he or she did—for instance, even if he or she had the desire to have an object with *that very color*—its content would derive from the content of his or her experience. So, one could not without circularity explain the fine-grained content of his or her experience in terms of the fine-grained content of such a desire.

[20] See Lewis (1983b).

these properties are response-dependent properties, primitive properties, or whatever, need not concern us. So it might be thought that the interpretation relation satisfies both the relationality constraint and the variation constraint.

But there are four problems. The first is familiar. Let *unique functional role* be the claim that, for every different possible experience, there is a unique functional role that belongs to it necessarily. There are reasons to think that this is false. Two individuals might have different experiences that have the same functional roles. For instance, there might be functionally identical individuals with different color experiences, as in spectrum inversion. Or there might be two simple creatures who have different pain experiences, but who are so wired that their pain experiences play exactly the same input–output functional role. Since the detail in sensory experience is so vast (think of listening to music, for example), it seems easy to imagine other cases of phenomenal differences between individuals without even potential functional differences. Even if functional role contributes to determining what properties we are conscious of, it is not the whole story. Internal factors play a role as well. But if *unique functional role* is false and such cases are possible, then the consciousness relation cannot be the interpretation relation. For, since the best interpretation of a population of individuals is only sensitive to the functional roles of their states, the best interpretations of the individuals in such cases would assign the individuals experiences of the same properties. So they would bear the interpretation relation to the same properties but bear the consciousness relation to different properties.

The second and third problems are independent of the well-trodden problem posed by spectrum inversion and other such problem-cases for *unique functional role*. If the functionalist rejects the relational view of sensory consciousness, and only recognizes monadic experiential properties, then once he or she establishes *unique functional role* the functionalist is home free. For then he or she can identify these monadic experiential properties with properties of the form: being in a state which plays functional role f. By contrast, on a relational view, even if *unique functional role* is true, the functionalist with reductive aspirations faces a difficult further issue. For, on a relational view, there also exists a dyadic consciousness relation to external properties which is involved in every sensory episode. As we have seen, there are good semantic, phenomenological, and epistemic reasons for believing that there is such a relation. Here I have been assuming an intentionalist version of the relational view according to which the relevant relation is an intentional relation between individuals and contents, in particular property-structures. So, on a relational view, the truth of *unique functional role* would not automatically vindicate reductionism about sensory consciousness. There would remain the problem of reducing the consciousness relation. The reductive functionalist must make it plausible that there is a relation defined in terms of functional role that satisfies the four constraints on the reduction of the consciousness relation, so that the consciousness relation might be identified with it. To make this plausible, he or she must be able to at least gesture at

an algorithm, *a*, for going from the functional roles of any actual or possible individual's internal states to the external properties the individual is conscious of. Then the reductive functionalist may say that the consciousness relation is identical with a relation of the following form: *x* is in a state with some functional role *f* and $a(f) = y$. The only way to define such an algorithm is *via* the notion of the *best interpretation* of a population. So the interpretation relation is the only relation of this form with which the consciousness relation might be identified. The second and third problems concern the notion of a best interpretation.

The second problem is that the interpretation relation is defined in terms of the property of being the best interpretation. But it is not even clear that there is such a property. So it is not even clear that there is such a relation as the interpretation relation.

Some philosophers have advocated best system theories of laws. Systems of generalizations vary in simplicity and strength. Since simplicity and strength are variable magnitudes, we might think that there is such a property as being the best system, although there are serious problems concerning when the virtues of simplicity and strength add up to an overall best system. But what variable magnitude do candidate interpretations of persons vary along such that the best interpretation may be defined as the one that maximizes this magnitude?

