2. The Internal State View: Experiences as Inner Physical Modifications

> The only properties of conscious experience with which we can make contact are intrinsic [neural] properties of subjects.
>
> --David Papineau (2016)

> Visual experience is intrinsically [essentially] spatial . . . if we do not use spatial properties in characterizing the visual [experience], we omit a subjective feature of the experience.
>
> --Christopher Peacocke (2008)

In the previous chapter, we considered the sense datum view of experience. On this view, when you view a scene, the entire space you experience is in fact a private mental arena, and the "objects" within this space are very life-like mental images, or "sense data". Changes in the character of your experience are changes in these "sense data" you experience.

By the 1950s and 1960s, everyone wanted to get rid of sense data because they would have to be strange, non-physical items. But if we reject the sense datum view, what view should we put in its place?

The *internal physical state* view (the *internal state view* for short) is the first alterative we will consider. Instead of holding that experiences are relations to non-physical sense data *created by neural states*, this view holds that experiences are identical with *neural states themselves*. Changes in the character of experience *just are* changes in neural states. After the sense datum view, this is a very natural next choice.

Recent proponents of the internal physical state view include Ned Block (2019), C. L. Hardin (1988), Geoff Lee (forthcoming), Brian McLaughlin (2016a), David Papineau (2014), Thomas Polger (2004), and Hilary Putnam and Hilla Jacobson (2014). Unsurprisingly, it is popular among neuroscientists. For instance, Christof Koch and Guillermo Tononi (2015) have proposed the "integrated information theory" of experience, which is a form of the internal physical state view.

The internal physical state view may seem obvious, almost inevitable. After all, the common factor between normal experience of a tomato and a hallucination of one is an internal physical state. But we will see that it also faces a problem. How does it accommodate the phenomenological fact of external directedness? For instance, visual experiences are quite different from internal

sensations, like headaches. Even if they depend on internal factors, they also *essentially* involve the seeming presence of items in space with various spatial features. Is the internal state view consistent with this fact?

The plan for this chapter is as follows. In section 2.1 we will learn more about what the internal state view is. In section 2.2, we will look at a possible argument for the internal state view based on its fit with internal dependence. In sections 2.3 through 2.5, we will consider problems for it concerning whether it can accommodate the essentially externally directed character of some experiences.

2.1 What is the Internal State View?

Recall that the central question of this book is the *character question:* what does the character of experience consist in? What do differences in the character of experience consist in? What kind of a thing is an experience? The internal state view gives a simple answer.

> *Internal physical state view:* every sensory-perceptual experience with a certain character is necessarily identical with an "internal" physical property, as it might be, a complex neural pattern. This property is an *intrinsic* property of the brain; it is not a relation to anything outside the brain. Differences in the character of experience consist in differences in such intrinsic physical-properties of subjects.

For instance, the vast difference between the experience of a color and the experience of a smell is just a difference an internal physical-computational difference in the brain, as it might be, a difference in the spatio-temporal pattern of neural firings.

I will continue to shorten the name of this view from "the internal physical state view" to the "internal state view", but it should be kept in mind that we are considering the view that types of experiences are necessarily identical with types of intrinsic *physical* states.

Our go-to example in chapter 1 was the experience of a humble tomato. To mix things up a bit, in the present chapter, let's consider an experience of an orange. This is a nod to one of the originators of the internal state view, J. C. Smart (1959), who illustrated the view with an experience of an orange.

As we will see in the next section, there is reason to believe that the character of your experience depends in some *systematic and regular way* on the character of your internal neural activity. In the end, this is what pins down the character of your experience. So if we knew the pattern of activity in some population of your neurons (the ones coding for color), and if we knew the systematic "neural code" for color, we could determine the character of your color experience as you view the orange. Likewise, if we knew the pattern of activity in some other population of your neurons (the ones coding for spatial features), and if we knew the systematic "neural code" for shape, we could determine that you have a round-experience rather than a square-experience. As the neuroscientist Stanislas Dehaene writes, "the [neural] code contains a full record of the subject's experience" and "if we could read this code we should gain full access to a person's inner world" (2014, 143-145). (See also Prinz 2012, 126-133 for an important discussion of this issue.)

Now if this is right, then a simple and natural hypothesis suggests itself: having an orange-experience (the kind of experience you in fact have when viewing the orange) *just is* undergoing a unique set of neural patterns in different parts of the brain – in the way that water *just is* H_20 or light *just is* electrical discharge.

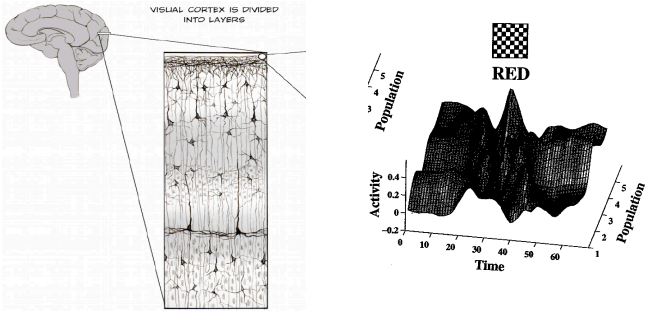

Figure 1. (A) A system of neurons. (B) representation of a spatio-temporal pattern of neural firing among a system of neurons (from McClurkin *et al*. 1996)

Suppose that the orange appears to change color, from orange to green. Then the pattern of firing in some of population of neurons changes (those coding for color), while the pattern in another population of neurons (those coding for shape) stays the same. On the internal state view, the change in the character of your experience is just *identical with* this change in the pattern of firing in your neurons. It is *not* a change in a "non-physical sense datum" that

is "created by" the change in neural pattern; it is just the change in the neural pattern itself. Likewise, if the orange appears to move, the change in the character of your experience just is another change in the neural pattern.[1]

On the internal state view, is your experience identical with the biological state, so that a robot couldn't have it? Or is it identical with a more abstract (but still intrinsic and physical) state that a robot could be in principle share? There are different versions of the internal physical state view. I will typically focus on a version that holds that different types of experiences are necessarily identical with different types of *patterns of neural activity*, or *neural states* for short.

Now you might think that the internal state view is obviously wrong. For instance, when you see an orange, you know that you have an orange-experience. But you don't know that you are undergoing a certain distributed neural pattern. When you inspect your experience, you don't find any neurons firing at all. Furthermore, there is an "explanatory gap": how could the activity of neurons in soggy grey matter constitute technicolor phenomenology?

But these quick arguments are suspect. To see this, consider an analogy. My daughter knows that there is water in the tub. But she doesn't know that there is H_20 in the tub. And when she inspects the water, she doesn't detect individual H_20 molecules. Nevertheless, water is H_20. In the same way, maybe having an orange-experience is identical with undergoing a certain type of distributed neural process, even though this is not evident to you.

The internal state view is the polar opposite of "naïve realism". On naïve realism, the explanation for the character of you experiences resides primarily in *your relations to the external world*. On the internal state view, the character of your experiences is entirely constituted by the character of *your internal neural processes*.

---

[1] Here is another example. Suppose you experience a blue square above a red circle and then a red square above a blue circle. There is some change in how your neural patterns coding for shape and color are "bound" together in the brain. The change in your experience *just is* this change in your total neural state, according to the internal state view. So, internal physical state theorists have a simple answer to the "many-property problem" (Jackson 1977).

The internal state view also differs from sense datum view. To see this, suppose you have a total hallucination. You hallucinate an orange rolling on a table. On the sense datum view, your brain causes an orange and round object to come into existence. It cannot be found within your physical brain but lives in a separate, private mental arena. And you experience this ghostly object. Internal physical state theorists avoid this "act-object" account of your hallucination by invoking the "seems gambit" (chapter 1). They hold, while there *seems to exist* an arena containing an orange and round object, there really is no such arena. However, your *experience* exists: you really have an experience in which it seems to that there is such an object. Your experience is identical with the relevant neural state in your brain.

