

3

The Introspectibility of Brain States as Such

PETE MANDIK

Paul Churchland has defended various bold theses throughout his career. Of particular interest to the current chapter is what I shall call Churchland's *Introspection Thesis*.

A person with sufficient neuroscientific education can introspect his or her brain states *as* brain states.¹

Is the Introspection Thesis true? It certainly isn't obvious. Introspection is the faculty by which each of us has access to his or her own mental states. Even if we were to suppose that mental states are identical to brain states, it doesn't follow immediately from this supposition that we can introspect our mental states *as* brain states. This point is analogous to the following. It doesn't follow immediately from the mere fact that some distant object is identical to a horse that we can perceive it *as* a horse. Further, it isn't obvious that any amount of education would suffice to make some distant speck on the horizon *seem* like a horse. It may very well be the case that no matter how well we *know* that some distant speck is a horse; as long as we are sufficiently distant from it we will only be able to *see it as* a speck. Analogously then, it may very well be the case that no matter how well we *know* that our mental states are brain states, we will only be able to *introspect them as* irreducibly mental.

Not only is the introspection thesis not obviously true, it is not obvious what it would be like for it to be true. We can easily imagine seeing a horse as a horse. Can we similarly imagine introspecting brain states as brain states? I think, indeed, we can. Though I think the case will be even clearer once we review Churchland's arguments, it will be useful, at this early stage, to get a sketch of what it is we are supposed to imagine. It helps to begin by noting the distinction between something that can be perceived and something that can be figured out based on what is perceived. If I stick my finger into a hot cup of coffee, I can perceive the heat of the coffee. If, without touching the coffee, I see the steam rising from it, it is the steam that I see and based on what I perceive, I figure out the approximate temperature

of the coffee. As will be elaborated below, for Churchland the crucial distinction here depends not on the degree to which theoretical knowledge is involved but on how *automatic* its application in perception is. If, on having a certain sensation, I come to *automatically* apply the concept of heat to the cause of the sensation, then, for Churchland, that counts as perceiving heat. If I learn, then, to apply the concept of heat automatically (that is, without going through some intermediary inference) to coffee on the sight of steam rising from it, then what I've learned to do is see the heat of the coffee. I may perceive the coffee as being hot even in situations in which I see it without feeling it. Churchland's defense of the Introspection Thesis depends on an analogous view of introspection. Suppose that, in addition to being able to apply the concept of heat to an external object as an automatic reaction to some sensation, I learn the concept that applies to the neural basis of that sensation. Or, to pick a different kind of sensation, suppose the neural basis of motion perception involves activity in area V5 of cerebral cortex and that I learn to apply the concept of activity in V5 as an automatic response to a sensation of motion. Under such conditions, then, I would be introspecting my brain states *as* brainstates and in this case I would be introspecting the sensation of motion as a pattern of activity in area V5 of cerebral cortex.

Why care whether the Introspection Thesis is true? Churchland cares about the Introspection Thesis because it provides him a defense of his favored brand of materialism against attacks by anti-materialists who would base their claims on introspection. Churchland is concerned to show, then, by showing how introspection itself can reveal that mental states are brain states, that introspection does not provide an unassailable refuge for the antimaterialist. Further interest in the Introspection Thesis is that it has deep implications for current work on consciousness and it impacts debates not only between materialists and antimaterialists but also debates among materialists. In particular, it arguably undermines *representationalism*, an approach to consciousness that has many materialist adherents.² Although this will be unpacked further later, to a first approximation representationalism is the view that there is nothing more to qualia – the phenomenal characteristics of conscious experience – than representational content. The tension between representationalism and the Introspection Thesis becomes apparent once we consider a thesis oft associated with representationalism: the *Transparency Thesis*.

When a person introspects his or her own conscious mental states he or she only has access to the properties those states represent objects as having.

To spell out the Transparency Thesis in terms of an example, it is the claim that when I have a conscious experience of a blue square, my introspective access to my experience only puts me in touch with features represented as instantiated in the environment – blueness and squareness. I have no introspective access, then, to features of the experience itself. In direct opposition to Churchland's Introspection Thesis, then, I can have no introspective access to any of the stuff going on in my brain when I have a conscious experience of a blue square on the wall. The metaphor of transparency is appropriate here insofar as when I examine my experiences I inevitably "look through" them to an external world of objects and properties that the experiences represent. The Transparency Thesis is oft appealed to as a premise in arguments for representationalism.³ Further, representationalism is arguably true only if the Transparency Thesis is true. As Kind (2003) puts the point, if we are able to introspect aspects of experiences other than their representational contents, then properties other than representational contents of experience figure into the phenomenal character of experience.⁴

My goal in this chapter is to adjudicate between the competing theses of Transparency and Introspection, arguing ultimately that Transparency is the weaker of the two. I discuss further what the prospects are for representationalism without the Transparency Thesis. The organization of the rest of this chapter is as follows. First I spell out Churchland's arguments for the Introspection Thesis. Next I spell out the Transparency Thesis and the related notions of qualia, consciousness, and representationalism. Finally I discuss the degree to which the tension between Transparency and Introspection is a problem for Churchland and what resources his larger body of work makes available to resolve it.

