

INNER AWARENESS: THE ARGUMENT FROM ATTENTION

Anna Giustina & Uriah Kriegel

Forthcoming in *Philosophical Studies*

Abstract

We present a new argument in favor of the Awareness Principle, the principle that one is always aware of one's concurrent conscious states. Informally, the argument is this: 1) Your conscious states are such that you can attend to them without undertaking any action *beyond mere shift of attention*; but 2) You cannot come to attend to something without undertaking any action beyond mere shift of attention unless you are already aware of that thing; so, 3) Your conscious states are such that you are aware of them. We open by introducing more fully the Awareness Principle (§1) and explicating the crucial notion of "mere shift of attention" (§2). We then develop the argument more fully, first in an intuitive form (§3) and then more formally (§4), before replying to a series of objections (§§5-7).

1. The Awareness Principle

Philosophical theories of consciousness differ along various axes. A central one concerns the *Awareness Principle*:

(AP) For any subject *S*, conscious mental state *M*, and time *t*, if *S* is in *M* at *t*, then *S* is *aware* of *M* at *t*.

Higher-order and self-representational theories of consciousness (e.g., Rosenthal 1997; Kriegel 2009) accept AP; other theories reject the principle, notably first-order representationalism (e.g., Dretske 1995).

The purpose of this paper is to present a new argument in favor of AP. Informally, the argument is this: 1) Your conscious states are such that you can attend to them without undertaking any action *beyond mere shift of attention*; but 2) You cannot come to attend to something without undertaking any action beyond mere shift

of attention unless you are already aware of that thing; so, 3) Your conscious states are such that you are aware of them. The purpose of this opening section is to get AP more clearly into view. To start, it might be useful to distinguish it from various neighboring theses, as AP is weaker than some and stronger than others.

Some philosophers have held that every conscious state is the object of explicit, *attentive* introspection by the subject (Armstrong 1968). This is not implied by AP, however. AP only claims that the subject is *aware* of each of her conscious states. But some awareness is non-attentive (as when, reading in a café, you are auditorily aware of music in the background, even as your attention is focused on your book). For all AP says, we may often or even *typically* be *non-attentively* aware of our conscious states—aware of them despite not *attending* to them.¹

Some philosophers hold that subjects have a conceptual, propositionally structured awareness of their concurrent conscious states – a *thought* about them (Rosenthal 1997). This too is not implied by AP, insofar as some awareness may well be non-conceptual and/or non-propositional: you may be visually aware of an armadillo without being aware of it *as* an armadillo and without being aware *that* it is an armadillo (*cf.* Dretske 1993). For all AP says, our awareness of our conscious states may be like this: S is aware of *M*, but S is not aware of it *as M*, and nor is S aware *that* she is in M (or *that M* occurs).

It is sometimes debated in the relevant literature whether S's awareness of M must be awareness of *being* in M, and therefore also awareness of S. David Rosenthal (1997), for instance, argues that the relevant awareness is an awareness of the particular *token* M, not just the *M-type*, and that there is no way to be aware of a *token* conscious state without being aware of the conscious subject of whom it is a state. Our own impression is that this depends on the metaphysics of states one is working with. If one takes states to be instantiations of a property by a substance at a time (compare Kim 1976 on *events*), then it would indeed seem impossible to be aware of a token state

¹ There is a way to use the term “attentive” such that a representation that mobilizes *any* amount of attention resources is “attentive” (Sebastian Watzl (2017), for example, argues that attention is the foreground/background structuring of the stream of consciousness, and that all conscious experiences are so structured; in this framework, every conscious experience is “attentive”). In that use, if you devote 90% of your attention to A and 10% to B, your awareness both of A and of B is “attentive.” We are using the term differently, to denote awareness that mobilizes *most* of the subject’s attention resources. Of course, all our claims could be translated into this other usage (e.g., instead of saying that AP requires only non-attentive awareness we would say that it requires only slightly-attentive awareness).

without being aware of the substance of which it is the state. But if one takes states to be special particulars (here the comparison on events would be to Davidson 1967), then it may be possible after all for S to be aware of M without being aware *of herself*; for in this picture S is not part of what makes M the token state it is. There is clearly no need for the proponent of AP to take a stand on the metaphysics of states *qua proponent of AP*. If you are particularly attracted to the property-instantiation view of states, then for you there would be a direct inference from AP to the stronger claim that whenever S is in M, S is aware of *being* in M (and/or of herself). But such claims do not fall out of AP itself.

AP is nonetheless quite a strong claim. It is stronger, for example, than the thesis that if S is in conscious state M, then S is “in a position” to be aware of M—where S is “in a position” to be aware of M just in case if S *considered* whether she is in M, then S would come to be aware of M (*cf.* Shoemaker 1994). AP is stronger in claiming that S is not just “in a position” to be aware of M, but actually *is* aware of M.

Similarly, some authors have claimed that conscious states are such that the subject is always *disposed* to be aware of them (Carruthers 2000 may be interpreted this way). But AP claims something stronger, namely, that one always has an *occurrent* awareness of one’s concurrent conscious states.

We hope that this series of contrasts gives more definition to AP. AP claims that whenever S is in conscious mental state M, S has actual, occurrent awareness of M, which may or may not be explicit, attentive, conceptual, or propositional and may or may not imply awareness of *being* in M.

Why believe in AP? Advocates often claim it to be phenomenologically evident (Goldman 1970; Zahavi 2005; Kriegel 2009). But AP-deniers too appeal to phenomenological considerations, notably the so-called transparency of experience (Harman 1990). It would appear the dispute needs to be settled on *theoretical* rather than phenomenological grounds. But arguments for AP are hard to come by in the extant literature, though they do show up occasionally—see notably the “Memory argument,” discussed e.g. in Williams 1998; Ganeri 1999; Perrett 2003; Thompson 2011; Kellner 2010; Kriegel 2019; Giustina 2022. To strengthen the case for AP, here we propose a new argument for it—we call it the *argument from attention*.

A cousin of the argument we will present is critically discussed by Daniel Stoljar (2023).² However, the argument Stoljar considers is different in a crucial detail from ours, a detail that proves dialectically very significant. We will discuss the difference at the end of §4. The argument we will present has not, to our knowledge, been discussed anywhere in print.³

The plan moving forward is as follows. First we elucidate the notion of “mere shift of attention,” which is central to our argument (§2). We then present and develop the argument more fully, first in an intuitive form (§3) and then more formally (§4). We close by considering a series of objections (§§5-7).

2. Mere shift of attention

Before presenting the argument, let us clarify a notion that will play a crucial role in it—the notion of “mere shift of attention.”

Shift of attention occurs whenever the focus of one’s attention moves from one object or event to another.⁴ Attention shifts often occur within perceptual awareness. Listening to a jazz piece, you may be auditorily aware of the trombone, the piano, the trumpet, and the bass, and the focus of your attention may roam from one instrument to another (or distribute evenly). This is a case of *intramodal* shift of attention. Sometimes attention shift occurs *intermodally*: you may shift your attention from the *sound* of the trumpet to the trumpeter’s virtuosic fingerwork, of which you are visually rather than auditorily aware. Attention shift is not restricted to the perceptual domain: it is possible to shift one’s attention from the jazz concert to one’s giddy mood, to one’s desire for a beer, or to some referee objection one is trying to figure out. In all these cases, one is still aware of the jazz, but now less attentively, having shifted attention away from it.

Often when one shifts attention from one thing to another, one must do so *by doing something else* (or at least *along with* doing something else). For example, while strolling in a museum gallery, one may shift one’s attention from one painting to

² Chalmers (2013) sketches a more distant relative of our argument, involving attention and phenomenal change; however, his argument supports the thesis that we can attend to our experiences (not just the objects thereof), rather than AP.

³ Another, more distant cousin of our argument is developed by Sebastian Watzl (2017), but that argument is subtler and its relationship to our own is less straightforward. Unfortunately we don’t have the space to explore these connections here. For relevant discussion see Chaturvedi (2022).