It would be a mistake to think that the internal state view resembles the sense datum view, with the only difference being that it holds that the orange and round object you experience exists *inside your brain* (so that, although you do not know it, you are seeing your own brain). This would still be an act-object theory, but where the "object" is interior to the brain. Internal physical state theorists reject the act-object view entirely. To repeat, their picture is this: your hallucinatory experience is your neural state *B*, and your neural state *B* makes it *seem* as if there is an orange and round thing, but *there is no such thing anywhere* - not even in your brain.[2]

In short, internal physical state theorists allow that the "act-object" view *seems* true, but they insist it is totally false. In that sense, they favor an *error theory*. The true nature of experience differs from how it seems.

You might have some residual questions about the internal state view. For one thing, I just said that internal state theorists invoke the "seems gambit". But since they deny that descriptions of how things seem are grounded in the literal properties of sense data in a private sense-field, they must give some alternative account of how things seem. What is their alternative account?[3]

---

[2] Russell (1927: ch.26) said that we don't know about the intrinsic nature of the physical world, including our own brains. See also Strawson 2020. But we know enough to know that there is not an orange and round thing in your brain moving to the right, fully constituted by neuronal firings!

[3] Internal state theorists can combine their view with an "adverbialist" semantic theory about the meaning sentences like "it looks *as if* I'm seeing a round thing, but I'm not

Relatedly, you might wonder, what do internal state theorists say about the whereabouts of the sensible properties that are necessarily bound up with the character of experience, such as sensible colors, audible qualities, and pain qualities? For instance, having an orange-experience necessarily involves "being presented with" an *orange* quality which appears to fill a round region in space. Where in reality is this quality to be found?

Some philosophers and scientists - for instance, Ned Block (2010: 24, 56n2), Semir Zeki (1983: 764) and Stephen Palmer (1999: 97) - seem to hold that this orange quality is *actually a neural property of the population of neurons* that codes for the chromatic aspect of your experience (as it might be, a unique spatio-temporal pattern of neural firings). The poet Oscar Wilde put it this way: "It is in the brain that the poppy is red, that the apple is odorous, that the skylark sings." Maybe internal state theorists could take this brain-based view?

This may seem to be a natural combination, but it faces an immediate difficulty, which we can appreciate by contrasting it with the sense datum view. On the sense datum view (e.g. Russell 1912, Peacocke 2008), when you have the orange-experience, the orange quality is a property of a *literally round* "sense datum". Internal state theorists reject the existence of such a thing, so they cannot take the same view. Instead, the brain-based view holds that this *same* quality is a neural property of a population of neurons that is *not* round. And now the difficulty is this: if the orange quality that is essentially involved in the orange-experience is really a neural feature of a population of neurons that *isn't* round, how come it *appears* to fill a round region?

In view of these problems, could internal state theorists about experience reject the brain-based view of sensible properties? For instance, might they

---

seeing a round thing". Roughly, on this semantics, this sentence means that you are sensing in a "way" that *generically* goes with seeing a round thing, but in this case you are *not* seeing a round thing. Notice that, whereas the internal physical state view is about the metaphysical structure of experience itself, "adverbialism" in this sense is a *semantic theory* about the meanings of *sentences* describing experiences. In fact, adverbialism in this sense is totally quiet about the metaphysical structure of experience. So it is not only compatible with the internal state view; it is compatible with all the other theories about the metaphysical structure of experience discussed in this book (Breckenridge 2018: 7).

instead accept an extreme version of the *rife illusion view* (section 1.5)? That is, might they say that experiences are neural states, and undergoing those neural states makes it *seem* that items arranged in space "out there" have sensible properties, but in fact *nothing* has these properties – not sense data (the sense datum view), and not even bits of the brain (the brain-based view)?

These questions are important. They concern the question of how internal state theorists might accommodate the "externally directed character" of experience. As I formulated the internal state view above, it is neutral on these questions. So we can understand it without having answers. We will take up the question of whether the internal state view comports with the externally directed character of experience later on (sections 2.4 through 2.7).

First we will consider a possible argument for the internal state view: that it is the simplest theory consistent with the evidence for the dependence of experience on internal factors.

2.2 An Empirical Argument for the Internal State View

The argument for the internal state view to be examined in this section has two steps. The step will argue for the general thesis "experiential internalism": every aspect of the phenomenological character of our experiences is directly and fully determined by our neural states.[4] This rules out "naïve realism". It moves us to a different ballpark of views that includes the sense datum view and the internal state view. The second step will use additional considerations to argue that, between these options, the internal physical state view is to be preferred.

Let's look at these steps in greater detail, and then turn to some problems with the argument.

First step: experiential internalism. The best case for experiential internalism is based on empirical findings. (In the previous chapter, we cited the same empirical findings in support of "illusionism" about the sensible properties.)

[4] In the introduction to this book I said that the central puzzle about sensory experience concerns how it can be both essentially externally directed and "internally-dependent". The thesis of experiential internalism is a particularly strong form of "internal dependence".

Start with the experience of pain. Even under normal conditions, there is no simple or systematic relationship between bodily disturbances and sensory pain intensity. By contrast, neuroscience has shown that firing rates *in the cortex* are "linearly related to subjects' perceived pain intensity" (Coghill et.al 1999: 1936). So, while pain intensity is only related to the bodily stimulus intensity in a rough way, it is more directly related to internal firing rates. In the domain of smell, similarities and differences in the smell qualities we experience are very poorly correlated with similarities and differences in the molecular-types that we smell. They are only well-correlated with distributed patterns of neural firing in the smell system (Youngentob *et al.* 2006; Howard *et al.* 2009). Likewise, in the domain of color, similarities and differences in color experiences fail to line up with similarities and differences in the ways external objects reflect light (Byrne and Hilbert 2003: 13; Thompson 1995: chapter 3). They only line up with similarities and differences in distributed patterns of neural firing in the color system (Bohon *et al.* 2016).

These findings support experiential internalism. True, they are limited to the experience of certain sensible properties: pain, smell, and color. But they provide reason to conjecture that *all* aspects of our experiences - including spatial and temporal aspects – are completely determined by our neural states.

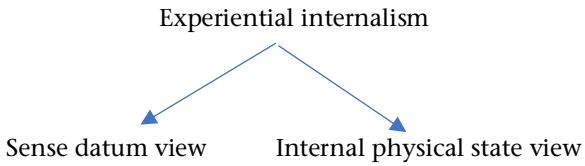

Figure 2

Second step: from experiential internalism to the internal state view. Both the sense datum view and the internal physical state view are forms of "experiential internalism". So once we accept internalism, we must choose between them (Figure 2). The second step of the argument uses more "philosophical" considerations to argue that, as between these options, we should favor the internal physical state view. After all, the sense datum view is a dualistic theory which holds that our internal neural states generate non-physical sense data that reside in private mental spaces. This is complicated and mysterious. The internal state view eliminates this extra step. It *directly identifies* experiences with our internal neural states. It is therefore simpler and less mysteri-

ous than the sense datum view. Furthermore, since the internal state view uses the "seems gambit" to avoid sense data (chapter 1), it avoids the myriad puzzles attending sense data: the percipi puzzle, the puzzle of location of sense data, and puzzles about indeterminacy.

That, then, is the argument for the internal state view. Is it convincing? In fact, both steps are open to criticism.

First, there is a gap between the empirical evidence and experiential internalism. From the fact that internal factors participate in the *determination* of *some* aspects of experience, it doesn't immediately follow that all aspects of experiences are fully determined by internal factors.

Second, even if we grant experiential internalism, there is a problem with the move from there to the internal physical state view. True, between the internal physical state view and the sense datum view, there is some reason to prefer the internal physical state view. But Figure 2 is misleading. If we accept experiential internalism, these are not the only options. They are just the only options we have discussed. There is a third option that should be added: "internalist representationalism", to be discussed in chapter 4. Like the internal state view, it fits with experiential internalism and avoids mysterious sense data. So, to complete the argument, internalist state theorists would need to provide additional reasons that break the tie and favor the internal physical state view over this third option. We will return to this issue in section 3.11.[5]

Still, the empirical facts cited above provide *some* reason to accept the internal state view. If in many cases the structural relations among experiences (similarity and difference, equal intervals, proportion) do not match the structural relations among the complex external physical properties that our brain is responding to, but they do match the structural relations among their internal neural correlates, then this certainly raises the probability of the view that our experiences are just identical with those neural correlates. For in-

[5] The argument for the internal physical state view based on experiential internalism presented in the text has a significant empirical component. Papineau (2014, 2016) gives a different, more philosophical argument for internal state view. In particular, he thinks that its only viable rival is representationalism. But he thinks that representationalism is ruled out by reflection on the nature of experience. We will examine Papineau's arguments against representationalism in sections 3.6 and 3.7 of the next chapter.

stance, maybe pains just are distributed neural states in the "pain matrix", with their intensity constituted by the average firing rate of neurons. Experiences of color and smell just are different distributed neural patterns. This provides a simple and natural answer to the "character question" that is in line with the empirical evidence.