Let me rule out some suggestions. First, the interpretationist cannot identify the property of being the best interpretation with the property of being the correct interpretation, on pain of circularity. Second, some interpretations are more reasonable than others, so one might think that the property of being the best interpretation is the property of being the most reasonable interpretation light of the functional evidence. But this would undercut the reductive aspirations of the theory, for now it is appealing to an unreduced notion of reasonableness of interpretation. In addition, this would make the interpretation theory very implausible, for we may imagine cases in which the most reasonable interpretation is mistaken. For instance, if an unfortunate species is so wired up that when it has an experience of an object straight ahead, it is disposed to reach slightly to the right, the most reasonable interpretation will be mistaken. Third, the property of being the best interpretation might be identified with the property of doing the best job overall of conforming to the constraints on interpretation. But this is not much of an advance. Since the constraints on interpretation are *ceteris paribus* (indeed, they are so vague it is not even clear that they express propositions), it is clear that different interpretations, which are intuitively not equally good, might all satisfy the few constraints on interpretation. So we would need an explanation of the key notion invoked here, the notion that some interpretations *better* conform to the constraints on interpretation. Since there is no account of the property of being a best interpretation, I think that the interpretation theory cannot even get off the ground. Nevertheless, I will raise a third and final problem, which applies even if this initial problem can somehow be overcome.

The third problem is as follows. Grant that there is such a property as the property of being the best interpretation, and so such a relation as the interpretation relation. I will argue that there is no interpretation of any human or non-human population that has the property of being the unique best interpretation. Since the interpretation relation is defined in terms of the notion of the best interpretation, it follows that individuals do not bear it to any properties. In other words, it is entirely without application. But, of course, individuals bear the consciousness relation to many properties. So even if there is such a thing the interpretation relation, it fails to satisfy the extensionality constraint.

Consider humans first. For example, suppose Mabel is disposed to sort two objects together. There are many different interpretations: she is conscious of two shades of red and desires to sort red objects together; she is conscious of two shades of green and desires to sort green objects together; she is conscious of two unitary colors and desires to sort objects with unitary colors together. In fact, once we remember that interpretations can differ in fine-grained ways—one might assign the experience of red_{17} and another might assign the experience of red_{18}—we see that the number of different possible interpretations that rationalize the sorting behavior perfectly well is fantastically large. There are two reasons to believe that none of these interpretations has the property of being the uniquely best interpretation. *First*, it is simply unbelievable. True, exactly one of the interpretations has the property of being the correct interpretation. However the property of being the best interpretation cannot be the property of being the correct interpretation, on pain of circularity. So the interpretation theory requires that there is some *other* property such that exactly one of the interpretations has this property and all the others lack it. But it is clear that there is no such property. *Second*, by their nature, the constraints on interpretation are insufficient to whittle down the interpretations to a single best interpretation. It is well-known that *Rationality* is insufficient on its own. If a man goes to a bar, one interpretation that rationalizes his behavior is that he wants a saucer of mud and believes he can get one at a bar. To rule out such perverse interpretations, defenders of the interpretation theory must appeal to another constraint, *Humanity*. We should assume that individuals have reasonable beliefs given the evidence available to them, and that they have desires that reflect the same values that we have. But in the case of sensory content this fix fails. As we have seen, in the case of sensory content as in the case of belief and desire content, there are indefinitely many interpretations that rationalize the individuals' behaviors. And here *Humanity* is inapplicable because it makes no sense to say that certain experiences are unreasonable on the basis of an individual's evidence while others are reasonable. So there is no constraint on interpretation available to whittle down all of these interpretations to a single best interpretation.

The problem extends to animals and aliens. Imagine that we discover a species on earth or on an alien planet that has sophisticated sensory systems and exhibits behavior that is finely tuned to its environment. But suppose its sensory systems