⁴ This does *not* presuppose a “spotlight” view of attention: on a “structuring” view (Watzl 2017), for example, shifting the focus of attention from A to B (roughly) amounts to B coming to the foreground and A receding to the background.

another *by turning one's neck*, or at least by moving one's pupils askance. Sometimes, however, it is possible for one to shift one's attention from *x* to *y* *without doing anything else*, for instance when one shifts one's auditory attention from the trombone to the piano while listening to the jazz piece (with eyes closed, say); this is what we call *mere* shifting of attention.

Attention shift may be *active* and *intentional*, as in the cases just described, but it may also be *passive*, a matter of one's attention being (externally) *drawn* rather than (intentionally) *directed* to something. (Cognitive scientists sometimes call the former "endogenous" attention and the latter "exogenous.") Attention shift is active when one forms the *intention* to change the focus of one's attention and acts on that intention; it is passive when one's attention is *captured* by something (e.g., a sudden loud noise), or drawn to something by *someone's intervention* (e.g., being told that a situation in the other room requires one's attention).

The active/passive distinction applies also to *mere* shift of attention. When you shift your auditory attention from the trombone to the piano while listening to jazz, you perform an *active* mere shift of attention. But there is also such a thing as *passive* mere shift of attention. In the case of passive attention shift, there are no intermediary *actions* in terms of which mere attention shift could be defined. Still, we might say that passive mere attention shift occurs when your attention shifts from *x* to *y* without you *undergoing any event* in virtue of which you come to attend to *y*—for instance if you hear a sudden loud noise and your attention is immediately drawn to the noise without anything *else* happening that *makes* your attention shift.

In any case, what will play a crucial role in our argument is the notion of *active* mere shift of attention. *Active* mere shift of attention occurs when one intentionally shifts one's attention from *x* to *y* without performing any action in virtue of which one comes to attend to *y*—beyond, that is, forming the intention to attend to *y* and acting on that intention to actually shift one's attention. We might express the propositional content of the relevant intention with something like "I shall attend to this" (or, if we prefer an imperative, to capture the action-guiding force of intention, perhaps "Attend to *this!*"). What we will argue is that one cannot come to attend to something through active mere shift of attention unless one is already aware of that thing, and that all conscious states are such that one can attend to them through active mere shift of attention.

It might be objected that, if the content of the intention to attend to *y* is demonstrative (“Attend to *this!*”), then forming the relevant intention requires not mere awareness but *attentive* awareness of *y*. The objection relies on the idea that demonstrative thought requires attention: for a subject to refer demonstratively to *y*, she must consciously attend to *y*. This is something that has been argued by a number of philosophers, including Campbell (2004), Dickie (2011), and Smithies (2011). The problem here is that if attention is needed to make sense of demonstrative reference then demonstrative reference cannot be used to make sense of attention shift, on pain of circularity—on so goes the objection.

Our response is twofold. First, while extant arguments in this area tend to show that some selection mechanism is needed for demonstrative reference, it is often assumed that attention is what will provide the relevant selection. But it has recently been argued that this assumption is misguided and that not all selection mechanisms are forms of attention (Wu 2023). So a non-attentional selection mechanism may well enable demonstrative reference, which would then enable the kind of attention shift we have in mind.⁵

Secondly, the idea that demonstrative reference requires prior attention to the demonstrative’s referent has been forcefully criticized (Martin 1997; Kelly 2004), and in fact some attention theorists have argued that, on the contrary, demonstrative reference is a *precondition* for attention (Mole 2011; Wu 2014): it is demonstrative reference that enables attention rather than attention that enables demonstrative reference. So, the claim that, necessarily, demonstrative thought requires prior attention is controversial. Furthermore, and more importantly, there are reasons to think that *not all* demonstrative thought requires prior attention. If attention were necessary for all demonstrative thought, it would be just impossible to intend to shift attention via demonstrative content: one could, at most, form the intention to shift attention to whichever *y* satisfies a given description (as in “Attend to what is behind my back” or “Attend to the woman with a red hat in the crowd”). But one can certainly also intend to attend to “this,” or to “this F,” as when you form the intention to shift your attention from the book you are reading to *this* mild backache that has been annoying you for the past hour, or from *this* part of the painting depicted in the book to *that* part.

⁵ We thank an anonymous referee for *Philosophical Studies* for suggesting this line of response to us.

Arguably, what is needed to enable demonstrative reference is not *attention* but *awareness*: plausibly, all that is needed for one to form the intention to shift attention from *x* to *y* is prior *awareness* of the *y* to which attention is intended to be shifted.⁶

To repeat, the reason to think it is awareness rather than attention that is needed to enable demonstrative thought is that if all demonstrative thought required prior attention to the demonstrative's referent, we would have no control over the way our attention moves over the array of items we are aware of, since intentional change of attention would be impossible—our attention would be prey to whatever *grabs* it, and every shift of attention from one item in consciousness to the next would be *passive* and exogenous. But we *can* control the way our attention moves around: we can *actively* decide to shift attention from one item to another just by forming the intention to attend to “*this*” or “*that*” item.

Someone might deny that this is something we can genuinely do: actively decide to shift attention between items just by forming a demonstrative attentional intention. Our first-person impression is that this is something we can and often do do. But given the elusiveness of first-person disagreement, for someone who denies this we also offer our previous response to the worry: that for all extant arguments have shown, a selection mechanism is all that may be needed for demonstrative reference – whether or not that mechanism constitutes a form of attention.

Our notion of mere shift of attention only partially resembles the notion, sometimes discussed in the philosophy of cognitive science, of *covert* (as opposed to *overt*) attention. Attention to *x* is *overt* when one comes to attend to *x* via some bodily movement aimed at orienting one's sense organ toward *x* (e.g., turning one's neck). When, instead, one comes to attend to *x* without any such bodily movement, attention is said to be *covert*. (Deliberately shifting attention across the visual field while keeping one's eyes fixed is an example.) This characterizes what we called mere shift of attention as well. However, while every mere shift of attention is covert, not every covert attention shift is a *mere* attention shift. For some attention shifts may be accompanied by no bodily movement but still involve intermediary *mental* actions or events. For instance, one may shift one's attention from screaming sounds at the

⁶ Within a framework that assumes a usage of the term “attentive” such that even a representation that mobilizes a minimal amount of attention resources is “attentive” (*cf.* fn. 1 above), the claim is that forming the intention to devote *highly-attentive awareness* to *x* requires prior *slightly-attentive awareness* of *x*.

neighboring apartment to the thought that one should call the police by inferring that someone may be in danger, reasoning that intervention may be needed, and so on; or one may shift one's attention from the smell of apple pie to the thought of one's grandmother via a series of mental associations.

So, mere shift of attention is not the same as, but rather a proper subset of, covert shift of attention; and it can occur either actively/endogenously or passively/exogenously. *Active* mere shift of attention from x to y occurs just if one moves one's attention from x to y without undertaking any (bodily or mental) action other than forming the intention to attend to y and acting on it, thus actually moving the focus of one's attention to y . This is the notion at play in our argument.

3. The Argument from Attention: The Intuitive Idea

Our argument for the Awareness Principle (AP) is based on two intuitive ideas. The first is this. Suppose you are sitting in a dentist's waiting room. In the normal go of things, you are attentively aware of the external-world *objects* of your experience: the coffee-table magazines, the other waiting patients, etc. But you can also—essentially at will—shift your attention to aspects of the conscious state you are in, such as your *irate mood*, your *toothache experience*, or *the way the coffee-table looks to you*—too aggressively orange, perhaps. Our central contention is that all a person in such a circumstance ordinarily needs to do, in order to come to attend to their concurrent conscious state, is perform what we called *active mere shift of attention*. There is no additional, intermediary action you must perform in order to move from attending to the orange chair to attending to your mood (say). Mere shift of attention, entirely unaccompanied by other actions (bodily or mental), can suffice to make you attentively aware of your conscious state.

An immediate objection is that the claim that we can always attend to our conscious experiences by active mere shift of attention seems (at least *prima facie*) to be in tension with the “transparency of experience.” We will discuss this more fully in §5, but briefly, our take is this. Since Harman's (1990) seminal article on transparency, many increasingly fine-grained distinctions have been introduced between different possible interpretations of the idea (Kind 2003; Siewert 2004; Gow 2019; Bordini 2023). As we will endeavor to show in §5, our argument is consistent with many of the

ideas commonly associated with the label “transparency of experience,” and only quite extreme interpretations are inconsistent with it.