In the rest of the chapter, we will look at problems for the internal physical state view. For one thing, many have objected to it on the grounds that it is inconsistent with the so-called "transparency observation" (section 2.3) Another problem is that it not consistent with the externally directed character of some experiences (sections 2.4 through 2.7).

2.3. Can the Internal State View Accommodate the "Transparency Observation"?

When you see an orange and a tomato on a table, you come to know things about the *objects that you experience*, for instance, that there is an orange thing and a red thing there. But you also come to know something about the character of *your experiences themselves:* for instance, that you are having an orange-experience and a red-experience.

The "transparency observation" is about the second type of knowledge: it is about how you attain knowledge about what your own experiences are like. It is about *introspective* knowledge.

The transparency observation is one of the most discussed ideas in recent philosophy of perception. Different philosophers formulate it differently. We will focus on Michael Martin's forceful discussion of the issue in his paper "Setting Things Before the Mind" (1998). In this paper, Martin offers a particular formulation of the transparency observation, and also argues that it rules out the internal state view. Following him, we will focus on the case of visual experience.

Here is how Martin formulates the thesis:

> *Transparency observation:* In general, "the way in which we learn what our [visual] experiences are like is by attending to the objects and features [in space] which are presented to us in perception" (xx). For instance, if I view an orange, I know what my experience is like by focusing on an orange and round thing in space.

This is called the "transparency", because the idea is that your experience itself is "invisible". You cannot attend to it *as distinct from* attending to the objects and features you are presented with.

This thesis fits poorly with the internal state view. It fits much better with the kind of "act-object" approach that was our focus in chapter 1.

To illustrate, suppose you have a normal experience of an orange and then you have a hallucination of an orange (maybe you have "Charles Bonnet Syndrome" discussed in section 1.4). On across-the-board sense datum theory (chapter 1), the relevant item in both cases is a sense datum created by your brain. On "normal-abnormal naïve realism" (section 1.3), the item is the physical orange in the first case and an orange and round sense datum created by the brain in the second case. Both of these versions of the act-object model predict the transparency observation: in both cases, you know what your experience is like by attending to the objects and properties you perceive.

But now consider the internal state view. This view holds, contrary to the act-object model, that the character of experience is not determined by any colored and shaped objects that you perceive. Rather, it is constituted by your internal *neural patterns*. Unlike the act-object model, the internal state view would not seem to predict the transparency observation at all. And this counts against it.

That is a first-pass statement of the argument from the transparency observation against the internal state view. Is it a good argument?

One problem is that the transparency observation has been much disputed. There are apparent counterexamples (see chapter 3). For instance, if you look at two dots on a piece of paper and move your attention from one dot to the other, you know that your experience changes. But it seems that you don't know this by focusing on some change in "what you experience". There is no change in *what* you experience. There is only a change in *how* you experience it. Or again, if you take off your glasses so that things look blurry to you, then you know that your experience changes, but it seems that you don't know this by focusing on some change in "what is presented to you". *What* you experience doesn't seem to change; there is only a change in *how* you experience it.

However, here I will set such apparent counterexamples aside and focus instead on what I think is a deeper problem with the transparency observation as Martin formulates it.

The problem I have in mind concerns illusion and especially hallucination. When we formulated the transparency observation, we did not restrict it to veridical experience. It was meant to apply to all visual experience. But consider an actual hallucination. One woman with "Charles Bonnet syndrome" described hallucinating "a colored flag in sharp focus . . . it looked exactly like a British flag" (Sacks 2012: 11). She surely knew what her experience was like. Now, the general transparency observation entails that she knew what it was like by attending to some existing *object*. If you attend to *o*, then *o* must exist. But since no physical object was present, this would have to be a non-physical object, namely, a flag-like sense datum created by her brain. So, on the face of it, the transparency observation as formulated above requires sense data!

But we have seen that there are very strong reasons to reject sense data. Therefore, the transparency observation as formulated above is very likely *false*, because there are some hallucinations where it is not the case that we know what our experience is like by attending to objects that we experience.

(It may be that we can draw a more general conclusion from cases of illusion and hallucination. If in non-veridical cases attending to objects is not part of what explains the justification we have for our introspective beliefs, then it becomes natural to think that in veridical cases too it is not part of what explains our having such justification.)

In fact, Martin seems to briefly acknowledge in passing that hallucination undermines the general transparency observation as formulated above. He acknowledges (xx) that "even in cases of hallucination, there is a way that one's experience is for one, and one can come to know what one's experience is like, *yet there are no objects [no "sense data"] of perception for one to attend to*" (my italics).

But if the general transparency observation as formulated above is false, then it cannot be used in a sound argument for the failure of the internal state view.

At this point, we might try to formulate an alternative version of the transparency observation, one that is not undermined by hallucination. And then perhaps we might use it in a sound argument against the internal state view. But what would that be? Martin does not officially offer a revised transparency observation. However, he does write:

> In as much as an hallucination may be indistinguishable for one from a genuine perception, it will still *seem to one as if* there is an array of objects there for one to scan and explore. (xx)

This suggests:

> *Seeming transparency:* as a matter of fact, *whenever* you have a visual experience and you know what it is like, then the following things at the very least *seem* true: it *seems* that that there are items in space for you attend to, and it *seems* that you know what your experience is like by attending to some items.

Maybe "seeming transparency" is true. In fact, maybe it is pretheoretically plausible. But now we need to ask whether it also has bite. That is, can it also be used in a convincing argument against the internal state view?

This is not clear. Even though internal state theorists can reject the general act-object model of experience, that doesn't automatically forbid them from saying that this assumption at least *seems true* to us when we reflect on what our experience is like. This is just an extension of the "seems gambit". They will just say that in this case things are not as they seem. In other words, introspection leads us astray about the nature of our own experiences. As I said before, in that sense, they advocate an "error theory". In that case, even if they must reject the transparency observation as initially formulated, they can accept that the transparency observation *seems* true. That is, they can accept "seeming transparency". At least, we have been given no reason to think otherwise. In short, once we water down the thesis in this way to accommodate hallucination, it is not clear that it is inconsistent with the internal state view.

Maybe internal state theorists can even explain seeming-transparency. Maybe there is seeming-transparency because our most natural way of describing experiences is indirect: our most natural way of describing experiences is in terms of the *worldly situations that bring about* our internal experiences (Papineau 2014: 24).

Now you might think we shouldn't quite yet give up on the idea that some *stronger* version of the "transparency observation" (stronger than mere "seeming transparency") *does* undermine the internal state view. You might think: we just have to formulate such a stronger version of transparency thesis, one that is (a) pretheoretically plausible and defensible (e. g. compatible

with hallucination), *and* (b) at the same time has *bite*, that is, is *inconsistent with* the internal state view.

Here we will not attempt to search after such a version of the transparency observation.[6] For, in the remainder of this chapter we will see that, to rule out the internal state view, we may not need to rely on any controversial "transparency thesis" about introspection. We may be able to rule it out on the basis of a related but much simpler idea, namely the idea that some of our experiences are *essentially* "externally directed". Unlike the transparency observation, this is not a theory concerning the thorny issue of *how we know* what our experiences are like. It is a more modest and defensible claim just concerning *what some experiences are like*.

I will begin by saying more about what essential external directedness amounts to (section 2.4). Then we will use it to construct a new argument against the internal state view (sections 2.5 and 2.6).