and behavior are radically different from ours. For instance, sometimes members of the species turn purple and expand. Now there is reason to believe that a member of this population, for instance *Blurg*, is conscious of indefinitely many fine-grained sensible properties, perhaps ones belonging to alien quality spaces. In the formal mode, if we were to say *Blurg is conscious of indefinitely many fine-grained sensible properties*, we would express a truth. So if *x is conscious of y* picks out the interpretation relation, as defenders of the interpretation theory say, then Blurg must bear this relation to indefinitely many fine-grained sensible properties that we cannot imagine. That is to say, the best interpretation of Blurg and the others in the population must assign to their brain states experiences of indefinitely many sensible properties. But, again, there is no such thing as the best interpretation. There are many interpretations that rationalize the bizarre behavior more or less well. Some differ radically: they assign experiences of different sensible properties from entirely disjoint quality spaces to the same brain states. Some interpretations differ less radically: they assign experiences of quite different sensible properties from the same quality space to the same brain states. Others differ less radically: they assign experiences of sensible properties that differ only in fine-grained ways to the same brain states. Nevertheless, many of these different interpretations could rationalize the strange behavior equally well. For the same two reasons given above in connection with humans, it is simply unbelievable that the functional roles of their brain states and the constraints on interpretation could determine that among these interpretations one that stands out as the uniquely best one.

The defender of the interpretation theory might reply that, in the human case and the alien species case, exactly one interpretation has the property of being the unique best interpretation, and attempt to explain our reluctance to believe this by saying that what interpretation has this property is deeply epistemically opaque. But this does not answer the problem. What could this nebulous property *being the best interpretation* be? As we have seen, exactly one interpretation has the property of being the correct one, but the defender of the interpretation theory cannot explain the property in this way, on pain of circularity. What the interpretation theory requires is that there is some *other* property that exactly one interpretation has and all the others lack. For the two reasons given above, the claim is completely unbelievable, and the present reply does nothing to make it more believable.

Another reply is that the best interpretation is determined by input-oriented functional role. Mabel and Maxwell bear the optimal cause relation to *r*, Yuck and Yum bear the optimal cause relation to *c*, and Mild and Severe bear the optimal cause relation to *d*. According to the present reply, the best interpretation will accordingly assign to them experiences of *r*, *c*, and *d*. But evidently this will not be the best interpretation. *If* there is a best interpretation, it will assign to them experiences of different external properties of objects and bodily regions, so as to provide the best rationalization of their radically different behaviors on the output

side. (Incidentally, it is unclear what these different properties of the objects and bodily regions might be: response-dependent properties, primitive properties, or whatever.) As we have seen, there is no best interpretation of this kind.

I have formulated the interpretation theory in terms of a unique best interpretation. The interpretationist might instead claim that there can be several different interpretations that are tied for best. An initial problem with this proposal is that, even if there were a unique set of interpretations that are tied for best, the arguments given previously show that they could radically differ as regards what properties an individual experiences. Depending on what the interpretationist says about such cases, the interpretation theory would entail that the individual is conscious of no properties at all, or that the individual is conscious of many properties, or that it is radically indeterminate what properties the individual is conscious of. None of these consequences agrees with fact. But there is a more serious problem. Not only is there no such thing as a single best interpretation, there is no such thing as a single set of interpretations that are tied for best. There will simply be many interpretations that rationalize individuals' behaviors. For the two reasons given above, no subset of these stands out as the interpretations that are tied for best.

So the third problem is that, even if the second problem can be overcome and some account of the property of being the best interpretation can be provided, individuals do not bear the interpretation relation to any properties at all. But, of course, individuals bear the consciousness relation to many properties. So even if there is such a relation as the interpretation relation, it fails to satisfy the extensionality constraint.

I conclude that there is *principled* reason to think that relations defined in functional terms will fail to satisfy the extensionality constraint. The problem with relations defined in purely internal terms was that there is no *scrutable* algorithm from the purely internal properties of any actual or possible individual to the external properties that the individual is conscious of. The problem with relations defined in purely functional terms is that there is *no algorithm at all* for going from the functional roles of any actual or possible individual to the external properties the individual is conscious of. The only possible algorithm proceeds *via* the defunct notion of a best interpretation. It might be said that I have forgotten algorithms defined in terms of both internal factors and functional factors. But such algorithms will simply face both of these problems.

10. HOW A PRIMITIVE RELATION MIGHT SATISFY THE CONSTRAINTS

By contrast, a primitive relation might easily possess all four of the properties possessed by the consciousness relation: relationality, variation, scrutability, and extensionality. This, together with the fact that there are systematic reasons to

think that no physical relation possesses these properties, gives us excellent reason to think that the consciousness relation is a primitive relation.