It may also be objected that, for one to *actively* shift one’s attention to (as opposed to one’s attention being passively captured by) a concurrent conscious experience of one’s, one needs at least to undertake the mental action of *forming the intention* to attend to such an experience. This is true but does not make the shift of attention at issue “non-mere.” For, as defined here, active mere shift of attention occurs when one moves one’s attention from *x* to *y* without undertaking any action *other than forming the intention* to attend to *y* and *acting on that intention*. There is a reason for carving out this exception. Forming the intention to shift attention is not an action *by which* one shifts attention the way, say, turning one’s neck is. Rather, forming the relevant intention is just required for the shift of attention to be *active*, implicating the *intentional action* of moving the focus of one’s attention from *x* to *y*.

The first intuitive idea behind our argument, then, is that whatever conscious state you are currently in, if you are not already attending to it, it should be possible for you to shift your attention to it without having to do anything else *by which* you would be shifting your attention. That is, it should be possible for you to come to attend to your conscious experience via active mere shift of attention. We will make the case for this in §5.

The second intuitive idea is that you cannot come to attend to something through active mere shifting of attention *unless you are already aware of that thing*. To get a sense of the idea, consider the case of active mere attention-shift from the previous section: you are listening to a jazz piece, attending first to the trombone, and then you actively and intentionally direct your attention away from the trombone and onto the piano. We may imagine that you are listening to this piece through your earphones while lying in bed with your eyes closed, unwinding after a long day of teaching. You are doing nothing with your body and your mind is focused only on the music and is otherwise about as blank as a human mind can be. When you shift your attention from the trombone to the piano, then, you are performing a mere shift of attention in our sense: you shift your attention without doing *anything* else. Here it is hard to imagine you could have had no auditory awareness whatsoever of the piano before shifting your attention onto it. For one thing, if you had listened to the “same” piece performed but without the piano component, your overall conscious experience of the piece would

have presumably been quite different. But even more fundamentally, it is hard to understand how you could intentionally direct your attention *onto* the piano-sound without being aware of there even *being* a piano-sound. Clearly, in this case your attention-shift presupposes a prior awareness, albeit non-attentive, of the piano-sound.

To be sure, coming to attend to something via *passive* mere shift of attention does *not* imply that one is already aware of it. When your attention is captured by a sudden loud sound, you undergo a mere shift of attention—the focus of your attention moves to the sound without you undergoing any intermediary event conducive to your attending to the sound—yet the sound is not something you were aware of before. However, the reason this is possible in the case of *passive* mere shift of attention is that the sound did not *exist* before the attention shift. It is precisely the loud sound's sudden *appearance* into your awareness that captures your attention. In contrast, when a person *actively* shifts attention onto *y*, *y* must already be in one's awareness. So, while passive or exogenous mere shift of attention does not imply antecedent awareness of what one shifts one's attention to, *active* or *endogenous* mere shift of attention does.

In general, our conscious field (i.e., the overall state of consciousness we are in at any one time) appears to be characterized by a certain center/periphery structure. In typical circumstances, our conscious awareness covers a dazzling multitude of objects and features, only a small minority of which manage to occupy the *center* of our awareness. That minority is in the *foreground* of our overall awareness, while much else is in the background or periphery—what William James (1890) called “the fringe of consciousness.” It is plausible to regard the distinction between center and periphery, foreground and background, as constituted ultimately by the distribution of attention (see Kriegel 2009; Watzl 2011): the “center” of our conscious awareness is simply whatever we are specially *attentively* aware of; the “periphery” is whatever we are *non-attentively* aware of (or: the center is what we are *more* or *highly* attentively aware of and the periphery what we are *less* or *slightly* attentively aware of). In this picture, the kind of bare attention-shift we called *active mere shift of attention* would appear to consist in a certain *restructuring* of one's field of conscious awareness: where initially the subject was attentively (highly-attentively) or “focally” aware of *x* and non-attentively (slightly-attentively) or “peripherally” aware of *y*, the mere attention shift yields—or perhaps just *is*—a change whereby the subject is attentively or “focally”

aware of *y* and non-attentively or “peripherally” aware of *x*.⁷ But both before and after the attention-shift, the subject is aware both of *x* and of *y*; the subject has merely restructured her conscious field so as to foreground what was previously in the background and background what was previously in the foreground. This, arguably, is what *mere* shift of attention consists in (we are tempted to say: is the *only* thing active mere shift of attention *could* consist in). If this is right, then it is in the very nature of active mere shift of attention that that-which-the-subject-comes-to-attend-to through the shift is something the subject was already aware of *before* the shift.

By way of summary of this discussion, consider that one can relate to something in three different ways, as far as awareness of that thing is concerned: either (i) one is *attentively* aware of *x*, (ii) one is *non-attentively* aware of *x*, or (iii) one is altogether *unaware* of *x*. Just now, you may have been *attentively* aware of the words you were reading, *non-attentively* aware of the soles of your shoes, and *unaware* of the window behind your back. Our claim is that in such a situation, you can come to attend to the soles of your shoes—more or less at will—through active *mere* shift of attention; but you cannot come to attend *to the window behind you* through active mere shift of attention. You can certainly come to attend to the window behind you, and it may perfectly well be rather effortless to do so. But it *would* require some additional action *by which* you would come to be aware of the window. (This is so not simply because the window is invisible to you. Even coming to attend to the window *imaginatively*, as opposed to *perceptually*, would be impossible through mere shift of attention, as we will show §6.2)

We conclude, then, that we cannot come to attend to something through active mere shift of attention unless we are already aware of that thing. This general principle acquires special significance when we remember the relatively straightforward observation that we seem able to come to attend to our concurrent conscious experiences without doing anything else, that is, through active mere shift of attention. For then the following piece of reasoning emerges:

⁷ As mentioned above (fn. 1), all our claims could be translated into the alternative usage of the term “attentive” such that a representation that mobilizes *any* amount of attention resources is “attentive.” Accordingly, what we call “attentive” awareness is “highly-attentive” awareness on the other usage, and what we call “non-attentive” awareness is “slightly-attentive” awareness. In the remainder, we omit the double label for ease of exposition.

- 1) Conscious states are such that their subject can come to attend to them without undertaking any action beyond active mere shift of attention;
- 2) A subject cannot come to attend to something without undertaking any action beyond active mere shift of attention unless the subject is aware of that thing; therefore,
- 3) Conscious states are such that their subject is aware of them.

This is the basic, intuitive idea behind what we call the *argument from attention*. In the remainder of the paper, we refine the argument through a series of clarifications and precisifications.

4. Some Clarifications

The argument concerns conscious states. But the term “conscious” is notoriously multiply ambiguous. Our interest here is in *phenomenally* conscious states—mental states of *subjective experience*, states in which there is something it is like to be. And of course we have in mind *present*, not past or future, phenomenally conscious states—it is the subject’s *concurrent* conscious states she can attend to by mere shift of attention. To attend to a *past* experience, the subject would need at the very least to also *recall* that experience; and to attend to a *future* experience, if such a thing is even possible, she would at least have to also *imagine* it.

There are important debates within philosophy of mind concerning the reach of phenomenal consciousness. On some views, phenomenality is the exclusive province of broadly sensory states, notably perceptual experiences and pleasure/pain. On other views, it extends also to conscious thoughts (which have a “cognitive phenomenology”), conscious decisions and actions (with a “conative” or “agentive” phenomenology), and other non-sensory mental states. Here we take no stand on the *scope* of phenomenality: if only sensory states exhibit such, then our argument’s conclusion will apply only to sensory states; if phenomenality is shared more widely, the argument will apply to some non-sensory states as well.

Both premises of our argument make use of the locution “coming to attend.” What we have in mind here is transitioning *into* a state of attentive awareness. As we use the term, to “attend” to something is to have attentive awareness of it. This attentive

awareness is the product of a certain process, and “coming to attend” is our way of referring to that process. What is this process? As noted, a subject can relate to something in one of the following three ways: attentive awareness, non-attentive awareness, and unawareness. To “come to attend” to x is to move from either non-attentive awareness or unawareness into attentive awareness of x .