2.4 The Essentially Externally-Directed Character of Some Experiences

Suppose that (for some reason) you have an experience of an orange moving to the right:

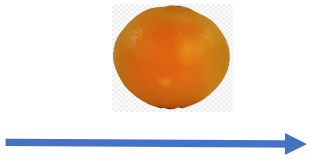

Figure 2: an orange moving to the right

[6] Another alternative formulation of the transparency observation, which avoids commitment to sense data, implies that when someone hallucinates the British flag they know what their experience is like by becoming aware of, and attending to, a "property-complex" even if there exists nothing that instantiates that property complex (not even an array of sense data). See Tye (2000: 48) and Dretske (2003: 73-74). But this is hardly a pretheoretical idea. And since property-complexes are abstract items that don't take up space (like numbers), it is hard to see how the hallucinator could be aware of them and attend to them. For discussion see Schellenberg 2018 and Tye 2019. It is difficult to formulate a version of the transparency observation that is both pretheoretically plausible (e. g. compatible with hallucination) and rules out the internal physical state view (Pautz 2007).

Call this type of experience (whether it occurs in normal experience or hallucination) the *orange-experience*.

Intuitively, the orange-experience is essentially externally directed in the following sense. It is part of the *essence* of the orange-experience that, if you have it, then it *seems* to you that there is *right there* an orange and round thing moving to the right. In that sense, the experience is "directed at" such a thing. To put it in a different way, it is in the essence of the orange-experience that it is an experience *as of* an orange and round item moving to the right.

Why call this essential *external* directedness? Because, if there really is an orange and round item there (a physical item or sense datum), it presumably exists somewhere external to your physical brain (after all, such items needn't exist in your physical brain while you are having the experience).

What does the term "directedness" mean? It's a metaphor. Your experience is metaphorically "directed" at an orange item moving to the right. You shouldn't read too much into the metaphor. In particular, the metaphor of directedness is sometimes use to explain the idea of "representation", as when a belief or a sentence is about ("directed at") something that may or may not exist. So the metaphor might suggest the representational theory of sensory experience that we will discuss in the next chapter. But in fact essential external directedness is not equivalent to the representational theory, although the representational theory accommodates it. It is a pretheoretical idea. And, as we shall see, nearly all major theories accommodate it, with the exception of the internal physical state view.

What does it mean to say that the orange-experience "essentially" involves the seeming presence of an orange and round thing? Water *is* essentially made of made up of H_20: it wouldn't be water if it didn't have this chemical composition. Or again, eight is essentially the successor of the number seven. In the same way, some facts about our experiences touch on their essence. The fact that the orange-experience is directed at an orange and round item is one of them. Part of *what it is* to have the orange-experience is for there to seem to be an orange and round item moving to the right. This is something that just happens to be true. Rather, in *any* possible situation, if any individual has this

experience, then they have experiences as of an orange and round thing moving to the right.[7]

This means that, even if someone should have the orange-experience during a hallucination, it seems to them that there is an orange and round item moving to the right. It is just that, in the hallucination case, no such physical object really is present.

I've focused on an ordinary visual experience. Other experiences are essentially externally directed, but not to the same degree as ordinary visual experience. We do not need to formulate a general thesis. The right thing to say is that we can roughly arrange experiences in a series depending on how externally directed they are. For instance, experiences of afterimages are somewhat externally directed, because they present images as located and as having shape; but their content is not as richly spatial as ordinary experiences, since they needn't present the images as being at any determinate distance. Bodily experiences are externally directed: they present qualities in bodily regions. Auditory experience present sounds as coming from certain directions. However, I will continue to focus on visual experiences.

So we have:

> *Essential External Directedness:* Some types of visual experiences essentially involve the *seeming* presence of items with certain shapes and other spatial features. They are essentially *as of* such items. This is so whether they are veridical or hallucinatory. For instance, having the orange-experience essentially involves the seeming-presence of a *round* thing that *moves to the right*.

[7] We could explain essential external directedness in terms of Fine's (1994) basic notion of "constitutive essence": it's in the constitutive essence of the relevant type of experience that it involves the seeming-presence of a round thing moving to the right. I would prefer to explain it in terms Dorr's (2016) notion of "real definition" or "identification": a correct "real definition" of what it is to have the orange-experience will have the form: to have the orange experience is to [. . . round . . . moves to the right]. (For Dorr, the real definition "to be F is to be G" is a generalized *identity*. This notion of real definition is to be distinguished from Rosen's (2015) notion which is explained in terms of "grounding".)

This formulation uses the term "seems". There is some controversy about what we mean when we say things like "it seems (or appears) that there is a round thing moving to the right" (xx). But we could sidestep these controversies by formulating essential external directedness without using "seems" talk:

> [#] For many experience-types, a correct real definition of them will use spatial terms. For instance, a correct definition of what it is to have the orange-experience will somehow include spatial terms such as *round* and *moving to the right*.

This is immediately plausible – just look at Figure 2. As Christopher Peacocke has put it, "visual experience is intrinsically [essentially] spatial . . . if we do not use spatial properties in characterizing the visual [experience], we omit a subjective feature of the experience" (2008: 10). Further, [#] is enough to refute the internal physical state view, as we will see in the next section. However, what follows, for ease of expression, I will use the "seems" formulation of essential external directedness.

In chapter 1, we saw that the starting point of the sense datum view was the *act-object* assumption. This assumption is very strong. It implies that having the orange-experience essentially involves the *real presence* of a round object, even in a hallucination case. We saw that there are reasons to doubt this starting point because it leads to sense data, which are problematic.

Essential external directedness doesn't face the same problem. It is the weaker claim that having the orange-experience essentially involves the *seeming* presence of an orange and round thing. There needn't be such a thing – not even a sense datum. You can accept it and avoid sense data by appealing to the seems-gambit. For instance, this is what "representationalists" (chapter 3) do. Compare: if you are searching for the fountain of youth, the content of your mental state essentially involves the notion of a *fountain*, even if the relevant fountain doesn't exist.

Likewise, (#) doesn't commit to sense data. For instance, if you have the orange-experience in a hallucination case, the term *round* might enter into the definition of the experience (because it enters into how things *seem* in having the experience), even though no *round* sense datum is present.

While essential external directedness is not committed to sense data because of the availability of the "seems gambit", it is compatible with the sense

datum view. In fact, the sense datum view implies and explains it (as we discussed in section 1.6). If orange-experience essentially involves the *actual* presence of an orange and round sense datum, this explains why it essentially involves the seeming presence of such an item.

The case for essential external directedness is based on reflection. It just seems to be an obvious description the phenomenological character of some types of experience. To see this, suppose you have a diffuse headache throughout your whole head. Here is an obvious comment about the phenomenology of this experience: it does *not* essentially involve the seeming presence of a round thing moving to the right. It is an equally obvious comment about the phenomenology of the orange-experience (in both normal experience and hallucination) that it *does* essentially involve the seeming presence of a round item moving to the right. It is part of the essence of the orange-experience that it is an experience as of a certain specific apparent *scene*. And this apparent scene can only be adequately characterized by using the terms *round* and *moving to the right*, where these terms are used with their normal meanings to express genuine spatial properties.

Here is another argument for essential external directedness. The following conditional claim is very plausible: if any individual (even the "brain-in-a-void" to be discussed in section 2.6) were to have all the same visual experiences as you, then they could have concepts of shapes and beliefs about the shapes of things. This requires that it is *built-in* to visual experiences that they involve the seeming presence of things with those shapes.

So even if, owing to hallucination, we deny the act-object assumption that in every case there must really be an orange and round item present ("sense data"), we should retreat to the weaker claim that in every case there at least seem to be such items present. That is why this book started in the introduction with essential external directedness, and not the stronger and more controversial act-object assumption.

The assertion that some visual experiences are necessarily externally directed is neutral on many questions. For instance, one question is: when you view a tilted penny, is the type of experience you have necessarily as of a thing that is elliptical, or of a thing that is tilted and round, or as of a thing that is "elliptical-from-here" (the view-point relative but objective property of having a shape that would be occluded by an ellipse placed in a plane perpendicular

to the line of sight)? Or are multiple answers correct? The discussion and arguments to follow are neutral on this issue.