CHURCHLANDISH INTROSPECTION

Churchland's argument for the Introspection Thesis depends on a particular view of perception and an analogy between perception and introspection. The view of perception at play here is that "perception consists in the conceptual exploitation of the natural information contained in our sensations or sensory states." (Churchland 1979:7). Analogously then, introspection is the conceptual exploitation of natural information that our sensations or sensory states contain about themselves. Fleshing out Churchland's views of perception and introspection requires us to flesh out what Churchland thinks the conceptual exploitation of natural information is. Crucial here

is a distinction Churchland draws between two kinds intentionality that sensations can have, that is, two ways in which a sensation can be a sensation *of* φ . A sensation can have “objective intentionality” as well as “subjective intentionality” and Churchland adopts the typographical convention of subscripts to distinguish “sensation of_o φ ” from “sensation of_s φ ”. Spelling out the distinction semiformally, Churchland provides:

Objective intentionality:

A given (kind of) sensation is a sensation of_o φ with respect to a being x if and only if

under normal conditions, sensations of that kind occur in x only if something in x 's perceptual environment is indeed φ .

Subjective intentionality:

A given (kind of) sensation is a sensation of_s φ with respect to a being x if and only if

under normal conditions, x 's characteristic non-inferential response to any sensation of that kind is some judgment to the effect that something or other is φ . (ibid, p. 14)

The objective intentionality of sensations is the information that sensations actually carry about the environment regardless of whether we exploit that information. The objective intentionality of sensations determines what it is that we are *capable* of perceiving. What we actually *do* perceive depends on subjective intentionality. That is, what we actually do perceive depends on what concepts we bring to bear in the judgments that our sensations noninferentially elicit. So, for example, whether I am capable of seeing the tiny insect on the far side of the room depends on whether I have states of my visual system that reliably co-vary with the presence of that object, and if my eyesight is insufficiently acute, I will lack such states. Whether I actually do perceive that object depends on more than just good eyesight. It depends on whether I actually do employ my conceptual resources to interpret my visual sensations as indicating the presence of an insect. Thus, enriching our conceptual repertoire allows us to better exploit, in perception, the information already contained in sensation (ibid: 16). For example, with sufficient education, we can move beyond the coarse-grained common-sense temperature concepts in virtue of which we feel things as ‘hot,’ ‘warm,’ and ‘cold’ and instead exploit scientific concepts in order to feel “that the mean kinetic energy of the atmospheric molecules in this room is roughly $6.2 \times 10^{-21} \text{ kg m}^2/\text{s}^2$ ” (ibid: 26). Multiplying examples, Churchland offers

that with sufficient conceptual augmentation we can hear “the occurrence and properties of compression wave trains in the atmosphere – most obviously of both their wavelength (from 15 m to 15 mm) and their frequency (from 20 to 20,000 cycles per second)” (ibid: 26) and we can see “the dominant wavelength (and/or frequency) of incoming electromagnetic radiation in the range $0.38\text{--}0.72 \times 10^{-16}$ m, and of the reflective, absorptive, and radiative properties of the molecular aggregates from which it comes” (ibid: 27). Our sensory states already carry this information and it is thus there waiting to be picked up by a suitably theoretically informed set of concepts.

Human perceivers are importantly analogous to measuring instruments, according to Churchland. Both have states that serve as reliable indicators of certain aspects of the environment. Further, in both cases reliable indication relies on interpretation functions that map distinct states onto distinct propositions (ibid: 38). In the case of measuring instruments, the interpretation function is determined when we calibrate the measuring instrument to map, for instance, the needle positions on an ammeter “onto distinct propositions such as ‘there is a 5 ampere (A) current flowing in the circuit’” (ibid: 38). In the case of the conceptual exploitation of sensory information, while Churchland acknowledges that we do not explicitly and consciously use an interpretation function to formulate our perceptual judgments, he nonetheless points out that,

insofar as our conceptual responses to our sensations do display determinate and identifiable patterns...we embody or model a set of interpretation functions...implanted in childhood as we learned to think and talk about the world... [and that] are just as properly subjects for evaluation, criticism, and possible replacement as are interpretation functions in any other context.” (ibid: 39; emphasis in original)

With the above view of perception in hand, Churchland goes on to spell out what introspection would amount to. Focusing on introspective judgments about sensory states “e.g. ‘I have a visual sensation of an orange circle’” (ibid: 40), Churchland describes introspection as involving “a temporary disengagement from the interpretation functions that normally govern our conceptual responses, and the engagement instead of an interpretation function that maps (what we now conceive as) sensations, etc., onto judgments *about* sensations, etc.” (ibid: 40; emphasis in original). One consequence of this view of introspection, important both to Churchland and for points I’ll raise subsequently, is that introspective judgments are no more likely to be incorrigible or infallible than perceptual judgments more

generally. Churchland illustrates by continuing the analogy to measuring instruments:

[C]onsider an ammeter with a graduated dial marked ‘5 A’, ‘10 A’, and so on. Suppose it [is] constructed so that at the flick of a switch it flips another dial into place behind the needle, a dial marked ‘0.01 gauss’, ‘0.02 gauss’, and so on. This second dial is so calibrated that the needle positions on the dial now *overtly* reflect the simultaneous strength of the variable magnetic field inside the instrument, the very field whose action moves the spring-loaded needle. Our ammeter is now operating in “introspective mode”. (ibid: 40)

A measuring instrument not only has states that carry information about its immediate environment, its states carry information about themselves and a calibration of the instrument can just as easily latch on to the one kind of information as the other. To use an example perhaps more accessible than those Churchland provides, the height of the column of mercury in a thermometer not only carries information about the temperature of the surrounding medium, but also information about how high the mercury is. We could put a mercury thermometer in “introspective mode,” then, by changing the marks on it from measurements of degrees in Celsius to measurements of height in millimeters. And again, there is no guarantee of accuracy, for the calibration scheme may very well say that the current height is 3 mm when in reality it is 3.5 mm. However, when the device is correctly calibrated, what indicates that the height is 3 mm is when the 3 mm mark is even with the top of the mercury column that is, in fact, 3 mm in height.