Importantly, it is not one of our assumptions that the distinction between attentive and non-attentive awareness is sharp and categorical. Attention comes in degrees, and there are no theory-independent bright lines to track here between a “fully attentive” and “fully inattentive” awareness. All the same, conscious awareness does have a center/periphery or foreground/background structure, and it is the distribution of attention that structures it that way (Watzl 2011). This is *all* the distinction between attentive and non-attentive awareness presupposes: awareness is attentive when its object is in the experiential foreground or center, non-attentive when it is in the experiential background or periphery.

Both premises of our argument are explicitly modal: they concern what a subject can and cannot do. What kind of modality is involved here? At the very least, the “can” and “cannot” of Premises 1 and 2 express *nomological* possibility and impossibility. Nomological possibility is a matter of consistency with the laws of nature. Here the relevant laws are the laws of psychology and perhaps neuroscience. The idea of Premise 2, thus interpreted, would be that attending to something of which one was previously completely unaware by active mere shift of attention is inconsistent with the neuro-psychological laws of nature: these laws are such as to require more than mere shift of attention if a subject is to actively attend to something they were previously completely unaware of. For instance, the neuro-psychological laws allow you to come to attend to the piano sounds in the jazz concert merely by shifting your attention from the trombone sounds you have been focusing on so far; but they do not allow you to come to attend to some sounds you have so far been completely unaware of merely by shifting your attention. The reason, arguably, is that to form an *intention* to attend to something (as required by *active* mere shift of attention), one needs to be somehow aware of that which one’s intention is directed at: it is psychologically impossible to form an intention about something you are utterly unaware of.

In fact, it is perhaps even *conceptually* impossible to actively shift one’s attention from x to y without being previously aware of y . For the very notion of “intending to

attend to *y*” seems to imply that, to form such an intention, one has to be aware of *y* in some way: plausibly, if you are to form an intention to attend to *y*, *y* needs to be somehow present to your mind. However, our argument does not need a stronger modality than nomological, so that is the modality we will work with.⁸

It might be felt that the “can” and “cannot” of our premises are intuitively supposed to capture a more robust phenomenon than mere consistency with certain laws. It might be felt, in particular, that such consistency is too “negative” a characterization, failing to capture the actual presence of a real psychological potentiality. We feel the force of this concern and propose the following more “positive” supplementation of the above elucidation. In addition to its occurrent phenomenology, a conscious state also has a *dispositional profile* (or “functional role”): certain things it is disposed to cause, and be caused by, in certain circumstances. To say that every conscious state is such that its subject can attend to it through mere attention-shift is to make a positive claim about one facet, or aspect, of every conscious state’s dispositional profile. Now, the standard way to individuate dispositions is by their triggering and manifestation conditions (e.g., fragility is that disposition which manifests in breaking, chipping, etc. and is triggered by dropping, bumping into, etc.). So here we might characterize the relevant disposition as follows: it is the disposition (i) whose triggering condition is *exhausted* by the subject (a) forming an intention to be attentively aware of her concurrent conscious state and (b) performing a shift of attention, and (ii) whose manifestation consists in the subject becoming attentively aware of her concurrent conscious state through mere shift of attention. Call the disposition individuated by these triggering and manifestation conditions “D.” What Premise 1 claims is that every conscious state has D as part of its overall dispositional profile.⁹

Finally, our argument’s premises are framed in terms of “conscious states.” This is naturally interpreted as a universally quantified claim, though it may also be fairly understood to express a generic (as in “Dogs have four legs”).¹⁰ It may be objected,

⁸ The “can” of Premise 1, however, needs to express nomological possibility, for mere metaphysical or logical possibility here would be too weak.

⁹ Presumably, the fact that a conscious state has D is internally connected to the fact that the neuropsychological laws allow the subject to become attentively aware of that state through mere shift of attention: depending on one’s general views of laws, one might either take the laws to ground the disposition or the disposition to partially ground the laws – but we do not need to take a stance on this for the purposes of our argument.

¹⁰ Three-legged dogs constitute a refutation by counterexample of the universally quantified statement “All dogs have four legs”; but they do not—not immediately, at any rate—threaten the generic “Dogs have four

however, that many subjects simply do not have the attentional capacities that would make the first premise true of them. This might include infants, brain-trauma patients, and some nonhuman animals. To avoid getting bogged down in discussions of animal and infant consciousness and subjecthood, we propose to restrict the premises, in the first instance at least, to *neurotypical human adults*—roughly, human beings who have reached complete cognitive development broadly characteristic of the species and whose cognitive capacities are not impaired or otherwise atypical.

With these clarifications in place, we reach a more formal argument, which we offer as our “official” *argument from attention*:

P1. For any phenomenally conscious mental state M of a neurotypical human adult subject S, the laws of nature are such that S’s (a) forming an intention to be attentively aware of M and (b) performing an active mere shift of attention is sufficient for S to become attentively aware of M;

P2. For any *x* and neurotypical human adult subject S, the laws of nature are such that S’s (a) forming an intention to be attentively aware of *x* and (b) performing an active mere shift of attention is sufficient for S to become attentively aware of *x* only if S is already aware of *x*; therefore,

C. For any phenomenally conscious mental state M of a neurotypical human adult subject S, S is aware of M.

If we focus on the “positive” modal characterization in terms of dispositional profile, a simpler presentation becomes possible: 1) every phenomenally conscious state M has D as part of its overall dispositional profile; 2) the only way for M to have D as part of its overall dispositional profile is if its subject is aware of it; so, 3) every phenomenally conscious state is such that its subject is aware of it. The idea is that a disposition is always grounded in some categorical basis, and the most plausible categorical basis for M’s having D involves the subject’s being already aware of M.

Note that the argument supports an Awareness Principle restricted to only neurotypical human adults. But first, this is already an important result, very much in tension with what most first-order representationalists, for instance, believe. And

legs.” What the best semantic analysis of generics is, is topic of lively debate in recent linguistics and philosophy of language (see Leslie and Lerner 2016).

secondly, there is a way to reason abductively from a restricted to an unrestricted AP. The idea is that the *best explanation* of the fact that neurotypical human adults are always aware of their conscious states is that it is somehow in the nature of conscious states as such to involve their subjects' awareness of them. The argument would be strengthened as follows:

P3. The best explanation of the fact that for any phenomenally conscious mental state M of a neurotypical human adult subject S, S is aware of M, is that for any subject S and phenomenally conscious state M whatsoever, if S is in M at *t*, then S is aware of M at *t*; therefore,

C+. For any subject S and phenomenally conscious mental state M, if S is in M at *t*, then S is aware of M at *t*.

Why believe P3? Recall that the motivation to restrict P1 to neurotypical human adults was the epistemic possibility of subjects who lack the cognitive infrastructure, and in particular the attentional capacities, needed to attend to their own internal states. But what attentional capacities control is the foreground/background structure of awareness: what one is attentively aware of and what one is non-attentively aware of. They are not what generates the existence of awareness in the first place. So, when a creature cannot be attentively aware of something, and the reason for this is only their lack of attentional capacities, the more minimal hypothesis is that their lack of attentional capacities prevents them transitioning from non-attentive to attentive awareness of that thing—not that it prevents them from being aware of that thing in the first place. That is not something we can pin on attentional capacities.

We noted in §1 that our argument from attention has a cousin discussed and dismissed by Daniel Stoljar (2021). That argument is formulated by Stoljar (2021: 13) as follows:

P1. If I am in a conscious state C, it is possible for me to attend to my being in C.

P2. If it is possible to attend to my being in C, I am aware of my being in C.

C. Ergo, If I am in a conscious state C, I am aware of my being in C.

The second premise of this argument, however, is false. As Stoljar objects:

I don't know what's in my pocket; it is nevertheless possible for me to attend to what's in my pocket because it is possible for me to come to know what's in my pocket, and attend to what I know. (Stoljar 2021: 14)

Indeed, it is obviously possible for one to attend to something one is currently unaware of. In our museum example, it is possible for one to visually attend to a painting that is currently *not* within one's visual field, simply by turning one's neck. In Stoljar's example, it is possible for him to attend to whatever is in his pocket, e.g. by sticking his hand in his pocket and discovering what's in there. In all these cases, however, what one needs to do to come to attend to the relevant thing is *more* than just *mere shift of attention*. To come to attend to what is in his pocket, Stoljar needs to put his hand in his pocket and thereby become *perceptually aware* of what is in it; to attend to the next painting, you need to turn your neck and thereby become *visually aware* of it. Here more than mere shift of attention is involved. By contrast, as we have tried to show, P2 is much more plausible when restricted to cases in which it is possible to come to attend to something by *active mere shift of attention*. In these cases, we have argued, if it is possible for one to attend to something just by active mere shift of attention, then one has been aware of that thing all along.