A primitive relation could obviously be a relation between minds and external property-structures. So it could satisfy the relationality constraint. In addition, what property-structures we bear it to might partly depend on internal and functional factors, and not depend only on what properties are tracked in the external world. (Whether the dependence here holds with metaphysical or merely nomological necessity is an issue addressed in section 12.) So, unlike A-type relations, it could also satisfy the variation constraint. The fine-grained internal processing and functional organization of an individual might fully determine what fine-grained properties the individual bears the relevant primitive relation to. So, unlike B-type relations defined in functional terms, it might satisfy the extensionality constraint. Finally, unlike an infinitely disjunctive B-type relation defined in internal terms, a primitive relation might satisfy the scrutability constraint. On one view, which was mentioned in section 7, the semantic value of *x is conscious of y* is the most natural relation that fits our use. The relevant primitive relation fits our use. And it is perfectly natural. Of course, it might *supervene on* an infinitely disjunctive, extremely unnatural relation of the kind discussed in section 7. On a more externalist view, it might *supervene on* a combination of internal factors and external factors. But the relation itself, as opposed to its supervenience-base, is perfectly natural. Since it is bound to be the most natural relation that fits our actual use of *x is conscious of y*, it is bound to be the semantic value of this expression. Once use plus naturalness determine that the expression *x is conscious of y* refers to this primitive relation, we will then be able to use it to state truths about instantiation of this relation in cases such as the case of Blurg (discussed in sections 7 and 10) which lie outside of actual use.

For reasons that I will not go into here, primitivism about the consciousness relation goes best with primitivism about colors, tastes, and pains. In principle, it might be combined with any of three versions of primitivism. In the case of color, they are as follows. First, *response-independent primitivism*. On this view, objects have certain response-dependent primitive colors, and they had these primitive colors prior to the evolution of color vision. However, this view faces an epistemic problem. On an A-type theory, misperception under optimal conditions is not possible. If we evolved so that we bear the optimal cause relation to a color of an object, we are bound to be conscious of that color. But, given coincidental variation, such A-type theories are mistaken. Internal factors play some role in determining what colors we are conscious of, so that misperception under optimal conditions is possible. For instance, if we evolved so that we bear the optimal cause relation to a dull color of an object, we might be conscious of a bright color, owing to our internal processing. That might be so if the object is an important food source. Now what internal wiring we evolved is insensitive to the response-independent primitive colors that objects had prior to the evolution of color vision. Instead, it was determined by the unique

set of selection pressures that operated on our ancestors, determined by their habits, dietary needs, and environments. So if we evolved internal wiring that makes us conscious of colors that occasionally coincide with the true response-independent colors of objects, then this can only be a lucky accident. Intuitively, this means that response-independent primitivism has the drawback of entailing that we cannot be credited with *knowledge* of the response-independent colors of objects in our environment, even if by a lucky accident we so evolved that we occasionally have true beliefs about the colors of those objects. In addition, response-independent primitivism is implausible for tastes and pains. Second, *response-dependent primitivism.* On this view, necessarily, an object has a primitive color just in case it is disposed to cause individuals to bear the consciousness relation to that primitive color under normal conditions. So, for instance, if on viewing a fruit Mabel is conscious of unitary red and Maxwell is conscious of red-yellow, then the fruit instantiates both of those primitive colors. Color vision and the primitive colors of external objects co-evolved. This view avoids the consequence of response-independent primitivism that we can only veridically perceive by accident, and so is compatible with the claim that we know the colors of things. But it violates our intuitions about color incompatibility, and it is unattractively complicated. Third, *projectivist primitivism.* On this view, which is the view I favor, we bear the consciousness relation to primitive colors, but nothing at all instantiates them. Colors live only in the contents of our experiences. I take a similar view of tastes and pains. It may be the common sense view that experience provides us with knowledge of the mind-independent sensible properties of things. On none of these three versions of primitivism about the sensible properties is this common sense view correct. The failure of the common sense view is an inevitable consequence of the combination of the relational view and coincidental variation.[21]