The Churchlandish introspection of brain states involves exploiting the information that a state of the nervous system carries about itself. Churchland offers possible examples of what this neurophysiologically informed introspection would be like. His remarks on these possibilities are worth quoting at length for they simultaneously serve to bolster the plausibility of the Introspection Thesis and cast doubt on the Transparency Thesis.

The considerable variety of states currently apprehended in a lump under ‘pain’, for example, can be more discriminately recognized as sundry modes of stimulation in our A-delta fibres and/or C-fibres (peripherally), or in our thalamus and/or reticular formation (centrally). What are commonly grasped as “after images” can be more penetratingly grasped as differentially fatigued areas in the retina’s photochemical grid, and the chemical behaviour of such areas over time – specifically, their resynthesis of rhodospin (black/white) and the iodopsins (sundry colours) – is readily followed

by suitably informed introspection. The familiar “phosphenes” can be recognized as spontaneous electrical activity in the visual nervous system. Sensations of acceleration, and of falling, are better grasped as deformations and relaxations of one’s vestibular maculae, the tiny jello-like linear accelerometers in the vestibular system. Rotational “dizziness” is more perspicuously introspected as a residual circulation of the inertial fluid in the semicircular canals of the inner ear, and the increase and decrease of that relative motion is readily monitored. The familiar “pins and needles” at a given site is more usefully apprehended as oxygen deprivation of the nerve endings there located. (ibid: 118–119)

Before moving on to consider how Churchland’s Introspection Thesis bears on discussions of the alleged transparency of conscious experience, it will be useful to summarize the key points from the above discussion and relate them to examples that are perhaps a bit easier to get an intuitive grasp of than Churchland’s examples of, say, perceiving red light as electromagnetic radiation in the range $0.38\text{--}0.72 \times 10^{-16}$ m. The crucial aspects of Churchland’s account of perception are those that allow for the reconstruction of the distinction between what is perceived without inference and what is inferred but not perceived. Let us consider the following situation to illustrate this distinction. Two friends, George and John, are lunching in a well-lighted location when, as part of some publicity stunt, a man in a realistic gorilla suit runs through the area. Suppose that both gorilla suit and gorilla act are quite realistic and convincing to the untrained eye. George, being a special effects expert for the film industry, is not fooled and can see quite clearly that this is indeed a man in a costume. John, however, is a novice and cannot help but be fooled: he sees this as a genuine gorilla, perhaps escaped from the nearby zoo. In fact, John the novice continues to see this individual as a genuine gorilla even after George the expert assures him that it is in fact a suited man. John may even come to believe George’s testimony for he trusts George’s expertise, but John cannot shake the impression that it is a real gorilla that is causing a ruckus in the restaurant. There are several key similarities and differences between John and George and Churchland’s account of perception helps to explain these similarities and differences. The first similarity is that there is a sense in which both John and George see the same thing. The first difference is that only George sees that thing *as* a man in a suit. The second similarity is that they both know that it is a man in a suit. The second difference is that in spite of his knowledge, John is incapable of seeing it as a man in a suit. The explanation of the first similarity is that John and George both have visual sensations with the same objective intentionality. They both have states of

their visual system that causally co-vary with, for example, the distinctive way that a man in a gorilla suit moves. The explanation of the first difference is that only George is able to automatically (without an intervening inference) apply the concept of a man in a gorilla suit to the thing causing his current visual sensation and thus only George's sensations have the subjective intentionality indicating the presence of a man in a gorilla suit. The explanation of the second similarity depends on nothing peculiar to Churchland: they both know that the thing is a man in a gorilla suit because they have justified true beliefs that it is a man in a gorilla suit. The explanation of the second difference is that, unlike George, John is incapable of automatically (without an intervening inference) applying the concept of a man in a gorilla suit to the thing causing his current visual sensation, and thus John's sensations lack the subjective intentionality indicating the presence of a man in a gorilla suit.

Let us briefly reconsider the example discussed at the beginning of this chapter, namely the distant horse that looks like a speck on the horizon. If the distant speck is indeed a horse and someone were incapable of automatically applying the concept of a horse to the cause of their visual sensation, then even if they knew it was a horse they would be incapable of seeing it *as* a horse. In contrast, if they were able to automatically apply the concept of a horse to the cause of their visual sensation, then they would be seeing the distant speck as a horse: the distant speck would seem like a horse to that person.

The appropriate analogy, then, to introspection would be the following. If a person knew that their mental states were identical to brain states, but was incapable of *automatically* applying the concept of a brain state to a mental state, then in spite of their knowledge they would be incapable of introspecting their brain states *as* brain states. In contrast, if they were able to automatically apply the concept of a brain state to their brain states then they would be introspecting their brain states as such: their brain states would seem like brain states to them.

Thus concludes my initial discussion of Churchland's Introspection Thesis and his defense of it. I turn next to unpack the Transparency Thesis and the opposition between it and Churchland's account of introspection.

THE ALLEGED TRANSPARENCY OF CONSCIOUS EXPERIENCE

Contemporary discussions of the notion that experience is transparent (or diaphanous) frequently trace the idea back to the following G. E. Moore quotation.