5. Objections to P1

5.1. Transparency

One objection we can surely expect questions the very idea of *being able to attend to a conscious state*. Based on transparency considerations, an objector may argue that we can only attend to objects, properties, and events in our surroundings, not to our experiences of them: if we try to draw our attention to an experience, all we end up attending to is what the experience is about (Harman 1990).

In discussing the transparency of experience, it is important to distinguish two different claims. The first—which we may call *modest transparency*—is that when we attend to our experiences, we can only become aware of their representational or intentional properties, what objects and properties the experiences present; so, when you introspect a hallucinatory experience as of a lemon, you are only aware of the “lemon” represented in your experience, but since there is no real lemon, all this means

is that you are aware only of your experience's representing a lemon. According to modest transparency, the relevant property of being lemon-representing is all you can become aware of when you introspect your experience. You cannot become aware of any "intrinsic," non-representational properties of your experience. The other transparency claim—call it *extreme transparency*—is that when we try to attend to our experience, we fail and become aware only of external objects and properties themselves; we cannot attend to *any* properties of our experiences, not even their representational properties.

The main point we would like to make is that modest transparency is perfectly compatible with everything we have said here, as it allows for the possibility of attending to one's experiences and their properties, insisting only that these must be representational properties. P1 of our argument implies that it is possible to attend to one's experiences, but it does not take a stand on whether what one thereby attends to are representational or non-representational properties of the experience. It is only the second claim—what we called *extreme transparency*—that is incompatible with our P1. If extreme transparency is true, then it is indeed simply impossible to introspectively attend to experience.

Extreme transparency is an extraordinarily strong claim, however. On a traditional conception, the cognitive aim of introspection is to acquire information about inner experience as opposed to external objects. At least by the lights of this traditional conception, extreme transparency amounts to rejecting the very possibility of introspecting one's conscious life. Proponents of extreme transparency could of course adopt an alternative conception of introspection, as indeed they do. But then their conception of introspection, and to some extent their very use of the word "introspection," become highly revisionary.¹¹ Moreover, *even if* extreme transparency could be defended for some experiences, it is exceedingly implausible that it could be for all. Proponents of transparency tend to focus on visual experience, where transparency intuitions are strongest, and work their way through a series of increasingly doctrinal steps to types of experience for which extreme transparency is not at all *intuitively* or *pre-theoretically* compelling (e.g., mood experiences). But in reality there are many experiences that are such that it seems perfectly possible to

¹¹ As pointed out by a referee, this presupposes an extreme and highly controversial form of externalism about phenomenal consciousness.

attend to their phenomenology, if only insofar as it is representational. This includes notably moods and emotional experiences such as anger and sadness. It may perhaps—*perhaps!*—be reasonable to claim that the phenomenology of moods and emotions is exhausted by their representational properties, or that those are the only properties that show up in introspection of moods and emotions. It is *not* reasonable to claim that we cannot introspect *any* properties of our moods and emotions—not even their representational properties—and when we try, we only manage to become aware of some part of the external world.

Yet this is what rejection of P1 in the argument from attention requires. To repeat, versions of transparency which only insist that, upon attending to our experiences, we become aware of their representational properties exclusively, do not pose a threat to P1 of the argument from attention; on the contrary, they *concede* that attending to our experiences is nomologically possible. P1 does not take a stand on what properties of our conscious states introspection presents. The only thing it commits to is that introspection presents *something* about our conscious states.

5.2. *Absent-minded experiences*

It may be objected that some kinds of conscious state (such as “background” experiences or experiences undergone in a state of “automatism”) constitute counterexamples to P1. A person who is engrossed in a cognitively taxing task—say, a philosopher reasoning through a complicated argument—may realize that the nearby refrigerator has been emitting a low background noise only when the noise abruptly stops. A long-distance truck driver, who has been driving for several hours, may suddenly realize that she “has driven many miles without consciousness of the driving” (Armstrong 1968: 93). Call such experiences “absent-minded experiences.” An objector may argue that the philosopher and the truck driver *could not* have actively merely shifted their attention to their respective absent-minded experiences. Yet, both are *conscious* experiences: the humming experience must have contributed to the phenomenology of the philosopher’s experience all along for her to notice its abrupt termination; the truck driver must have been visually aware of the road all along for her to be able to keep the car on the road. If so, some conscious states are such that their subject cannot come to attend to them by active mere shift of attention.

Here too, however, it is important to distinguish two views on such absent-minded experiences. The first is that they are *particularly difficult* to attend to by mere

shift of attention. The second is that it is *strictly impossible* for us to do so. The objection that absent-minded experiences constitute counterexamples to P1 requires the second, more assertive claim, which involves commitment to the idea that it is *nomologically impossible* to come to attend to such experiences by active mere shift of attention. But this idea is very implausible: why should coming to attend to such experiences by active mere shift of attention be inconsistent with the laws of nature? Perhaps shifting attention to such peripheral experiences is more *difficult*, more *unlikely*, or more *infrequent* than shifting attention to those that lie closer to the center of consciousness. But this does not imply that such attention shift is ruled out by the psychological laws that govern neurotypical human adults' cognitive processes.

The first take on absent-minded experiences is certainly the more conservative. Rather than *impossible*, coming to attend to absent-minded experiences is just *harder* or *less likely* than attending to experiences more central in one's overall field of consciousness. This might be explained by a correlation between easiness or likelihood of active mere attending to an experience and the closeness of that experience to the center rather than periphery of consciousness: the closer something lies to the center, the easier or more likely it is for it to become the target of active mere shift of attention. This is quite plausible, if we consider that the center/periphery structure of the conscious field (partly) depends on the degree of *salience* that its non-central items bear with respect to the central ones (*cf.* Watzl 2011). In this framework, the more salient a conscious mental state, the more likely it is (other things being equal) that it will become the target of an active mere shift of attention. Absent-minded experiences lie at the farthest peripheries of the conscious field, and therefore are particularly unlikely to become such targets; but unlikelihood is not impossibility.

5.3. Phenomenal Overflow

It may be objected that P1 is inconsistent with the Sperling experiment (Sperling 1960). Subjects in this experiment are shown, for a very short duration (a twentieth of a second), nine letters arranged in a three-by-three grid. If asked which letters were displayed, subjects can typically recall accurately only four or five. So far, this is consistent with subjects having phenomenal representations of only four-five of the nine letters, with the rest never being phenomenally represented. In a crucial tweak, however, Sperling accompanied the display with one of three tones, telling subjects that if a high tone is heard they would be asked about the top-row letters, if a low tone they

would be asked about the bottom-row letters, and if a middle tone about the middle-row letters. In these circumstances, subjects were essentially always able to accurately recall all the relevant letters. Sperling concluded that subjects could recall *any* of the letters, just not *all* of the letters. Years later, Ned Block (1995) interpreted this to mean that subjects had phenomenal consciousness of every letter but had access-consciousness to only some of them; and therefore that phenomenal consciousness “overflows” access consciousness.

What we would like to consider is an objector who claims that Sperling subjects have phenomenal experiences they are not capable of attending to: they have phenomenal representations of each letter but it is nomologically *impossible* for them to attend to every one of these phenomenal representations.

We think this objection puns on the each/every distinction. It certainly appears true that Sperling subjects cannot attend to *every* phenomenal letter-representation they have. But each phenomenal letter-representation they have is such that they can attend to *it*. This, in fact, seems to be precisely what the experiment shows: insofar as reporting on an experience requires attending to it, and any of the subject’s phenomenal letter-representations can end up among those the subject reports, it would seem any of the subject’s phenomenal letter-representations can be attended to.

We conclude that phenomenal overflow does not threaten the nomological-possibility claim made in P1.