[21] For response-independent primitivism, see Campbell (1993). For response-dependent primitivism, see McGinn (1996). For projectivist primitivism, see Pautz (2006: 235), Pautz (MSa), Pautz (MSb), and Chalmers (2006). It should be noted that the epistemic problem raised in the text for the response-independent view applies equally to a pluralistic version of this view according to which before the evolution of color vision every object had a *cluster* of similar response-independent colors, but not every single color: for instance, on this view, an object might have various shades of red, but no shades of green, yellow, or blue (Kalderon 2007: 581). Given the role of internal factors in determining color-consciousness, we might have so evolved that, on viewing an object, we bear the consciousness relation to color properties that lie entirely outside of its color cluster. And if we so evolved that occasionally we bear the consciousness relation to properties lying within the color clusters of objects, this must be counted a lucky accident. So this view entails that we can never be said to have knowledge of the colors of things. (Of course, the problem is avoided by an even more radically pluralist response-independent view which makes veridicality assured by holding that every object has *every* color, but such a view is not to be taken seriously.) The pluralist version also violates our intuitions about color-exclusion because it holds that objects have all the colors different normal perceivers perceive them to have. One possible reply is that different perceivers are always conscious of colors from disjoint color families, and that color exclusion only holds within a color family (Kalderon, 2007: 583). But since some of the primitive colors one person perceives will exactly resemble some of the primitive colors another person perceives, this view goes against the principle

11. WHAT SHOULD WE CONCLUDE?

There are principled reasons to think that reductionism about the consciousness relation is false and primitivism is true. But sometimes it is not reasonable to follow an argument wherever it leads. The reductionist might say that the argument for reductionism is stronger than the argument I have presented against it. There are two reductionist views which might be adopted in face of the variation argument: mysterian reductionism and compromise reductionism.

Mysterian reductionism is the conjunction of three claims. First, there is a consciousness relation, and the constraints on its reduction I have put forward are correct. Second, reductionism is correct: there is a physical relation which satisfies the constraints, and with which the consciousness relation is identical. Third, we cannot even gesture at what this physical relation is because we are still in the early days of the reductionist program.[22]

To evaluate this response to the variation argument against reductionism, we must consider the argument for mysterian reductionism and the argument against it. The chief argument for mysterian reductionism is that the rival view of primitivism requires danglers (brute laws connecting the internal and functional properties of individuals with what sensible properties they bear the primitive consciousness relation to) and faces problems with mental causation. So even if we cannot come close to finding a physical relation that satisfies the four constraints, maybe we should simply conclude that we have not looked hard enough.

Another argument is that existing theories show promise in handling simple cases and this provides reason to think a suitably elaborate and detailed descendant of one of these theories will work for the more complicated examples.[23] Consider the optimal cause theory. And consider a simple case concerning belief rather than sensory consciousness. Jack and Jill view a cow. Jack sees it from close up and says *that is a cow*. Jill sees it from far away and says *that is a horse*. Intuitively, the right verdict in this case Jack and Jill bear the belief relation to different propositions. The optimal cause theory applied to belief accommodates this intuition. In the case of Jack, optimal conditions obtain, so the content of *that*

that for universals exact resemblance entails identity. And in any case it is not clear how it answers the intuitive objection. To take an example from section 4, intuitively, the unitary blue color that John is conscious of and the green-blue color that Jane is conscious of *exclude*, even if we say that they are from different families. These problems cast doubt on the claim that color pluralism is the best view of variation consistent with our pretheoretical conception of colors (Kalderon 2007: 584). Further, in its primitivist version, it is extremely complicated, requiring a kind of dualism at the surfaces of objects. Again, in my view, projectivism is the most reasonable view on color.

[22] For mysterian reductionism, see Byrne and Tye (2006: 252) and Lycan (2006). This form of mysterianism must be distinguished from that of McGinn (1989), which is instead a view about a priori deducibility.