[T]he moment we try to fix our attention upon consciousness and to see what, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous. (Moore 1903: 25)

However, for the sake of historical accuracy (at least), it is worth noting that while Moore discusses the Transparency Thesis, he does not actually endorse it. Transparency is introduced in the contemporary literature (and endorsed) by Harman:

When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience. Nor does she experience any features of anything as intrinsic features of her experiences. And that is true of you too. There is nothing special about Eloise's visual experience. When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree... (Harman 1990: 667)

Harman's interest is in a defense of functionalism (wherein mental states are type identified in terms of their causal relations, not, *pace* typical qualiaphiles, in terms of their intrinsic properties). Along the way, he defends a kind of representationalism: the objects of experience are intentional objects. Other adherents of the Transparency Thesis who utilize it in the defense of representationalism include Tye (1995, 2000) and Dretske (1995).

Although the metaphor of transparency is visual and thus most appropriate for visual experiences, defenders of the Transparency Thesis intend it to generalize to all conscious experience. So, for example, as Dretske writes, "If one is asked to introspect one's current gustatory experience... one finds oneself attending, not to one's experience of the wine, but to the wine itself (or perhaps the tongue or palette)" (1995: 62).

We can get a further understanding of what is being affirmed and denied by the Transparency Thesis by seeing how disagreement over it divides various approaches to understanding consciousness. Advocacy of Transparency (frequently) goes hand-in-hand with First-Order Representationalism and goes against Higher-Order Representationalism. Roughly, First-Order Representationalism explains consciousness in terms of mental representations of aspects of the environment. Thus, according to

First-Order Representationalists, meta-representational states are strictly irrelevant to phenomenal consciousness. As Tye puts the point, “Cognitive awareness of our own feelings itself feels no special way at all. Phenomenal character attaches to experiences and feelings (including images), and not, I maintain, to our cognitive responses to them” (Tye 2000: 36–37). Dretske states his agreement regarding the irrelevance of meta-representational states for phenomenal consciousness as follows:

Conscious mental states – experiences, in particular – are states that we are conscious *with*, not states we are conscious *of*. They are states that make us conscious, not states that we make conscious by being conscious of them. They are states that enable us to see, hear, and feel, not states that we see, hear, or feel. (Dretske 1995: 100–101)

According to First-Order Representationalism, to have a conscious experience of a blue square on the wall it suffices to have a (certain kind) of mental representation of a blue square on the wall. What kind of mental representation will suffice to give rise to consciousness is something that various First-Order Representationalists may disagree on. But in spite of their differences they agree that the mental representation in question need not itself be represented by any other mental representation in order to give rise to a conscious state.

Higher-Order Representationalism, in contrast, explains consciousness in terms of mental representations of other mental states. A key principle appealed to by Higher-Order Representationalists is the Transitivity Principle.

In order to have a conscious mental state, one must be conscious *of* that mental state.

Thus, according to advocates of the Transitivity Principle such as Lycan (2001) and Rosenthal (2002), if one has a conscious experience of a blue square, it is insufficient to simply have a mental state that represents a blue square – having only a mental representation of a blue square would be having only an unconscious mental representation of a blue square. One must additionally have a mental representation of the mental representation of the blue square, that is, a second mental representation which represents the first representation, which, in turn, represents the blue square.⁵

The Transitivity Principle gets its name from the fact that consciousness in the intransitive sense of the term (e.g., “Mary’s experience was conscious”) is being explained by consciousness in the transitive sense of the term (e.g., “Mary was conscious of her experience”). The English word “conscious”

has several uses in the construction of verb phrases, some of which yield transitive verb phrases (e.g., “John was conscious of the smell of coffee”) and some of which yield intransitive verb phrases (e.g., “John was conscious” and “John’s desire was conscious”).

The tension between transparency and transitivity becomes apparent when we note that the higher-order representations must represent aspects of the first-order states themselves. If so-called higher-order states simply had the same contents as their first-order targets, then they wouldn’t really be higher-order after all. What makes a mental representation first-order is that it isn’t meta-representational – it doesn’t represent itself or any other mental representations but instead represents, for example, aspects of the creature’s environment or body. If the so-called higher-order state didn’t represent aspects of the first-order state itself, but instead represented what the first-order state represents, then the so-called higher-order state would be representing, for example, aspects of the creature’s environment or body and would thus itself be a first-order state.

Another way of putting the previous point is in terms of a distinction between representational content and representational vehicle. I may have, at 3 P.M., a memory of something that happened at 2 P.M. – I may remember that at 2 P.M. someone told me a particularly funny joke. Occurring at 3 P.M. is a property of the representational vehicle, it is a property of the memory itself. Occurring at 2 P.M. is a property of the content – it is a property of what was remembered, namely, that a funny joke was told. A second class of examples of the vehicular properties of a mental representation includes the neurophysiological properties of a mental representation. The neurophysiological properties of a first-order representation are typically vehicular properties of that representation. For example, the pattern of neural activation that constitutes my perception of a green bottle three feet away from me is neither green, a bottle, nor three feet away from me. With the distinction between content and vehicle thus in hand, the main point here is as follows. If a representation doesn’t represent any of the vehicular properties of some other representation, but simply has similar contents to the second representation, then the first representation isn’t a representation of the second representation, and thus isn’t a higher-order representation.

The vocabulary of “content” and “vehicle” allows us to formulate the opposing theses of transparency and transitivity as follows. According to the Transparency Thesis favored by First-Order Representationalists, when we have a conscious first-order representation, all we can be conscious of are the contents of that representation, we are thus incapable of becoming

conscious of any of the vehicular properties of that representation. According to the Transitivity Principle, then, when we have a conscious first-order representation we must be conscious of (among other things) vehicular properties of that representation.