5.4. The Destructive Effects of Attention

A final objection to P1 comes from an observation, which goes back at least to Franz Brentano (1874), that attending to a conscious experience may alter and even destroy it. Suppose you are infuriated with the fact that *p*. In this state you are engrossed in *p* and consumed with rage. If you then turn your attention away from the fact that *p* and onto your feelings about *p*, something happens that changes substantially your overall emotional state. In taking this step back from the angering fact and focusing on your anger instead, you are taking some distance from the anger. The anger becomes an object of your contemplation rather than something you are consumed by. You are no longer lost *in* your anger, but instead have an “external” perspective on it and are thus more detached from the offending fact. Your anger may then lose much of its phenomenological bite. And it may fairly be said that your emotional state has changed: you were in emotional state E_1 before, now you are in emotional state E_2 .

The objection we have in mind leverages this phenomenon to argue that some conscious states *cannot* be attended to by their subject, insofar as the mobilization of attention tends to *destroy and replace* them. It is not the same conscious experience which is attended to that previously existed unattended. More generally, the objector may insist, the individuation of conscious experiences is not so straightforward as to allow us to assert that every conscious state is such that one can attend to *it* – as opposed to attending to some state that succeeds it in the wake of attention-shifting.

There are three points we want to make in response. First and most importantly, it is crucial to keep in mind here that P1 only requires the nomological *possibility* of non-destructive attention-shift. Even if human psychology were such that turning attention to one's anger *often*, or even *typically*, replaced a conscious state with another (and this is far from having been established), there would be nothing in the phenomena to suggest that it is strictly *impossible* for a human being to turn their attention to their anger without the anger going out of existence in consequence. As long as it is consistent with the laws of psychology that a person can turn their attention to their experience without destroying it, P1 stands.

Secondly, the objection obviously relies on certain assumptions about experience individuation that involve difficult issues. Although we can't get into this here, one can readily envisage principles for experience individuation under which E_1 and E_2 are the same experience after all. For instance, it is possible to distinguish in phenomenal properties between a qualitative and a quantitative dimension, and claim that experience individuation is sensitive only to qualitative difference. If your headache intensifies over the course of two hours, it is natural to say that it is still the same headache, and that it was *it* – *this* headache – which has intensified. The reason this is natural, arguably, is that quantitative change that fails to entrain qualitative change does not result in a new experience, but only in one and the same experience changing over time. It is phenomenological evolution, not phenomenological revolution. Now, this line of thought is certainly contestable, but our point is just that matters of experience individuation are so difficult that it's unclear whether they will ultimately support the objection.

Thirdly, however, it's not clear that questions of individuation touch the underlying idea of our argument. Suppose shifting attention onto E_1 results in the destruction of E_1 and its replacement with E_2 . There is clearly still a special relationship

that E_2 bears to E_1 that it does not bear to other experiences. Perhaps we could label this the “successor relation”: E_2 is numerically different from E_1 , under the present supposition, but it is nonetheless a successor of E_1 . The argument may then be reframed explicitly in these terms, perhaps as follows: every conscious experience of ours is such that we can attend either to it or to its successor through active mere shift of attention; we wouldn’t be able to come to attend to the experience or its successor through active mere shift of attention if we weren’t already aware of the experience; therefore, every conscious experience of ours is such that we are aware of it. There is every reason to suspect this argument will be just as probing as the original argument. Arguably, this is because the way our argument works is by focusing on an alleged *symptom* – the “attendability” of conscious experiences – to uncover an alleged *ground* of that symptom – the subject’s awareness of their conscious experiences. Quarrels about experience individuation may force us to reframe the symptom (e.g., as the “attendability” of an *experience-or-its-successor*), but this cannot affect the proposed underlying ground.

6. Objections to P2

6.1. *Blindsight*

Against P2 it may be objected that blindsight patients lack conscious visual awareness of objects in (part of) their visual field, but seem able to attend to these objects, as suggested by the fact that their correct behavioral responses to the shape and position of those objects are above chance (Kentridge, Heywood, and Weiskrantz 1999). It is possible to respond that what blindsight patients are using cannot be properly described as attention, perhaps precisely because it is an unconscious phenomenon. But we do not wish to quarrel about terms. However one uses the term “attention,” the argument from attention can be run safely with specifically *conscious* attention. If we replace “attend” with “consciously attend” in both premises of the argument from attention, the conclusion still follows. (Note: it is not circular to appeal to the concept of conscious attention in the argument’s premises, if only because the premises only make claims about the *possibility* of conscious attention, whereas the conclusion asserts *actual* awareness.)

6.2. *Non-attentive awareness*

P2 asserts that coming to attend to x via mere shift of attention requires prior awareness of x . Arguably, such prior awareness must be *non-attentive*: to *become*

attentively aware of x via mere shift of attention, S needs to be *non-attentively* aware of x . It may be objected, however, that it is impossible to be non-attentively aware of anything, because consciousness requires attention.¹² The objection is partly motivated by some empirical literature on “inattention blindness” (Mack and Rock 1998; Simons and Chabris 1999; Rensink 2002), aiming to show that unattended items or features cannot be consciously perceived.

Our response is threefold. First, the relevant empirical results are not conclusive. On the one hand, slight changes in the experimental setting deliver a drastic downscaling of inattention blindness (Memmert 2006; Ro *et al.* 2001). On the other hand, it has been argued that such empirical arguments face serious conceptual issues, including begging the question by assuming an attention-entailing definition of consciousness (Schwitzgebel 2007; Mole 2008).

Second, as mentioned in §5.3, the experimental paradigm introduced by Sperling (1960) has been taken to provide support for “phenomenal overflow,” i.e., the idea that what one can be phenomenally aware of exceeds what one can cognitively access—including what one can *attend to* (Block 2007). Although the Sperling-experiment-based argument for overflow has been criticized (see, e.g., Naccache and Dehaene 2007; Kouider *et al.* 2010; Cohen and Dennett 2011; Cova *et al.* 2021), many have forcefully argued that the empirical evidence is at least compatible with phenomenal overflow (see, e.g., Lamme 2003; Block 2011; Wu 2014). So, at the very least, the empirical evidence does not seem to deliver conclusive results against the possibility of non-attentive awareness.

Finally, and most importantly, P2 does *not* imply that pre-shift awareness must be utterly non-attentive. As mentioned in fn. 1 and in §3, what most fundamentally matters for the notion of “mere shift of attention” that is at play in our argument is that conscious awareness has a center/periphery or foreground/background structure that depends on the distribution of attention: the closer to the center the object of awareness, the more attentive the awareness itself. So, we may conceive of shift of attention from x to y as consisting in restructuring one’s awareness in such a way that, while before the shift one is *highly attentively* aware of x and *slightly attentively* aware of y , after the shift one becomes *highly attentively* aware of y and *slightly attentively* aware

¹² We thank an anonymous referee for pressing us on this point.

of *x*. In this framework, *all* awareness is attentive, because all awareness mobilizes *some* amount of attention resources. We therefore do not have to commit to the possibility of non-attentive awareness to accept P2.

6.3. Imaginative awareness

Suppose it is true that without prior awareness of an object, we cannot actively mere-shift our *perceptual* attention to it. Attentive awareness need not be perceptual. When we imagine a blue camel, we are in some sense aware of an imaginary blue camel, and aware of it attentively. We are attentively aware of the blue camel, moreover, as soon as we imagine it. Obviously, however, we could not have been aware of the imaginary camel *before* attending to it, since, in the sense in which there is an imaginary blue camel once we imagine it, there was no such camel for us to be aware of before we became attentively aware of it.

There are two things we want to say about this case. First, we must draw a distinction between two variants of the case. In the first, the imagining comes to us unbidden, and we suddenly find ourselves imagining a blue camel. This is not a case of *active*, but of *passive*, shift of attention; so we can set it aside. What we have to consider is the case where one forms the intention to imagine a blue camel and then acts on that intention, thereby coming to be (attentively) aware of a blue camel. This does not strike us as a genuine case of *mere* shift of attention. In order to become aware of a blue camel, the subject has to do more than just shift attention—she must also *form the image* of the blue camel, that is, generate the relevant imaginative experience. Since intentionally forming a mental image is a mental action, this is a case of *covert* attention-shift, but not of *mere* attention-shift. In mere attention-shift, by definition, the subject exercises *only* their attentional capacities, turning her attention from one thing to another. In the present case there are additional capacities that are mobilized, namely, imaginative capacities.