[23] See Byrne and Tye (2006: 253–4).

is a cow in his belief-box is that there is a cow there. In the case of Jill, optimal conditions do not obtain. She is viewing the cow from far away. But if they did, *that is a horse* would only be tokened in her belief-box if there really had been a horse there, so this is the content of this sentence in her belief-box. But this provides no reason to believe that the optimal cause relation satisfies the variation constraint in the coincidence cases I have presented. In the case of Jack and Jill, the optimality clause saves the day. Not so in coincidence cases. For in these cases the relevant individuals track the same properties *under optimal conditions*. This is so however the vague notion of optimal conditions is elaborated. But the most reasonable view is that in at least some of these cases the individuals involved bear the consciousness relation to different properties. So even if the optimal cause theory handles simple cases, there is absolutely no chance that a suitably elaborate version will handle coincidence cases.

So the chief argument for mysterian reductionism is that the rival view of primitivism requires danglers and faces problems with mental causation. But there is also an argument against it. As we have seen, all the physical relations we can think of fail to satisfy one or another of the four constraints (see Figure 2.3). So mysterian reductionism requires that the alleged macro-level physical relation which satisfies the constraints, and with which the consciousness relation is identical, is a relation which we cannot presently think of. And there is a problem with this view. Occasionally, mysterians about the mind–body link say that we cannot form concepts of certain hidden *micro-level* physical properties, the categorical bases of microphysical dispositions. This is not entirely implausible because there is a sense in which the categorical bases of microphysical dispositions are undetectable. But the mysterianism being contemplated now is implausible because *macro-level* physical relations are perfectly detectable. So the mysterian reductionist's claim that the consciousness relation is identical with a macro-level physical relation that we cannot think of is very implausible. What could prevent us from thinking of it?

In response, the defender of mysterian reductionism might attempt to provide an alternative explanation of why we cannot think of the alleged hidden physical relation which satisfies the four constraints. The explanation is provided by the fantastic complexity of the sensory systems, the fact that there are huge gaps in our knowledge of how sensible properties are represented in the brain, and of the selection pressures driving the evolution of sensory systems.[24]

Against this, no discovery of what happens in the brain will enable us to think of a physical relation between individuals and external properties that we could not think of before and with which the consciousness relation might be identified. Such discoveries may tell us a great detail about the details of the neural content-carriers; but they will not tell us anything about how these content-carriers get their contents. Consider an analogy: no amount of studying

[24] This is almost a direct quote from Byrne and Tye (2006: 252).

of the shapes of Chinese characters will enable one to discover what makes it the case that those characters carry the meanings they do among Chinese speakers. As for discoveries of the selection pressures driving evolution, it is impossible to see how they might reveal anything relevant here. In coincidence cases, the relevant differences are adaptations to different selection pressures, so that the sensory systems of the relevant individuals, although different, operate in accordance with design. Surely such cases are possible. Nothing we could discover about our actual evolutionary history could cast doubt on the claim that such cases are possible. I conclude that mysterian reductionism must be rejected.

Compromise reductionism is more concessive than mysterian reductionism. The compromise reductionist grants an inconsistency between the four constraints and reductionism about the consciousness relation. There is, on this view, no unknown physical relation that satisfies all the constraints. Since the extensionality constraint and the scrutability constraint are non-negotiable, this means we must choose between the relationality constraint, the variation constraint, and reductionism. But, whereas I hold that the most reasonable course is to keep the relational view and coincidental variation and give up reductionism, the compromise reductionist holds that the most reasonable course is to keep reductionism and give up the relational view or coincidental variation. For instance, he or she might keep reductionism and give up the relational view. Then the compromise reductionist would have no problem with coincidental variation. In particular, he or she might say that experiences are necessarily identical with internal brain states, which differ between the individuals in coincidence cases. Or he or she might keep reductionism, and reject coincidental variation rather than the relational view. In particular, the compromise reductionist might say that the consciousness relation is an A-type relation such as the optimal cause relation that is held constant in coincidence cases. This would entail that coincidental variation is false. In every possible coincidence case, on this view, the individuals involved bear the consciousness relation to exactly the same properties and have exactly the same experiences, in spite of the radical neural and functional differences between them.

