Ch 2. Introspecting Brain States as Such

§1. Identity and Neuro-introspection

If the neuro-reductionism of the previous chapter is true, then also true is the following thesis about introspection: *when* we introspect our conscious states, *what* we introspect are actually brain states. However, if we can't in fact introspect brain states, then that fact, the alleged non-introspectibility of brain states, would count against neuro-reductionism.

One point about introspection that even a neuron-reductionist must grant is that when we introspect it doesn't usually seem like we are introspecting brain states. Now, if it is additionally true that how conscious states *are* is identical to how conscious states *seem*, and it is further true that conscious states *never* (not just usually, but in no possible situation) seem like brain states, then those propositions would form the bases for a very powerful case against neuron-reductionism.

In this chapter, I plan to defend neuron-reductionism against such a threat by defending what I'll call the thesis of Neuro-introspection:

> (NEURO-INTROSPECTION): A person with sufficient neuroscientific education can introspect his or her brain states *as* brain states.[16]

---

[16] See, e.g., Rorty (xxxx) and especially Churchland (1979). See also Churchland (1985, 1986).

There are two main sources of resistance against neuro-introspection: non-reductioinism and Transparency. Non-reductionism having been dealt with in the previous chapter, my focus in the present chapter will be Transparency.

The remainder of the chapter is as follows. First I'll say more about Transparency. Next I'll spell out Neuro-introspection. Then I'll defend the superiority of Neuro-introspection over Transparency.

§2. Transparency

Recall that Transparency, as defined in Ch. 0, is the view that when one has a conscious experience all one is conscious of is what the experience is an experience of. We can see how Transparency might pose a threat to Neuro-introspection as follows. If introspection is a way of being conscious of something, then if Transparency is true, when I attempt to introspect my conscious experiences all I thereby have access to are things that the experiences are experiences of.

To spell this out 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 square-ness. I have no introspective access, then, to features of the experience itself. In direct opposition to Neuro-introspection, 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 First-Order representationalism (FOR).[17] Further, FOR 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.[18]

Contemporary discussions of the notion that experience is transparent (or diaphanous) frequently trace the idea back to 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, p. 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:

---

[17] See, in particular, Tye (2000, pp. 45-51).

[18] 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.

> 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, p. 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 qualiophiles, 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, p. 62).

We can get a further understanding on 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, pp. 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, pp. 100-101).

According to the 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.

In contrast, Higher-Order Representation theories of consciousness—HORs—already discussed in chapter 0, explain consciousness in terms of mental representations of other mental states. Recall, as mentioned previously, that a key principle appealed to by HORs is Transitivity (A state is conscious only if one is conscious of this state). Thus, according to advocates of Transitivity 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.[19]

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" and its cognates have 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").

---

[19] While the Transitivity 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. See, e.g., Kreigel (xxxx)

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, e.g. aspects of the creature's environment or body and would thus itself be a first-order state.[20]

Another way of putting the above point is in terms a distinction between representational content and representational vehicle. I may have, at 3pm, a memory of something that happened at 2pm—I may remember that at 2pm someone told me a particularly funny joke. Occurring at 3pm is a property of the representational vehicle, it is a property of the memory itself. Occurring at 2pm 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

[20] See Guzeldere (xxxx).

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 Neuro-introspection.

§3. Inspecting Neurointrospection

Following Churchland (1979), the case for Neuro-introspection may be built on a foundation that consist in two components. The first is an account of perception designed primarily to account for a distinction between what is perceived and what is inferred from what is perceived. The second component is an account of introspection that views it as

analogous to perception. Neurorintrospection is true *because* of a crucial analogy between introspection and perception.

Less schematically, yet still briefly, a sketch of the case is as follows. Regarding the distinction between something that can be perceived and something that can be figured out based on what is perceived, consider the following. If I stick my finger into a hot cup of coffee, I can perceive the heat of the coffee. However, if, without touching the coffee, I see the steam rising from it, it is the steam that I perceive and based on what I perceived, I figure out the approximate temperature of the coffee. As will be elaborated below, 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, 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 through 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. Neuro-intropspection may be defended in terms of 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.

I turn now to flesh this out in further detail, closely following Churchland's own formulations. 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, p. 7; emphasis in the original). 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 semi-formally, 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 or not 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 non-inferentially 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, p.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 6.2×10^{-21} kg m^2/s^2" (ibid, p. 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, p. 26) and we can see "the dominant wavelength (and/or frequency) of incoming electromagnetic radiation in the range $0.38—0.72 \times 10^{-6}$ m, and of the reflective, absorptive, and radiative properties of the molecular aggregates from which it comes" (ibid, p. 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.

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 consider a different, namely a 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.

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, p. 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, p. 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* as 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, p. 39; emphasis in the 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, p. 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, p. 40; emphasis in the original). One consequence of this view of introspection, important both to Churchland and for points I'll raise below, 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, p. 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 3mm when in reality it is 3.5mm. However, when the device is correctly calibrated, what indicates that the height is 3mm is when the 3mm mark is even with the top of the mercury column that is, in fact, 3mm 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 Neuro-introspection and cast doubt on Transparency.

> 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, pp. 118-119).

If a person knew that their mental states were identical to brain states, but was incapable of *automatically* applying the concept 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.

§4. Neuro-introspection over Transparency

Neuro-introspection involves a middle-ground between Transparancy and Transitivity because it 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. Neuro-introspection 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 Neuro-introspection 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 Neuro-introspection entails that you can, (but don't have to) be aware of vehicular properties of conscious states.

Not only does Neuro-introspection conflict with the Transparency Thesis, but it threatens the larger project of FOR. As Amy Kind (2003) has argued, if the Transparency is false then FOR 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 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 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 Transparency, 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, p.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 Transparency and the Neuro-introspection. 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 Neuro-introspection more plausible than Transparency, but the premises upon which Neuro-introspection is based can also be used to explain whatever initial plausibility the thesis of Transparency 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 may Transparency seem plausible without being true.

I turn now to briefly address the question of whether in spite of undermining FOR, Neuro-introspection might be consistent with a kind of representational theory (one that is neither FOR nor HOR).

That brain states are 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 Neuro-introspection is ultimately in error. To see this most clearly it is useful to consider how, as already mentioned above, introspection may be 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 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).

I have pitted the less-than-obvious thesis of Neuro-introspection against the allegedly obvious yet opposing theses of Transitvity and Transparency. Both the Transitivity and Transparency 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, Neuro-introspection, in spite of its bold and surprising content, turns out to be the most plausible of the three.

In the next chapter I turn to mount a quite different attack against Transparency and Transitivity as well as the related theories of FOR and HOR.