Essential external directedness is quite minimal in another way. It is not committed to the claim that, necessarily, whenever someone has the orange-experience, it seems to them that there is a "*mind-independent*" orange and round item – an object whose existence and character is independent of you. That is good because this stronger "mind-independence" version of essential external directedness is problematic. Imagine that you have just come into the world and your first experience is the orange experience. In accordance with my minimal form of essential external directedness (#), it is correct to say that there is ostensibly an *orange* and *round* item in your visual field *moving to the right*. But, since this is your first experience and you know nothing of the objective world, it doesn't yet seem "mind-independent" to you. For all you know, it is a transitory mind-dependent sense datum. So the stronger "mind-independence" version of essential external directedness is mistaken (Farkas 2013, Byrne 2013).

The thesis of essential external directedness differs from the transparency thesis in a couple of respects. First of all, the transparency thesis is a controversial and general claim about *how we know* what *all* visual experiences are like. We saw that there are potential counterexamples involving attention shifts and blurry vision. We also saw that it faces a problem about hallucination. By contrast, essential external directedness is only a claim about what *some* experiences are like. So attention shifts and visual blur are not problems. And hallucination is not a problem either, as we have seen.

Second, the transparency observation (at least the form discussed in the previous section) is not obviously inconsistent with the internal state view. By contrast, essential external directedness is apparently inconsistent with the internal state view, as we shall now see.

2.5 The Argument from External Directedness Against the Internal State View

The argument from essential external directedness against the internal state view a simple logical rule called *Leibniz's law*. This rule states that if what is true of A is not true of B, then A ≠ B. The argument says that different things are true of types of experiences and types of neural states, so they cannot be literally one and the same:

1. A correct definition of what it is to have the orange-experience will include the spatial terms *round* and *moves to the right*.
2. This is not true of the underlying neural state *N*. It not the case that a correct definition of neural state *N* will include the spatial terms *round* and *moves to the right*.
3. Therefore, having the orange experience is not identical with underlying neural state *N*, even if it may be dependent on that neural state.

To appreciate the argument, look at Figure 2 illustrating the orange-experience. Premise 1 says that a definition of what it is to have the orange-experience will include the spatial terms *round* and *moves to the right*. This is pre-theoretically very plausible. It is the same as (#) from the previous section. Now look at Figure 1b illustrating neural pattern *N*. Premise 2 says that neural pattern *N* is different. No definition of what it is to undergo the neural pattern *N* will include the spatial terms *round* and *moves to the right*.[8] Because different things are true of them, having the orange-experience cannot be *identical with* undergoing neural pattern *N*, even if they are intimately connected. Because it concerns spatial features, we might also call this the *spatial argument* against the internal state view.

At this point, the internal state theorist might respond as follows:

> I want to agree with premise 1. But what is the case for premise 2? Why can't I just reject it? Of course, undergoing the neural pattern *N* doesn't involve the real existence of a round thing *in the brain* moving to the right. But why can't I say – contrary to Premise 2 - that part of the essence of undergoing the neural pattern is that it involves there *seeming to be* a round thing moving to the right? In

[8] Some philosophers, for instance Russell (1913, 79) and Chalmers (2010, 443; 2012, 296-297) accept essential external directedness but then go on to argue on the basis of physics that experienced spatial features are not really instantiated by physical objects (just as many argue that color qualities are not really instantiated by physical objects). We will discuss this "illusion" view in section 3.10. Even if it is correct, the argument from essential directedness against the internal physical state view is sound. Even if the relevant spatial features are uninstantiated, if *they enter into the definition of experience-types but not neural types*, it follows that experiences are distinct from neural-types. That is the key point.

> that case, the argument collapses. The internal state view is totally consistent with essential external directedness.

Here is the reason why the internal physical state theorists cannot make this response. Look again at Figure 1. What it is to undergo the neural pattern can be *completely* described in terms of *types of neurons* and the *times*, *directions* and *intensities* at which they fire. This is just what it is to be a neural pattern. Therefore, in accordance with premise 2 the definition of what it is to undergo the neural pattern can be fully characterized *without* mentioning spatial terms *round* or *moves to the left*. (It is true that states can have hidden essences but we know that the essence of the neural pattern doesn't involve these spatial features.) This rules out the response that it is part of the definition of undergoing the neural pattern that it involves the seeming-presence of a *round* thing *moving to the right*.

The argument from essential external directedness against the internal state view here is analogous to an argument that all will accept.

1. A definition of what it is to *think* that something is *round* and *moving* will mention *round* and *moving right;* for to think this is to attribute *round* and *moving to the right* to something.
2. This is not true of making the noise "something is round and moving"; this noise can be fully characterized without mentioning the spatial features *round* and *moving to the right*.
3. So thinking that something is round and moving is not identical with making the noise "something is round and moving"; it might sometimes *involve* making this noise, but it is *something more* than making this noise.

The argument from essential external directedness against the internal physical state view seems straightforward. Suppose we accept it. What alternative view might we accept? What views are consistent with essential external directedness?

In fact, all the other views we discuss in this book endorse the argument and its conclusion:

> *Sense datum view* (chapter 1). To have the orange-experience is to experience a sense datum that is *round* and *moving to the right*, in accordance with essential external directedness. By contrast, neural

> states can be defined entirely in terms of types of neurons and the times, directions and intensities at which they fire. So, on the sense datum view, to have the orange-experience is not just to have a neural state, since they have different definitions. The orange-experience is dependent on a neural state but it is something more than a neural state.
>
> *Representationalism* (chapters 3-4). To have the orange-experience is to "experientially represent" that something is round and moving to the right. (If someone has the orange-experience in a hallucination, there is in fact no such object - not even a "sense datum".) So spatial terms enter into a definition of what it is to have the orange-experience. This is not true of any neural pattern. So, on representationalism, to have the orange-experience is not just to have some neural state. The orange-experience may be dependent on a neural state but it is something more than a neural state.
>
> *Contemporary naïve realism* (chapter 5). To have the orange-experience is to *either* really experience a physical thing that is orange and moving to the right (in normal perception) *or* to be in a state that is indiscriminable from experiencing such a thing (in illusion or hallucination). Since this is not the definition of any neural state, to have the orange-experience is not just to undergo a neural state. The orange-experience is something more than a neural state.

Let me conclude with some comments.

(i) There is a long history of Leibniz's law arguments against the identification of experience-types with internal neural-types (Smart 1959). For instance, suppose that you smell some mint tea. Against the internal physical state theory, it might be argued that your sensation involves the smell quality *minty*, but your underlying neural state *S* doesn't involve this smell quality (just neuronal firings), so they cannot be one and the same. But the "spatial" Leibniz's law argument above is superior to this traditional kind of Leibniz law argument. For in response to this traditional Leibniz law argument, the internal physical state theorist can say that the neural state *S does* essential involve the small quality minty, because that smell quality *just is* a neural pattern involved in *S* (even if this is not introspectively evident). By contrast, a parallel response to the spatial Leibniz law argument above is not possible. For no one

thinks that the spatial features *round* and *moving to the right* involved in the orange experience turn out to be neural properties involved in the underlying neural state *N!*

(ii) The argument from essential external directedness is only supposed to rule out the *internal physical state view*, which makes the strong claim that all experience-types are *identical with* neural-types. By this argument, experiences and neural states have different *natures*. So even if neural states are sufficient for experiences, experiences cannot be *identical with* neural states. The argument is not meant to rule out experiential internalism – the weaker claim that, for every experience-type, there is an internal neural-type that *necessitates* it. In fact, it doesn't rule out experiential internalism.

To see this, notice that there are views that accommodate experiential internalism but that are not ruled out by essential external directedness. One such view is sense datum theory (see especially box 1.1 in chapter 1). On this view, the orange-experience depends on a neural pattern but it has a different nature from the neural pattern: it is a relation to a round and moving sense datum. So, unlike the underlying neural pattern, the experience is essentially externally-directed. Another such view is "internalist representationalism". We already mentioned this view in passing in section 2.2 and will consider it in detail in section 4.8. On this view, too, the orange-experience depends on the neural pattern but it has a different nature from the neural pattern: it consists in *experientially representing* a round and moving thing. So, unlike the underlying neural pattern, the experience is essentially externally-directed. We will look at this view in detail in chapter 4.