But we cannot make sense of the phenomenological, semantic, and epistemic facts about sensory consciousness unless we accept the relational view. And I cannot bring myself to deny coincidental variation. Imagine meeting Yuck and Yum, Mild and Severe, or Mabel and Maxwell. To say that they have the same experiences in spite of all the evidence against this would be unreasonable. So the case for combining the relational view and coincidental variation is overwhelming.[25] By contrast, reductionism is an extremely speculative metaphysical

[25] This is one problem with combining a relational view such as intentionalism with an A-type reductive theory of the consciousness relation such as the optimal cause theory. Bad external correlation creates another, independent, problem that does not involve hypothetical coincidence cases. Given bad external correlation, a person might judge that one of his or her pains is twice as

claim. The chief argument for it is that it avoids danglers, providing a pleasingly simple view of the world. But it would be dogmatic to suppose that our world *must be* simple in this respect. So given the conflict between the relational view, coincidental variation and reductionism, I believe that the reasonable course is to keep the relational view and coincidental variation and to reject reductionism.

12. CONCLUDING REMARK

Primitivism does not automatically lead to the rejection of physicalism—at least if physicalism is a mere thesis of supervenience. G. E. Moore held that goodness is primitive yet supervenient on the natural as a matter of metaphysical necessity. Likewise, one could hold that the consciousness relation is primitive yet supervenient on the physical with metaphysical necessity. On this view, Zombies are impossible. This would yield what we might call *primitivist physicalism*. On this view, the consciousness relation is not a physical relation in the sense introduced at the beginning of section 3. It is not a complex relation constructible from the fundamental physical and functional relations of the world. It is an *extra* relation. But if primitivist physicalism is true, then the consciousness relation qualifies as physical in a broader sense because on this view it supervenes with metaphysical necessity on the physical way the world is. Alternatively, once one accepts primitivism, one could hold that the primitive consciousness relation supervenes on the physical with only nomological necessity. On this view, Zombies are possible. This would yield *property dualism*. Ontologically, primitivist physicalism and property dualism are identical, since both admit that the consciousness relation is an extra element of the world. They differ only modally. Which of these views should the primitivist adopt?

Some hold that reductionism about manifest properties fails in general in the sense that manifest properties cannot be identified with hugely complex properties built up from the fundamental physical and functional properties of the world. As noted, Moore held that reductionism fails in the case of the property of being good. And some would say that it fails even in the case of such

great as a second pain, even though the bodily disturbance the person bears the optimal cause relation to in having the first pain is *less than* twice as great as the bodily disturbance he or she bears the optimal cause relation to in having the second pain. Even under optimal conditions, there is *response expansion* (section 2). On a relational view such as intentionalism, truths about phenomenology are truths about content. So this combination of views runs the risk of entailing that John's introspective judgment about the phenomenal relationship among his pains is *false*. Since there is bad external correlation in general, the problem is general. For instance, such a combination of views also runs the risk of entailing that our introspective judgments about the resemblances among our color experiences and their unitary-binary structure are false. For a reply to this problem in the case of the unitary-binary character of color experience that involves complicated non-linear functions, see Tye and Bradley (2001). They do not explain how their reply applies to judgments about *resemblances* among color experiences or colors; nor do they address the problem as it arises for pain and taste.

unexciting properties as the property of being a mountain. But they still believe that there is an argument for believing that they supervene (with metaphysical necessity). Likewise, it might be said that, even if sensory consciousness fails to reduce, there is an argument for believing that it supervenes.

But this is a mistake. I do not think that the property of being good or the property of being a mountain fails to reduce. But if even we knew that they fail to reduce, we would have an a priori justification for believing that they supervene. It is inconceivable that a world that is a physical or natural duplicate of our world should differ from our world with respect to pattern of instantiation of the property of being a mountain or the property of being good. But consciousness is an exception. In the case of consciousness, we lack such a priori justification for supervenience.