In addition, although there may be a sense in which when we imagine a blue camel we are aware of an imaginary blue camel, there is also a sense in which the term “aware of” does not feel perfectly felicitous in this context. For there is a certain connotation of *responsiveness* in the locution “aware of” as we use it to describe someone being aware of the laptop in front of them—that person is responsive to the presence of the object, in that the existence of her state of awareness is explained by the existence of what there is in front of her. In imagination there is no such

responsiveness—the “existence” of the imaginary blue camel is if anything explained in part by the existence of the imaginative act rather than the other way around. Now, perhaps standard use is flexible enough to tolerate “awareness” of imaginary objects. But then there would still be this fundamental difference between two *kinds* of awareness, one “responsive” and one “generative” (if you please). And so the argument from attention could be reframed entirely in terms of “responsive” awareness. In particular, the basic idea behind P2 would be understood as the idea that one cannot become attentively *responsive*-aware of x by active mere shift of attention unless one is already responsive-aware of x . And the conclusion would be that conscious states are such that we have responsive-awareness of them. This is the conclusion we want anyway, since proponents of AP don’t have in mind that conscious states are such that sometimes we are aware of them by way of imagining them.

6.4. Far-fetched cases

Blindsight and imagination are of course recognized mental phenomena, and so it is important to understand that they do not pose any genuine challenge to the argument from attention. But perhaps one could also cook up more far-fetched cases, through thought experiments, to counterexemplify P2.

Imagine, for instance, a subject who forms the intention to be attentively aware of some F that occupies a certain portion of their visual field, but just when they perform an active mere shift of attention, something else—some G that was not there before—appears in the targeted portion and occludes or replaces the F , with the result that the subject becomes attentively aware of the G instead of the F . In this case, it might be claimed, the subject actively merely shifts attention toward something—it turns out to be a G —of which she was not previously aware.

However, this is not a successful case of *active* mere shift of attention. Active shift of attention is a form of intentional (mental) action. So to actively shift her attention, the subject needs to have an intention. Now, if the content of that intention is something like “attending to that F ,” and the subject never becomes aware of an F , then the subject has never managed to perform the shift of attention she set out to perform; any shift of attention that resulted from her action would not qualify as an instance of active shift of attention.

We can, of course, envisage cases involving other intention contents. For instance, the subject may decide to shift her attention to region R of her visual field, and

it may so happen that by the time she has shifted her attention, R is occupied by a *G*. Nonetheless, in this case the subject *was* aware of R all along—at least nothing in the case suggests otherwise. So this would be a case in which the subject managed to actively mere-shift her attention to R having been previously non-attentively aware of R (so not a counter-example).

An even more *recherché* case is one where a subject forms the intention to become attentively aware of *whatever will occupy a certain portion of their visual field by the time they are done shifting their attention*. We may imagine that during the attention shift some *G* that was not there before appears in the relevant region of the visual field. In this case, it might be objected, the item one ends up attending to is what the intention is about (albeit “descriptively”); and yet the subject was *not* aware of it all along. So here we *do* get a counterexample to P2.

In addressing this case, we have to appreciate first of all the sophisticated air that surrounds the case, with its highly particular construction. That’s not always a vice in philosophy, but often such cases can be handled by carving out the right kind of exception, in a way that does not undermine the significance of the fundamental idea. In this case, it is noteworthy that although P1 claims that we can always attend to our conscious state by active mere shift of attention, the kind of mental act envisaged is one where we attend to our *concurrent* conscious state. So the intention definitive of the active shift of attention does not involve a future-tensed content (or, for that matter, a past-tensed one). Call an active shift of attention involving an intention with a present-tensed content “a simultaneous active shift of attention.” P1 could perfectly well be stated in terms of this more specific kind of mere shift of attention. That would then allow us to restrict P2 to these kinds of cases as well: it would claim only that a subject cannot become attentively aware of *x* through a *simultaneous* active mere shift of attention unless the subject is already aware of *x*.

7. Objections to P3

The third premise of the argument from attention is the one that takes us from ubiquitous inner awareness in neurotypical human adults to ubiquitous inner awareness in any form of conscious life. As noted, we find the argument from attention interesting enough without this extra step. After all, the thesis that inner awareness is present for all conscious states of neurotypical human adults does not involve the claim

that inner awareness is *absent* in other cases. It just goes a long way toward the unrestricted AP and opens the way for a shorter dialectical route toward AP. It is the task of P3 to take us this final bit of road.

Any objection targeting P3 must have the following form: it must claim that even on the assumption that inner awareness is ubiquitous in neurotypical human adults, there are cases where a conscious state is not accompanied by inner awareness of its occurrence: perhaps the conscious state of a snail, or of a neonate, or of a person suffering from depersonalization. Now, for many of these cases there are dedicated discussions in the literature. But more fundamentally, we think it highly instructive that in practice, *nobody* who rejects AP accepts ubiquitous inner awareness in neurotypical human adults. Opponents of the Awareness Principle, such as first-order representationalists and qualia theorists, never concede this much. Why? We think the reason is simple. Surely all or almost all theorists working on the nature of consciousness are themselves neurotypical human adults. To concede that inner awareness is universally present in neurotypical human adults' conscious life would be to concede that AP is true and exceptionless for all the cases we theorists actually know of from our personal experience, and is only false for cases about which we are effectively reduced to speculation. We can speculate about the inner lives of snails, neonates, and depersonalization patients (whose actual reports are by and large extremely ambiguous—see Billon and Kriegel 2015); but if all the cases we actually know do conform to AP, then it is *very* natural to suppose that the ubiquity of inner awareness in neurotypical human adults is just the “visible” facet of the ubiquity of inner awareness in all consciousness. This would be the best explanation for the exceptionless truth of AP in all the cases we are actually familiar with.

The alternative explanation would presumably have to be that it is something about the cognitive architecture peculiar to neurotypical human adults that makes possible the presence of inner awareness in conscious life: something about being human, being neurotypical, and/or being adult that guarantees inner awareness. But this is a groundless hypothesis. Setting aside atypical human adults, suppose we accept that certain nonhuman animals and human neonates lack the “cognitive firepower” needed for inner awareness. This in itself in no way tells us whether (i) they have phenomenally conscious states unaccompanied by inner awareness or (ii) they lack phenomenally conscious states (in virtue of lacking the inner awareness necessary for

such states). The proponent of AP would obviously plump for (ii), precisely on the grounds that in all cases we know of from the first-person perspective inner awareness is present. It is unclear what grounds there are for preferring (i) other than the desire to reject AP despite its exceptionless application to all the cases we know of from the first-person perspective.¹³

We conclude that once one accepts that AP applies without exception to all conscious states of neurotypical adult humans, it becomes very strained to deny that it applies in other cases. And this is what rejecting the argument from attention by denying P3 amounts to.

Conclusion

We opened by noting that AP-proponents tend to find AP phenomenologically compelling, while AP-opponents do not, and that some *arguments* would be needed to generate some dialectical pressure in one direction or another. The argument from attention is intended to do that. Naturally, it is not intended to settle the issue. There may also be direct arguments *against* AP, as there may be objections to either our first or second premise that we have not considered. Nonetheless, we hope that the argument from attention, as set out here, produces genuine input into the dialectic and casts AP in a more compelling light.¹⁴

¹³ Even in the case of atypical adult humans, it *has* been argued that the mental states of depersonalization patients, for instance, might lack phenomenal consciousness, precisely because they lack inner awareness (Billon 2011); though much more often, it has been argued that while certain kinds of self-awareness are absent in depersonalization, the minimal kind of inner awareness necessary for phenomenal consciousness is still present (Gallagher 2000).

¹⁴ For comments on a previous draft, we are grateful to Daniel Stoljar, as well as to three referees for *Philosophical Studies*. The paper also benefitted from helpful discussions with Davide Bordini and Arnaud Dewalque.