(iii) It may be that the *only* way for internal state theorists to save their view in the face of this argument would be to reject premise 1. If this is right, then the internal physical state view requires that it is *not* the case that spatial terms like *round* and *moving to the left* enter into a correct definition of what it is to have the orange experience (because they don't enter into a definition of what it is to have internal the neural pattern with which this type of experience is identical). It is worth pausing for a moment to appreciate what this implication would amount to.

Return to an example we have already used before: having a diffuse headache throughout your head. Clearly, you can define what it is to have an experience with *this* phenomenal character without using the spatial terms *round* and *moving to the right*. So if the internal state view implies the rejection of

premise 1, then what it implies is that having the orange experience (Figure 1) is like having a headache in this respect. And this amounts to saying that the experience "really" has a character other than the character it seems to have. For it certainly seems to have a character radically different from the character of a headache, a character that can only defined by using *spatial terms* like *round* and *moving to the right*.[9] We will return to this issue in section 2.7 when we consider David Papineau's response to the argument from essential external directedness.

2.6. Could a Brain-in-the-Void Have a Favorite Shape?

The argument of the previous section, if sound, shows that essential external directedness and the internal state view are inconsistent. If the orange-experience is essentially externally directed (Premise 1), but the neural pattern is not (Premise 2), the orange-experience cannot be identical with the neural pattern. Now we will sketch a further argument for thinking that the internal state view and essential external directedness are in tension. The argument concerns our ability to be *mentally related* to *properties* – for instance, our ability to think about shapes.

Let me first explain the idea of a property (van Inwagen 2004, Yi 2018). A property is a *way things might be*. For instance, purple is a way things might be. Properties – ways things might be - are more abstract than ordinary things (somewhat as numbers are). For instance, a specific shade of purple cannot be located in any particular place; if it is anywhere, it is wherever there is a thing with that shade. Properties are not created by the mind, any more than numbers are created by the mind. They are "objective". Even before minds came on the scene, external objects were certain ways: they had certain shapes, distances, orientations, and so on. There are properties that nothing has. For in-

[9] Of course, internal physical state theorists who reject essential external directedness don't have to say that the orange-experience is like a head-pain in *every way*. For instance, they can say that the orange-experience is a neural pattern with *more complexity* than the neural pattern that they identify with the pain. But they do have to say that the orange-experience is like a head-pain in this respect: its essential nature can be fully characterized without mentioning spatial features like *round* and *moving to the right*.

stance, in usual circumstances people can hallucinate unusual colors that nothing has (see section 5.5); those colors still exist because they are ways things *might* be.

By having experiences, we become *mentally related* to properties in various ways. We mentally represent them. For example, when you see an orange, you believe it is *round*. You mentally attribute the property of being round to the orange. You can think about its round shape.

Here now is a question: *how* do you become mentally related to properties in such ways? Can these mental relationships be identified with some kind of physical relationships? Or are they spooky non-physical relationships?

Some theories of experience may be able to explain, in non-spooky physical terms, how we become mentally related to shape properties and other properties. One example is naïve realism (chapters 1 and 5). On this view, to have the orange-experience is just to experience an *example* of roundness in the physical world. You experience the roundness of the orange because your visual system is causally sensitive to the roundness of the orange (via the light coming from the orange). It is only by having such experiences that you are able to think about the property of being round. On naïve realism, then, it is natural to think that *mentally representing* roundness can be identified with a complex *causal* or *informational* relationship to roundness. There is a pattern of firing in your cognitive system that is normally caused by round things in the world. Analogy: a thermometer's ability to represent temperatures is reducible to a causal relationship between its levels of mercury and examples of those temperatures in the world (Figure 4a). This kind of approached has been developed by Fodor (xx), Dretske and Neander.

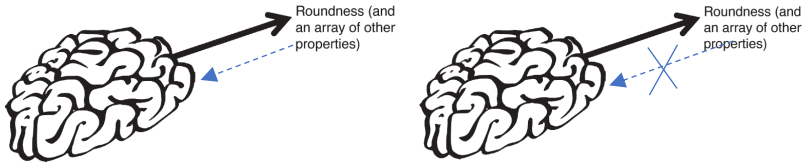

Figure 3: *A*. Some - Fodor, Dretske Neander - think that your brain enables you to think about and mentally represent roundness (solid arrow) because there is a pattern of neural firing that is normally *caused by* round things. *B*. Given the internal state view and external directedness, the BIV represents round without any causal-informational relation to round, contrary to the causal theory of representation.

But now suppose that we reject the externalist approach of naïve realism. Instead, suppose that we accept the internal state view. And suppose that we also accept external directedness (McLaughlin 2016a: 856-857).

In addition to being subject to the Leibniz's Law argument presented in the previous section, it implies that mental representation of properties cannot be explained in causal-informational terms and, indeed, that it is a spooky, non-physical relationship.

The argument for this is based on the *brain in the void* thought experiment. Imagine that, in some other "possible world", there happens to be a brain just like your own, and it happens to undergo the same neural activity as you own brain (Figure 3b).

Now, naïve realists, because they hold that experience is fundamentally a relationship to the world, are not committed to saying that the BIV has any experiences at all. By contrast, internal physical state theorists must say that BIV has exactly the same experiences as you, because it has the same neural states. And, supposing for the sake of argument that they also accept essential external directedness, it seems to the BIV that there are variously-shaped things in space, just as it seems to you. For instance, if you have an experience as of a round orange on a table, then BIV has the same experience as of a round orange on a table.

Therefore, the combination of the internal state view and essential external directedness implies that the BIV can mentally represent the shape *round*. In fact, the BIV could have a favorite shape just like you do. The BIV may not see any items (physical objects or even "sense data") that possess this shape. But the shape *property* still exists – it is a way things might be – and the BIV is mentally related to it.

(Indeed, the BIV *knows* things about shapes. True, the BIV cannot know that there is a round thing in front of it – there is no such thing. Most of its beliefs about the world are false. But the BIV *can* know some more abstract things, such as that round is more like oval than square. Strange as this may seem, hallucination can be a source of knowledge of non-mental reality. More on this in section 5.5.)

Here now is the argument:

(1) If both the internal state view and essential directedness are correct, then an isolated BIV might bear the *mentally represents* relationship to roundness.

(2) But the BIV doesn't undergo a neural state that is normally caused by the occurrence of roundness in the world. Indeed, the BIV is in every way cut off from *any* relevant physical relations to roundness

(3) So if both the internal state view and essentially directedness are correct, then the *mentally represents* relationship the BIV bears to roundness cannot be identified with any *physical* relationship between the BIV and roundness.

Actually, given internal state view and essential external directedness, we can establish that, when *you* view an orange and mentally represent (think about) its round shape, then your *mentally representing* the shape *round* cannot be identified with a physical relationship, such as a causal-informational relationship (the dotted arrow in Figure 2a). For the isolated brain in the vat (Figure 2b) stands in the *same* mental relationship to the shape round, but doesn't bear any causal-informational relationship or other relevant physical relation to the shape round - all such physical relations to roundness have been removed.

As Jeff Speaks (2015: 272) says, the combination of the internal physical state and essential external directedness "would have as a surprising (and presumably unwelcome) consequence the irreducibility of [representational] relations." Likewise, Ned Block says "it may be right that we internalists should acknowledge an irreducible representation relation" (2019: 426).

So the internal physical state view and essential external directedness together entail that brains simply have an innate, intrinsic capacity to mentally represent shapes and other perceptible properties. And here mental representation cannot be explained in physical terms; it cannot, for instance, be explained in terms of causal-informational relation (the dotted line in Figure 3a). This is somewhat mysterious.

If this is correct, it means that the internal physical state view and essential external directedness do not sit together very well. Part of the argument for the internal physical state view is that it provides an attractively simple "reductive physicalist" theory of *experience* (section 2.1). But now we see that, *if it is combined with essential external directedness*, it implies a somewhat mysterious "non-reductive" theory of *mental representation*. This result will be unwelcome to internal physical state theorists (e. g. McLaughlin 2016a and Papineau 2014) who want a reductive physicalist theory of the mind.

To sum up. We have been considering the question of whether internal physical state theorists can accept essential external directedness. We now have two reasons to think that they cannot. In the previous section, we saw that, given essential external directedness, the internal state view can be ruled out by a Leibniz's Law argument. In the present section, we saw that, if they accept essential external directedness, they must accept a strange, non-reductive theory of mental representation of shapes and other perceptible properties – one that is quite antithetical to their reductive theory of experience.