These two theses, while exclusive, are not exhaustive. The middle ground left open by merely denying transparency without necessarily affirming transitivity is where we find Churchland's Introspection Thesis, since Churchland's thesis entails that we *can* be conscious of vehicular properties of our first-order representations but does not entail that we *must* be conscious of vehicular properties of our first-order representations. Churchland's thesis states that a suitably educated individual can become aware of their own brain states as such and as I have argued above this means that a suitably educated individual can become aware of the vehicular properties of their first-order representations. It does not, however, entail that everyone who has conscious states must be aware of the vehicular properties of their first-order representations because it leaves open, as it should, that perhaps not everyone is suitably educated in the relevant neuroscience. The Transparency Thesis states that when one has a conscious state one cannot be conscious of the state itself, and as I have argued this entails that one cannot be conscious of the vehicular properties of the state. The Transitivity Principle states that one can have a conscious mental state only if one is conscious of that state, and as I have argued this entails that one must be conscious of the vehicular properties of the state. Thus does Churchland's Introspection thesis occupy a middle ground between Transparency and Transitivity. Transparency entails that you cannot be aware of vehicular properties of conscious states; Transitivity entails that you must be aware of vehicular properties of conscious states; and Churchland's thesis entails that you can, (but don't have to) be aware of vehicular properties of conscious states.

Not only does Churchland's Introspection Thesis conflict with the Transparency Thesis, but it threatens the larger project of representationalism. As Amy Kind (2003) has argued, if the Transparency Thesis is false then representationalism itself is false. If we can have introspective access to conscious states themselves and not just their representational contents, then there must be more to the phenomenal character of a conscious state than its representational contents. To be clear, the representationalism impugned by the falsity of the Transparency Thesis is First-Order representationalism. If we have introspective access to more than the contents of a first-order representation, then there is more to the character of consciousness than those contents. Of course, the possibility remains that the character

of consciousness is still fully determined by representational content, but if transparency turns out to be false, the content in question would include the content of higher-order representations.

It is instructive to see what the transparency thesis looks like when stated in Churchland's vocabulary. It becomes the thesis that while the objective intentionality of a sensation may include information about both itself and states external to it, the subjective intentionality of a sensation is limited to states external to the sensation. The interpretation functions imposed by the conceptual exploitation of sensations may map sensations onto states external to them but can not possibly map sensations onto themselves. Spelling this out further in terms of the analogy to measuring instruments, the claim of the Transparency Thesis becomes tantamount to claiming that while it is possible to calibrate a thermometer so that mercury column heights indicate temperatures, it is impossible to change the marks on the thermometer so that the mercury column heights indicate mercury column heights. That such a reinterpretation of our brain states should be absolutely impossible seems implausible. The implausibility is further heightened when we consider that the Transparency Thesis is supposed to be introspectively and/or pre-theoretically obvious. That something like Churchlandish introspection is impossible seems an odd candidate for something that we would have introspective or pre-theoretic access to.

Once we have the Transparency Thesis stated in a Churchlandish vocabulary, it is apparent that it is less plausible than Churchland's Introspection Thesis. Once we grant Churchland's general view of perception and introspection, namely, that both involve a procedure for mapping sensations onto judgments, it follows quite naturally that, contra the transparency thesis, it would be possible for a suitably educated person to introspect his or her own brain states as brain states. That is, once we grant that sensations carry information about lots of things including themselves, and that perception involves interpreting sensations in ways so that we conceptually exploit the information already contained in the sensations, then there is no reason for it to be impossible to interpret sensations in ways so that we conceptually exploit the information that sensations carry about themselves.

Given the dependence of the introspection thesis on Churchland's views concerning perception and introspection, a natural move for the friend of transparency would seem to be to question such views of perception and introspection. However, it is not clear that such a move would actually be available to the current defenders of transparency. For example, Tye would seem to be hard pressed to deny such views since they seem very close to

his own. Consider, for example, Tye's description of the introspection of our own thoughts and experiences:

[I]ntrospection of thought contents is a reliable process that takes as input the content of the thought and delivers as output a belief or judgment that one is undergoing a state with that content... We acquire introspective knowledge of what it is like to have such-and-such an experience or feeling via a reliable process that triggers the application of a suitable phenomenal concept or concepts. This reliable process... takes as input the direct awareness of external qualities (in the perceptual case)... (Tye 2000: 53)

The view that introspection involves a mapping process is common to both Churchland and Tye. Of course, whereas Churchland uses the language of "mapping" x 's onto y 's Tye instead speaks of processes that have x 's as inputs and y 's as outputs. However, I do not suppose that there is any difference between mapping and input-output processing, so whatever disagreement there must be between Churchland and Tye concerning introspection it must be a disagreement not about the relation involved, but instead about what the admissible relata are. And further, it seems that Churchland and Tye agree that the introspection of sensations would deliver as outputs judgments about sensations. So, whatever Tye could disagree about here would be limited to what the introspective judgments could be about. However, this disagreement simply is the disagreement between the Transparency Thesis and the Introspection Thesis. Therefore, a First-Order Representationalist such as Tye cannot object to Churchland's argument for the Introspection Thesis on grounds concerning the general nature of introspection, that is, whether it is a reliable process that yields judgments about sensory states.

Not only is Churchland's Introspection Thesis more plausible than the thesis of Transparency, but the premises upon which the Introspection Thesis is based can also be used to explain whatever initial plausibility the Transparency Thesis enjoys. The natural explanation that emerges is the following. Transparency is plausible because the mappings of sensations onto propositions that people typically acquire first are mappings that involve judgments about external world objects. Children learn to call objects blue, red, and so on way before (if ever) they learn that there are such things as blue sensations, red sensations, and so on. Further, this kind of mapping is relatively entrenched: it takes a bit of (philosophical?) sophistication for it to occur to any one to map things in any other way, that is, to map sensations onto judgments about sensations as opposed to judgments

about external world objects and their properties. Thus, transparency may seem plausible without being true.