In fact, I believe that reflection reveals that in the case of consciousness the only possible argument for supervenience proceeds by way of reduction. In slogan form: no justification without reduction. The argument from simplicity (avoiding danglers) and the causal exclusion argument might give us reason to accept reductionism about consciousness in a broad sense that includes reduction to functional properties. And reductionism entails supervenience. These arguments do not support supervenience independently of reductionism; they do not support primitivist physicalism. For, since primitivism is like property dualism, it faces the same problems about danglers and mental causation, as we shall see.

Now I have argued that reductionism about sensory consciousness fails. Even if one rejects my argument, one must at least admit that we are not overall justified in accepting reductionism. At the very least, we should suspend judgment. So if the only argument for supervenience proceeds by way of reductionism, we are left without any argument for accepting supervenience in the crucial case of consciousness. So we are left without any argument for even a minimal form of physicalism. We are also left without an argument for accepting what we might call *mysterian primitivist physicalism,* which has recently been defended by some philosophers.[26] This view combines primitivist physicalism with the claim that the supervenience of consciousness on the physical is not a priori to us now but would be a priori if only we knew more about the physical world. (Of course, this view must be distinguished from mysterian *reductionism* discussed in section 11.) Once we accept primitivism, there is no argument for accepting this view because there is no argument for accepting supervenience in the first place.[27]

[26] McGinn (1996) defends primitivist physicalism, and McGinn (1989) defends mysterianism.

[27] One might think that I have overlooked an argument: all other properties and relations of the manifest image supervene with metaphysical necessity, so we have inductive reason to think that the consciousness relation supervenes as well—even if it fails to reduce. This argument fails for two reasons. First, it is not clear that all other properties and relations of the manifest image supervene with metaphysical necessity. Consider sensible properties like color, sound, and taste: the gap between the ostensible sensible properties of external objects and physical properties is

In fact, I think we can say something stronger. Once we accept primitivism, there are two reasons for preferring property dualism over primitivist physicalism. First, we have the intuition that Zombies are possible. On any view, this provides *some* evidence that supervenience fails. Now, typically, defenders of supervenience respond that we have countervailing reasons to accept supervenience and to doubt this intuition. But, as we have seen, once we accept primitivism, we have no countervailing reasons to accept supervenience. So we no longer have any reason to doubt this intuition. Second, there is the Humean dictum against necessary connections between wholly distinct existences. Perhaps there are counterexamples to this dictum (for instance, being red seems to necessitate being extended), but one might think that in those cases where we have no reason to believe that distinct existences are necessarily connected we are justified in believing that the connection is only contingent.

But I should say that I do not find the modal issue between primitivist physicalism and property dualism very interesting, because these views are very similar and face the same problems. Property dualism requires nomological danglers: fundamental laws that dangle from the rest of the body of nomological truths. Primitivist physicalism requires modal danglers: necessary connections between wholly distinct properties that dangle from the rest of the body of modal truths. So the views seem on a par with respect to complexity. And, unlike some reductionist views, both views face the dilemma between overdetermination and epiphenomenalism. Of course, there are proposals on how to dodge this dilemma, but they seem available to the property dualist as well as the primitivist physicalist.[28]

In my view, the real interesting issue is the one that divides reductionism and primitivism. The views provide radically different pictures of our world. Here I have argued for the relational view and coincidental variation, and I have argued that these claims lead to primitivism.[29]

just as wide as the gap between consciousness and physical properties. Second, the properties of the manifest image that supervene also arguably reduce. So once we accept primitivism about the consciousness relation, we are admitting that it is very different from other properties and relations of the manifest image, and this considerably weakens the inductive inference.

[28] See Bealer (2007).

[29] Earlier versions of this chapter were presented at the New York University Friday Forum in 2002; at the universities of Michigan, Iowa, Texas, Arizona, Massachusetts, and Colorado in 2004; and at the inauguration of the Centre for Consciousness at the Australian National University in 2004. Thanks to the audiences on those occasions. I would especially like to thank David Barnett, George Bealer, Anna Bjurman-Pautz, Ned Block, David Chalmers, Rob Koons, Stephen Schiffer, and Michael Tye for comments and other help.