For these reasons, proponents of the internal physical state view are under pressure to reject essential external directedness. [10]

2.7 Papineau's Reply: Rejecting Essential External Directedness

David Papineau (2014, 2016) is a proponent of the internal physical state view who does exactly that: he rejects the essential external directedness (personal discussion). In this way, he avoids the problems raised in the previous sections for the internal state view based on external directedness.

For instance, take the argument from essential external directedness. Papineau accepts premise (2): neural patterns don't essential involve spatial features like *round* and *moving to the right* in any way (they don't essentially involve the seeming presence of items with these spatial features, and they are not essentially "as of" items with these spatial features). To block the argument, he denies premise (1), that is, essential external directedness (or "essential spatial character"): he denies that such spatial features *essentially* enter into the characterization of our visual experiences (they don't essentially involve the seeming presence of items with these spatial features, and they are not essentially "as of" items with these spatial features). That is, he thinks that visual experiences are *like* neural patterns in that spatial features like *round* and *moving to the right* do *not* essentially enter into their characterization. If so,

[10] For more detailed discussion of BIV argument and the Leibniz's Law argument for thinking that the internal physical state view and essential external directedness are incompatible, see Speaks (2015: 271-272); Pautz (2010: 265ff); and Block (2019), and Papineau (forthcoming OUP book).

there is no good "Leibniz's law" argument for thinking that visual experiences are different from neural patterns.

Likewise, Papineau's response to the BIV argument is to deny essential external directedness. So, while he accepts that a BIV could indeed have the same rich visual and other experiences as you (including the orange experience represented in Figure 2), he denies that the BIV would thereby be mentally related to external shape properties. Even though it has all the same rich visual experiences as you, BIV cannot think about any shapes (so BIV cannot have a favorite shape).

In effect, the arguments of the previous sections attempt to demonstrate an inconsistency between the internal state view and essential external directedness. Papineau agrees that they demonstrate an inconsistency. But instead of concluding that the internal state view is false, he concludes that essential external directedness is false, because he thinks that the case for the internal state view is overwhelming. In other words, since experiences are neural states, and since neural states don't essentially involve spatial features like *round* and *moving to the right*.

In the introduction, I started off with two initial assumptions: essential external directedness and some form of internal dependence. I said that the traditional puzzles in the philosophy of perception can be summed up in this way: how can both of these things be true? Papineau is saying that one of our initial assumptions – essential external directedness – turns out to be mistaken. We should accept internal dependence but not essential external directedness. And the view that best explains internal dependence is the internal physical state view.

Papineau doesn't just reject essential external directedness and leave the matter there. Although he denies essential external directedness, he suggests a few points that might be thought to soften the blow of this denial.

(1) Papineau says (2016: 340) that internal physical state theorists can be "quite happy to agree with" the following "alternative" to essential external directedness: necessarily, many of our visual experiences contain *"phenomenal objects"* with various *shapes** and *colors**. For instance, the orange-experience (Figure 2) essentially "contains" a *round* phenomenal object*. So, necessarily, if a thinker has this experience, she is in a position to refer to and think about the property *being round**. Call this *Papineau's replacement thesis*.

Papineau seems to be thinking along the following lines: "We internal state theorists must deny the initially plausible thesis of essential external directedness, but at least we can accept a replacement thesis that *comes close* accommodating essential external directedness". In that case, rejecting essential external directedness *comes to appear more acceptable.*

But does Papineau's replacement thesis really come close to accommodating essential external directedness, making his rejection of that thesis more acceptable?

To answer this question, we first need to understand Papineau's replacement thesis. *On the face of it*, it does *look* close to essential external directedness. For, to formulate it, he uses *similar-looking* spatial vocabulary, such as "round*". But whether it *really is* close to essential external directed depends on what Papineau means by this vocabulary. What in the world does means by a "phenomenal object" and "round*" and so on? In fact, Papineau doesn't really say. So we need to try to figure that out on our own.

One interpretation is a *sense datum interpretation*. Let *N* be the neural state that, on the internal state view, the orange-experience is identical with. On the sense datum interpretation, Papineau holds that being *N* necessarily causes coming-into-existence of an orange and round non-physical sense datum or "visual field region" (Peacocke 2008). This non-physical sense datum is literally *round:* it has *edges that are equidistant from a common point*. And this is what Papineau means by a "phenomenal object": it is just another word for a sense datum or Peacocke-style visual-field region.

But, of course, this this cannot be Papineau's idea, because it inconsistent with the internal state view and its motivations. It is the sense datum theory, which the internal state view was meant to avoid. Further, if this were Papineau's idea, he would have *no* need to reject essential external directedness and move to his replacement claim in the first place, because the sense datum theory fully accommodates essential external directedness (as noted in section 1.6).

Evidently, by "phenomenal object", Papineau must mean some kind of *physical object*, since the whole point of the internal physical state view is to avoid non-physical sense data. But what kind of physical object? One clue is provided by the fact that Papineau says that a *visual experience* "contains" a phenomenal object. Now Papineau holds that a visual experience is just a type

of a neural state, *N*. So he must think that a phenomenal object is some kind of physical object that is "contained in" neural state *N*. What could this be?

One possibility is that Papineau has in mind something like a *population of neurons*. Let us adopt this interpretation.

Now we can also figure out what Papineau means by his starred vocabulary, such as *round**. He uses this expression to characterize the phenomenal object contained within a visual experience. We have just seen that for Papineau such a phenomenal object must be something like a population of neurons. So *round** must refer to a property *P* that characterizes such population of neurons, as it might be, the property *firing in pattern Z*. That is, for Papineau, "is round*" doesn't literally mean *being round*, that is, *having edges equidistant from a common point*. It refers to a radically different kind of property.

It follows that Papineau's use of spatial vocabulary like "round*" is somewhat misleading. He is not using it to pick out properties with a spatial structure, like *having edges equidistant from a common point*. So even if he packages his replacement claim in vocabulary that looks similar to that used to formulate essential external directedness, it is in fact a radically different thesis.

In the end, here is what Papineau's replacement thesis amounts to:

> *Papineau's replacement thesis*. Necessarily, the orange-experience is nothing but a neural state *N* that "contains" a *neuro-computational* object with internal physical properties P1, P2, . . . (as it might be, *firing in pattern Z*). So having this experience only gives thinkers the capacity to think about these neural properties. This is true *even when a brain in the void* has the experience.

Now that we have clarified Papineau's replacement thesis, we are ready to answer our question: does Papineau's replacement thesis make the internal state theorist's reject of essential external directedness any more acceptable? Does it allow the internal physical state theorist to say that there is *a sense* in which having the orange-experience essentially involves the feature *round*.

According to essential external directedness, having the orange-experience necessarily involves the seeming presence of an item that *is round*, *that is, has edges roughly equidistant from a common point*. Here, in describing how things

seem, "round" picks out a genuine spatial feature – *it certainly doesn't pick out neural property P*.

As we saw in section 2.4, essential external directedness just seems obvious. It just seems to be an obvious commentary on what the orange-experience is like. It is also needed to explain things we all accept. For instance, we are all inclined to accept the following conditional (if-then) claims: *if* a BIV had all the same experiences as you, but no items with shapes were present, its experiences would be non-veridical. If we accept essential external directedness, we can explain this: *it would seem to the BIV that there are items out there with various shapes* when this is not so. Further, *if* a BIV had *all* the same experiences as you, it could think and know about shapes. For instance, it could know Euclidean geometry. And it could have a *favorite shape*. These conditional (if-then) claims are hard to deny. But they presuppose essential external directedness. So internal physical state theorists like Papineau who reject essential external directedness must deny them.

There is, then, reason to think that Papineau's replacement claim may not make the denial essential external directedness any more acceptable. It doesn't change the fact that it requires the denial of these obvious-seeming truths. The replacement claim just adds: "instead, having the orange-experience only necessarily involves a neural object with neural properties *P1*, *P2*, . . ." Because this replacement thesis is not very similar to essential external directedness – it is not a close replacement to the real thing - it doesn't make the denial of essential external directedness any more palatable.