CHURCHLAND IN TROUBLE?

The strength of the Introspection Thesis not only spells trouble for representationalists such as Tye and Dretske, but it may very well spell trouble for Churchland himself, since there is evidence from other parts of his corpus that he himself may be a representationalist. I will here briefly review the *prima facie* evidence, both pro and con, regarding whether Churchland is indeed a representationalist. I will argue that the cons will outweigh the pros: in spite of a few superficial appearances, Churchland is ultimately *not* a representationalist.

One general consideration that favors regarding Churchland as a representationalist is that he is, in general, quite sympathetic to representational (and computational) approaches to cognition and it would thus not be incongruous for Churchland to think that qualia were amenable to a representational/computational analysis. However, this is not the sole consideration that favors reading Churchland as a representationalist. One of the most striking pieces of evidence implicating Churchland's sympathy for representationalism comes from the article "Some Reductive Strategies in Cognitive Neurobiology." In a section entitled "The Representational Power of State Spaces", the bulk of the discussion is concerned with the topic of qualia. The relevant notion of state spaces and their neural implementation is conveyed in the following quotation

The global state of a complex system of n distinct variables can be economically represented by a single point in an abstract n -dimensional state space. And such a state-space point can be neurally implemented, in the simplest case, by a specific distribution of n spiking frequencies in a system of only n distinct fibres. Moreover, a state-space representation embodies the *metrical* relations between distinct possible positions within it, and thus embodies the representation of *similarity* relations between distinct items thus represented. (Churchland 1986: 102, emphases in original)

The upshot of the discussion that ensues can be conveyed in terms of one of Churchland's major examples: color. He endorses Land's view that the perceptual discriminability of reflectances by humans is due to the reception of three kinds of electromagnetic wavelength by three different kinds of cones in the retina. Further, he states that the degrees of perceived similarity

between colors are due to the degrees of proximity between the points in neural state space that represent those colors. Churchland is quite clear in his intent to identify color sensations with the neural representation of colors (the neural representation of spectral reflectance). Thus he endorses, or at least entertains,

... the hypothesis that a visual *sensation* of any specific color is literally identical with a specific triplet of spiking frequencies in some triune brain system. If this is true, then the similarity of two color sensations emerges as just the proximity of their respective state-space positions. And, of course, there are an indefinite number of continuous state-space paths connecting any two state-space points. Evidently, we can reconceive the cube [depicting the three dimensions of coding frequencies for reflectance in color state space] as an internal “qualia cube”. (ibid: 105)

Churchland’s endorsement of the identification of color sensations with the neural representation of color seems like a straightforward endorsement of First-Order Representationalism, at least as far as color is concerned. Further, the rest of his discussion in “The Representational Power of State Spaces” section of “Some Reductive Strategies in Cognitive Neurobiology” makes it quite clear that he intends the representational approach to generalize to other sensory qualia, since he goes on to discuss gustatory, olfactory, and auditory qualia (ibid: 105–106). I think however, that these remarks can ultimately be read as consistent with the falsity of both First-Order Representationalism and the Transparency Thesis.

When wondering whether Churchland’s identification of sensations with neural state space representations is in tension with the Introspection Thesis it is useful to note that in “Some Reductive Strategies...” he explicitly portrays the two views as compatible. He writes

The “ineffable” pink of one’s current visual sensation may be richly and precisely expressible as a 95Hz/80Hz/80Hz “chord” in the relevant triune cortical system. The “unconveyable” taste sensation produced by the fabled Australian health tonic Vegamite [*sic.*] might be quite poignantly conveyed as a 85/80/90/15 “chord” in one’s four-channeled gustatory system (a dark corner of taste-space that is best avoided). And the “indescribable” olfactory sensation produced by a newly opened rose might be quite accurately described as a 95/35/10/80/60/55 “chord” in some six dimensional system within one’s olfactory bulb.

This more penetrating conceptual framework might even displace the commonsense framework as the vehicle of intersubjective description and spontaneous introspection. Just as a musician can learn to recognize the

constitution of heard musical chords, after internalizing the general theory of their internal structure, so may we learn to recognize, introspectively, the n -dimensional constitution of our subjective sensory qualia, after having internalized the general theory of *their* internal structure. (ibid: 106)

Note, however, in the examples in the quotation that what are introspected may very well be conceived of as vehicular properties. The chords in multidimensional systems that he discusses are chords in multidimensional *neural* systems. Another way of putting the point of how Churchland's remarks about qualia and representations can be consistent with the falsity of First-Order Representationalism is by noting how the word "representation" can often pick out only a representational vehicle. That is, while sometimes "representation" picks out the process of x 's representing y , at other times "representation" can be used to pick out only x itself, the thing which represents as opposed to the act of representing. What Churchland is doing is identifying sensations with certain representations. However, this identification is consistent with the view that we have introspective access to aspects of the representations other than their representational contents.

I turn now to briefly consider two considerations in favor of considering Churchland as an antirepresentationalist (aside from whatever is implied by his endorsement of the Introspection Thesis). The first consideration is that it seems that Churchland thinks that the identity of a sensation is due to its intrinsic features and that whatever intentionality it has is due to extrinsic features. Churchland illustrates the point in terms of an extended thought experiment concerning beings who perceive temperatures, but in virtue of what in us would be the visual sensations we associate with light and dark (Churchland 1979: 8–14). In his summary of the crucial points of the thought experiment, Churchland notes the following:

It is clear [...] that neither the objective nor the subjective intentionality of a given kind of sensation is an intrinsic feature of that kind of sensation. Rather, they are both relational features, involving the sensation's typical causes in the former case, and its typical (conceptual) effects in the latter. And it is equally clear that both the "of_o-ness" and the "of_s-ness" of one and the same kind of sensation can vary from being to being, and even over time within the history of a single individual, the variation being a function of differences or changes in sensory apparatus in the case of objective intentionality, and of differences or changes in training and education in the case of subjective intentionality.

[...]

[T]he intrinsic qualitative identity of one's sensations is irrelevant to what properties one can or does perceive the world as displaying. (ibid: 15)

⋮

The tension between this quotation and any representationalist interpretation of the remarks on sensations as state-space representations should be relatively obvious, but in case it requires spelling out, the salient contrasts are two fold: the intrinsic/extrinsic contrast and the intentional/nonintentional contrast. The view expressed in "Some Reductive Strategies..." identifies sensations with points in a neural state space and points in a space are individuated extrinsically by their relation to all of the other points in that space. In contrast, the qualitative identity of sensations is characterized as intrinsic in the 1979 quote. Second, and most important, the view expressed in "Some Reductive Strategies..." identifies sensations with states having a certain kind of intentionality. In the case of color sensations, the sensations are identified with points in neural state-space that represent spectral reflectance. In contrast, the 1979 quote presents the intrinsic qualitative identity of a sensation as distinct from both its objective intentionality and its subjective intentionality.

A second consideration that casts some doubt on whether Churchland is committed to representationalism comes from an article co-authored with Patricia Churchland, "Functionalism, Qualia, and Intentionality." Of particular interest is a section of that article entitled "The Problem of Distinguishing States with Qualia from States Without" wherein the Churchlands affirm and attempt to explain how it is that a sensation has qualitative character whereas a belief does not. In brief, the Churchlands' explanation of the difference is that, in introspection, "Sensations are identified by way of their intrinsic properties; beliefs are identified by way of their highly abstract structural features" (Churchland and Churchland 1981: 33). According to the Churchlands, "the number of possible beliefs is at least a denumerable infinity" whereas the finite number of continua (and positions on them) that characterize distinct qualia "is sufficiently small" for noninferential discriminatory mechanisms to exploit the intrinsic qualities that define sensations (ibid: 32). The problem for discriminating beliefs, allegedly, is not that beliefs lack characteristic intrinsic properties, but instead that there are too many of them "for us to have any hope of being able to discriminate and identify all of them on such a one-by-one basis" (ibid: 32). Ultimately, the suggestion here seems to be that what makes a sensation a state with qualia is not merely that it has intrinsic qualities (because beliefs have those

too), but that it may be introspectively identified in terms of its intrinsic qualities.⁶ While the Churchlands do not explicitly address the question of whether sensations are to be distinguished from beliefs in terms of sensations lacking intentionality, the point that emerges of relevance to Paul Churchland's antirepresentationalism is that if a sensation can be introspectively identified in terms of its intrinsic qualities, this entails that it need not be introspectively identified in terms of its intentionality, since whatever intentionality it has would be extrinsic to it. This is in strict contrast to the view of the introspection of sensations advocated by representationalism, especially adherents of the Transparency Thesis, since on the latter view one may only introspect the representational contents of sensations, and *not* any of their intrinsic properties.

It is worth briefly noting an addendum that closes this section in the version anthologized in Paul Churchland's (1989) *A Neurocomputational Perspective*. In it Paul Churchland writes

I must now express a loss of confidence in this argument. The problem is that sensations now appear to be decidedly more various than I had originally estimated and to have a much more intricate combinatorial structure than I had earlier supposed (see [the "The Representational Power of State Spaces" section of "Some Reductive Strategies in Cognitive Neurobiology"]). Accordingly, the contrasts on which the preceding argument places so much weight now appear spurious: what seemed a large difference in kind now seems a mere difference in degree. (Churchland 1989: 33)

Whatever, precisely, Churchland is abandoning in this addendum, this much remains clear: Whatever it is that is being trumped, the considerations that are doing the trumping come from the "The Representational Power of State Spaces" section of "Some Reductive Strategies in Cognitive Neurobiology," and as I have already argued above, that material is entirely consistent with viewing Churchland as being opposed to representationalism.

One way of getting a handle on Churchland's position is in terms of a distinction between two meanings of "representationalism" that have emerged in the contemporary philosophical and cognitive scientific literature. The first sense is the broad view that all mental states and processes are representational. The second is the more narrow view concerned with consciousness and has also been the primary focus of this paper: it is the view that the properties of perceptual consciousness are the representational contents about environmental objects and properties. To give more precise labels to these

views we could use “representationalism about the mind” for the first view and “first order representationalism about phenomenal consciousness” for the second view. Churchland is an adherent of the first view but not an adherent of the second.

That Churchland is not an adherent of the second view should be relatively obvious by now, since it has been the major task of this chapter to argue this point. However, it may take further work to make it clear how Churchland is an adherent of the first view, especially in light of his commitment to the possibility of the “direct introspection of brain states.” That brain states are *directly* introspectible may very well make it seem like the vehicular properties of experiences are entering into consciousness and thus the overall character of one’s mental life contains more than just representational contents, but also includes vehicular properties themselves. I think, however, that this interpretation of Churchland is ultimately in error. To see this most clearly it is useful to consider how, as mentioned previously, Churchland regards introspection as fallible. The possibility of erroneous introspection is explained in terms of the possibility of introspectively misrepresenting sensations. It is natural to suppose, then, that in such cases, what enters into consciousness are not the sensations themselves, but the ways in which the sensations are represented (which may include inaccurate as well as accurate ways of representing them). What follows, then, from the Introspection Thesis is not that the vehicular properties of first-order states enter into consciousness, but that what enters into consciousness in the direct introspection of brain states are the *contents* of higher-order representations (which are, of course, representations of the vehicular properties of the first-order representations).