(2) Here is a second point of Papineau's which might seem to soften the blow of the internalist state theorist's rejection of essential external directedness: he notes that internal physical state theorists can at least happily accept what we might call *"inessential* external directedness" (2014: 24).

For instance, let *N* be the neural pattern that is necessarily identical with the orange-experience-type *E*, on Papineau's view. Papineau can say that *N* (and hence *E*) is "as of" a round item, and involves there seeming to be a round thing, in the sense that it is normally caused by the presence of a round thing. So he can say that *N* (and hence *E*) is connected to the shape *round*, but in a way that doesn't touch on its essentially nature.

Here is an analogy. Suppose some aliens live on another planet with sudden hailstorms. They get a distinctive head pain when and only when a round hail pellet hits their heads. When the aliens have this head pain, there is a

sense in which they have an experience as of a round thing, and they are experiencing as if a round thing is present: for they are then having the kind of experience that normally goes with the presence of a round thing hitting their head. Papineau's point is that internal state theorists can at least accept that visual experiences are contingently connected to spatial features, in the way that that the aliens' head-pain is. So we can indirectly or obliquely describe experiences by referring to spatial properties.

However, accepting inessential external directedness may not help internal state theorists very much. For the fact remains that they still must reject *essential* external directedness. And this is a problem because essential external directedness seems obviously right. As we have noted, visual experiences *differ* from head pains *in the following respect:* while spatial features (*round, moving to the right*) don't *essentially* enter into the characterization of a head pain, they do *essentially* enter into the characterization of a normal visual experience. If visual experiences really were purely internal states akin to head-pains, then we should be no more inclined to think that spatial features essentially enter into their characterization than the aliens in the above example are inclined to think that they enter essentially into the characterization of head-pains.

(3) I said that Papineau thinks that the only possible response to the argument from essential external directedness against the internal state view is to reject essential external directedness. However, in his work, he gives the impression that internal state theorists only need to reject the controversial *representational view* of experience. We will look at this theory in detail in the following chapter. Briefly, on this theory, having the orange-experience essentially involves "representing" the presence of a round thing. Papineau also associates this theory with the somewhat obscure claim that experiences "lay claim" to the world (2016: 341).

If essential external directedness is equated with a controversial and obscure *theoretical* claim about experience, then it would no longer look so bad if internal state theorists must reject it.

But it would be wrong to equate essential external directedness with representationalism or any other theoretical claim. It just says that having the orange experience essentially involve the seeming-presence of a round item moving to the right. This is not a *theory;* it is a pre-theoretical claim formulated in non-theoretical language that is compatible with a number of theories:

sense datum theory, the versions of naïve realism, and all the versions of the representational view.

We have critically evaluated three points of Papineau's that might seem to soften the blow of the internal state theorist's need to reject essential external directedness. In fact, they don't seem to much soften the blow.

But what if internalist state theorists like Papineau still decided to reject it? For instance, what if they said, contrary to essential external directedness, that the orange-experience (Figure 2) is like a head-pain in the following respect: since it is just a neural pattern, its essential nature can be fully characterized without mentioning spatial features like *round* and *moving to the right*.

Against this, it will seem obvious to many that that this is incorrect. It is also worth emphasizing again that, in the history of philosophy, nearly all major theories of experience (naïve realism, the British empiricists' "theory of ideas", sense datum view, Peacocke's related visual field view, representationalism, multiple relation theory) have accommodated essential external directedness, even if they have provided different theoretical accounts of it.[11] In fact, while Papineau rejects it, other internal physical state theorists such as McLaughlin (2016a: 856-857) are favorable towards it (even if the previous two sections brought out problems with combining the internal state view with essential external directedness). The widespread acceptance of essential external directedness further testifies to its truth.

---

[11] It might be thought that there are figures in the history of the philosophy who did reject essential external directedness. One candidate is Thomas Reid (1785). But while it is true that Reid thought that phenomenal character is partly determined by "sensation" which is not essentially externally directed, he held that it is also determined by "conception" which *i*s essentially external directed. See van Cleve (2005: 468). So, given Reid's view, it is natural to take "the orange-experience" to refer to a hybrid state involving a color sensation and a conception. In that case, it essentially directed at a round thing. Another candidate is Chisholm (1957). Chisholm is often called a proponent of "adverbialism" or the "multiple relation theory". (See Jackson 1977: 63, 90.) But Chisholm's main point was to reject "the sense datum fallacy" (1957: 151). He denied that having the orange-experience essentially involves the presence of a round sense datum. But he nowhere explicitly denied essential external directedness: that it essentially involves there *seeming* to be a round item. (Thanks to James van Cleve for discussion of these matters.)

Maybe it would be reasonable to accept the internal physical state view and reject essential external directedness if there were decisive problems with *all* the alternative views that accommodate essential external directedness: the sense datum view, representational view, the theory of appearing, naïve realism, and so on. Papineau (2014, 2016) thinks that the representational view, in particular, faces debilitating problems. We will address the representational view, and the problems for it, in the next chapters.

Summary

The sense datum view (chapter 1) provided a neat solution to the puzzle of how experiences can be essentially externally directed and also dependent on internal processing. But it required strange non-physical items, sense data. The desire to avoid sense data led naturally to the internal state view examined in the present chapter. Rather than holding that experiences are relations to non-physical sense data *created by* neural states, internal state view holds that experiences should be directly identified with *neural states themselves*. It fits nicely with empirical findings about the role of internal factors in shaping experience. It is also in line with the worldview of "reductive physicalism" and has the virtue of simplicity. However, there is one major rub: it is inconsistent with the essentially externally directed character of some types of experiences.

Therefore, we are still without a totally satisfying solution to the puzzle of perception that started off this book. What we need is a theory that simultaneously accommodates the role of internal factors as well as the essentially externally directed character of experience, while avoiding the postulation of "sense data".

While we do not have a solution, we have made significant progress. In the present chapter, we emphasized essential external directedness. For instance, having the orange-experience essentially involves the seeming presence of a round thing moving to the left. In the previous chapter, we learned that, in some cases, even if it seems to you that a round item is present, no such item really is present – not even a "mental image" or "sense datum". Putting these lessons together, we arrive at the result that some experiences essentially involve the seeming presence of an *F* item, even if they do not essentially involve the real presence of an *F* item. As Fred Dretske (2003) puts it, "there

needn't be anything orange or pumpkin-shaped in (or outside) the head at the time the experience is occurring in order for us to have an experience as of an orange pumpkin". If we further accept the existence of "properties" (see 2.6), this means that experiences relate us to the property of being *F* even if they do not necessarily relate us to an item that hast the property of being *F*.

Now here is something interesting: this is a feature that perceptual shares with *mental representation*, as when you represent the world to be a certain way in thought. For instance, it is in the nature of thoughts that they are externally directed at things, and involve the attribution of properties to things. Thoughts can also misfire: you can think that something is *F* even if nothing really is *F*. This suggests an intriguing idea: perhaps perceptual experience should be understood as a species of mental representation. This would explain why they are essentially externally directed, without requiring "sense data". We might then add that in some cases how we perceptually represent the external world is due to our own internal processing, rather than to the character of the world itself. This would explain why experiences are internally dependent as well as externally directed.

In this way, our discussion so far naturally leads to the representational view. We have already mentioned this alternative to the internal physical state view in the present chapter (sections 2.2 and 2.5). It will be the subject of our next two chapters. We will first see (chapter 3) that it can accommodate the essentially directed character of experience without sense data". Then we will see (chapter 4) that it may also be able to accommodate the fact of internal dependence.

Further Reading

For recent defenses of the internal physical state view, see Papineau 2014 and 2016, McLaughlin 2016a, and Block 2019. The view derives from Place 1956 and Smart 1959. It is associated with an "adverbialist" account of our descriptions of experience (see fn. xx). For a recent defense of adverbialism, see Breckenridge 2018.

In this chapter we looked at the much-discussed "transparency observation". It comes in many different versions. Some sources are Moore (1903: 449-450), Price (1932: 5), Geach (1957: 126-128), Armstrong (1981: 85) and Harman (1990: 39). For some recent defenses of different versions, see Tye 2000 and Byrne 2018. For criticism, see Kind 2003.