What, then, are we to make of the “direct” in “the direct introspection of brain states”? I think that, for Churchland, the directness here is that the introspective judgments are *noninferential*. That is, they are not inferred from introspective evidence but *directly* caused by the occurrence of their target sensations.

I close, then, by briefly summarizing what I take myself to have shown in this chapter. My primary aim was to unpack Churchland’s Introspection Thesis and pit it against the ultimately inferior Transparency Thesis of the representationalists. I then considered the question of whether the tension between the Introspection Thesis and representationalism plays itself out in Churchland’s corpus. I argued that the evidence that Churchland is a representationalist can be easily explained away and is further overwhelmed by contrary evidence. The position that emerges from these considerations is a kind of representationalism about consciousness but not in a sense

equivalent to either First-Order or Higher-Order Representationalism. As I read Churchland, the qualitative character of consciousness is always identical to the content of some representation or other but, contra First-Order Representationalism, it need not always be the content of a first-order representation and contra Higher-Order Representationalism, it need not always involve the presence of some higher-order representation. If there is, however, a tension remaining in Churchland's work concerning the qualitative character of consciousness it is a tension concerning whether he ultimately thinks that the intrinsic properties of neural states themselves can enter into consciousness or whether it is the *representation of* intrinsic properties that enters into consciousness. I offer that the latter option is the superior view, and if it isn't what Churchland explicitly has in mind, then it should be.

I have pitted Churchland's bold and surprising Introspection Thesis against the allegedly obvious yet opposing Transitivity Principle and the Transparency Thesis. Both the Transitivity Principle and the Transparency Thesis are supposed by their proponents to be pre-theoretically intuitively obvious, but once we see what they entail, it is not clear how they can be pre-theoretically intuitively obvious. Transitivity entails that it is *necessary* that we are aware of vehicular properties of our conscious states and Transparency entails that it is *impossible* for us to be aware of such properties. Both claims seem too strong to be accessible to pre-theoretic intuition. In contrast, Churchland's Introspection Thesis, in spite of its bold and surprising content, turns out to be the most plausible of the three.

ACKNOWLEDGMENTS

This work was supported in part by grants from The National Endowment of the Humanities and The James S. McDonnell Foundation's Project for Philosophy and the Neurosciences. I am grateful for feedback on this material from Brian Keeley and Josh Weisberg and from members of the William Paterson University Cognitive Science Reading Group, especially Mike Collins, Thomas Hogan, and Alexander Vereschagin.

Notes

1. Churchland's most extended treatment of the Introspection Thesis is Churchland (1979) but see also Churchland (1985, 1986).
2. See, for example, Tye (1995, 2000) and Dretske (1995).
3. See, in particular, Tye (2000: 45–51).

4. Although, as I'll argue later, just because representational contents *of the experience* do not figure into phenomenal character, this doesn't mean that phenomenal character is anything besides representational content. It may involve, for instance, the representational contents of higher-order states.
5. While the transitivity principle is typically taken to be satisfied by a second representation, it is at least *prima facie* possible for the principle to be satisfied with a single representation that is, in part, self-representational.
6. This remark perhaps suggests a *dispositional* higher-order representational theory of qualia insofar as a state has qualia in virtue of certain dispositions there are for its uptake by the higher-order representations employed in introspection. I will not explore this possibility further beyond noting it here.

References

- Churchland, P. M. & Churchland, P. S. (1981). "Functionalism, qualia and intentionality." *Philosophical Topics* **12**: 121–32. Reprinted in *A Neurocomputational Perspective* (Cambridge, MA, MIT Press, 1989).
- Churchland, P. M. (1979). *Scientific Realism and the Plasticity of Mind*. Cambridge, Cambridge University Press.
- Churchland, P. M. (1989). *A Neurocomputational Perspective: the Nature of Mind and the Structure of Science*, Cambridge, MA, MIT Press.
- Churchland, P. M. (1985). "Reduction, qualia and the direct introspection of brain states." *Journal of Philosophy* **82**: 8–28. Reprinted in *A Neurocomputational Perspective* (Cambridge, MA, MIT Press, 1989).
- Churchland, P. M. (1986). "Some reductive strategies in cognitive neurobiology." *Mind* **95**: 279–309 Reprinted in *A Neurocomputational Perspective* (MIT Press, 1989).
- Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA, MIT Press.
- Harman, G. (1990). "The Intrinsic Quality of Experience." In *The Nature of Consciousness*, Ned Block et al, eds. Cambridge, MA, MIT Press, 663–75.
- Kind, Amy. (2003). "What's So Transparent About Transparency," *Philosophical Studies* **115**: 225–244.
- Lycan, W. G. (2001). "A simple argument for a higher-order representation theory of consciousness." *Analysis* **61**: 3–4.
- Moore, G. E. (1903). "The Refutation of Idealism." *Philosophical Studies*. Totowa, NJ, Littlefield, Adams & Co., 1–30.
- Rosenthal, D. (2002). "Explaining Consciousness," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, New York, Oxford University Press, 406–21.
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA, MIT Press.
- Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA, MIT Press.