References

- Armstrong, David M. 1968. *A Materialist Theory of the Mind*. New York: Humanities Press.
- Billon, Alexandre and Uriah Kriegel 2015. "Jaspers' Dilemma: The Psychopathological Challenge to Subjectivity Theories of Consciousness." In Rocco Gennaro (ed.), *Disturbed Consciousness*, 29–54. Cambridge, MA: MIT Press.
- Block, Ned. 1995. "On a Confusion About a Function of Consciousness." *Behavioral and Brain Sciences* 18 (2): 227–47.
- . 2007. "Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience." *Behavioral and Brain Sciences* 30 (5): 481–548.
- . 2011. "Perceptual Consciousness Overflows Cognitive Access." *Trends in Cognitive Sciences* 15 (12): 567–75.
- Bordini, Davide. 2023. "Seeing through Transparency, Clearly." *Oxford Studies in Philosophy of Mind* 3: 263–96.
- Brentano, Franz. 1874. *Psychology from an Empirical Standpoint*. Translated by Antos C. Rancurello, Dailey B. Terrell, and Linda L. McAlister. London: Routledge, 1973.
- Campbell, John. 2004. "Reference as Attention." *Philosophical Studies* 120 (1/3): 265–76.
- Carruthers, Peter. 2000. *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Chalmers, David J. 2013. "The Contents of Consciousness: Reply to Hellie, Peacocke and Siegel." *Analysis* 73 (2): 345–68.
- Chatuverdi, Amit. Forthcoming. "Attentional Structuring, Subjectivity, and the Ubiquity of Reflexive Inner Awareness." *Inquiry*.
- Cohen, Michael A., and Daniel C. Dennett. 2011. "Consciousness Cannot Be Separated from Function." *Trends in Cognitive Sciences* 15 (8): 358–64.
- Cova, Florian, Maxence Gaillard, and François Kammerer. 2021. "Is the Phenomenological Overflow Argument Really Supported by Subjective Reports?" *Mind & Language* 36 (3): 422–50.
- Davidson, Donald. 1967. "The Logical Form of Action Sentences." In *The Logic of Decision and Action*, edited by Nicholas Rescher. Pittsburgh: University of Pittsburgh Press.
- Dickie, Imogen. 2011. "Visual Attention Fixes Demonstrative Reference by Eliminating Referential Luck." In *Attention: Philosophical and Psychological Essays*, edited by Christopher Mole, Declan Smithies, and Wayne Wu, 292–322. Oxford; New York: Oxford University Press.
- Dretske, Fred. 1993. "Conscious Experience." *Mind* 102 (406): 263–83.
- . 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Ganeri, Jonardon. 1999. "Self-Intimation, Memory and Personal Identity." *Journal of Indian Philosophy* 27 (5): 469–83.
- Giustina, Anna. 2022. "A Defense of Inner Awareness: The Memory Argument Revisited." *Review of Philosophy and Psychology* 13: 341–63.
- Goldman, Alvin I. 1970. *A Theory of Action*. Princeton: Princeton University Press.
- Gow, Laura. 2019. "Everything Is Clear: All Perceptual Experiences Are Transparent." *European Journal of Philosophy* 27: 412–25.
- Harman, Gilbert. 1990. "The Intrinsic Quality of Experience." *Philosophical Perspectives* 4: 31–52.
- James, William. 1890. *The Principles of Psychology*. London: Macmillan.
- Kellner, Birgit. 2010. "Self-Awareness (Svasaṃvedana) in Dignāga's Pramāṇasamuccaya and -Vṛtti: A Close Reading." *Journal of Indian Philosophy* 38: 203–31.
- Kelly, Sean Dorrance. 2004. "Reference and Attention: A Difficult Connection." *Philosophical Studies* 120: 277–86.

- Kentridge, R. W., C. A. Heywood, and L. Weiskrantz. 1999. "Attention without Awareness in Blindsight." *Proceedings of the Royal Society of London B: Biological Sciences* 266 (1430): 1805–11.
- Kim, Jaegwon. 1976. "Events as Property Exemplifications." In *Action Theory*, edited by Myles Brand and Douglas Walton. Dordrecht: Reidel.
- Kind, Amy. 2003. "What's so Transparent about Transparency?" *Philosophical Studies* 115 (3): 225–44.
- Kouider, Sid, Vincent de Gardelle, Jérôme Sackur, and Emmanuel Dupoux. 2010. "How Rich Is Consciousness? The Partial Awareness Hypothesis." *Trends in Cognitive Sciences* 14 (7): 301–7.
- Kriegel, Uriah. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford University Press.
- . 2019. "Dignāga's Argument for the Awareness Principle: An Analytic Refinement." *Philosophy East and West* 69: 143–55.
- Lamme, Victor A. F. 2003. "Why Visual Attention and Awareness Are Different." *Trends in Cognitive Sciences* 7 (1): 12–18.
- Leslie, Sarah-Jane, and Adam Lerner. 2016. "Generic Generalizations." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. <https://plato.stanford.edu/archives/win2016/entries/generics/>.
- Mack, Arien, and Irvin Rock. 1998. *Inattentional Blindness*. Cambridge, MA: MIT Press.
- Martin, M.G.F. 1997. "Sense, Reference and Selective Attention." *Aristotelian Society Supplementary Volume* 71 (1): 75–98.
- Memmert, Daniel. 2006. "The Effects of Eye Movements, Age, and Expertise on Inattentional Blindness." *Consciousness and Cognition* 15 (3): 620–27.
- Mole, Christopher. 2008. "Attention and Consciousness." *Journal of Consciousness Studies* 15 (4): 86–104.
- . 2011. *Attention Is Cognitive Unison: An Essay in Philosophical Psychology*. Oxford; New York: Oxford University Press.
- Naccache, Lionel, and Stanislas Dehaene. 2007. "Reportability and Illusions of Phenomenality in the Light of the Global Neuronal Workspace Model." *Behavioral and Brain Sciences* 30 (5–6): 518–20.
- Perrett, Roy W. 2003. "Intentionality and Self-Awareness." *Ratio* 16 (3): 222–35.
- Rensink, Ronald A. 2002. "Change Detection." *Annual Review of Psychology* 53 (1): 245–77.
- Ro, T., C. Russell, and N. Lavie. 2001. "Changing Faces: A Detection Advantage in the Flicker Paradigm." *Psychological Science* 12 (1): 94–99.
- Rosenthal, David M. 1997. "A Theory of Consciousness." In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen J. Flanagan, and Güven Güzeldere, 729–53. Cambridge, MA: MIT Press.
- Schwitzgebel, Eric. 2007. "Do You Have Constant Tactile Experience of Your Feet in Your Shoes? Or Is Experience Limited to What's in Attention?" *Journal of Consciousness Studies* 14 (3): 5–35.
- Shoemaker, Sydney. 1994. "Self-Knowledge and 'Inner Sense': Lecture II: The Broad Perceptual Model." *Philosophy and Phenomenological Research* 54 (2): 271–90.
- Siewert, Charles. 2004. "Is Experience Transparent?" *Philosophical Studies* 117 (1–2): 15–41.
- Simons, Daniel J., and Christopher F. Chabris. 1999. "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events." *Perception* 28 (9): 1059–74.

- Smithies, Declan. 2011. "Attention Is Rational-Access Consciousness." In *Attention: Philosophical and Psychological Essays*, edited by Christopher Mole, Declan Smithies, and Wayne Wu, 247–73. Oxford; New York: Oxford University Press.
- Spelring, George. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs* 74 (11): 1–29.
- Stoljar, Daniel. 2023. "Is There a Persuasive Argument for an Inner Awareness Theory of Consciousness?" *Erkenntnis* 88 (4): 1555–75.
- Thompson, Evan. 2011. "Self-No-Self? Memory and Reflexive Awareness." In *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*, edited by Mark Siderits, Evan Thompson, and Dan Zahavi, 157–75. Oxford: Oxford University Press.
- Watzl, Sebastian. 2011. "Attention as Structuring of the Stream of Consciousness." In *Attention: Philosophical and Psychological Essays*, edited by Christopher Mole, Declan Smithies, and Wayne Wu, 145–73. Oxford; New York: Oxford University Press.
- . 2017. *Structuring Mind. The Nature of Attention and How It Shapes Consciousness*. Oxford: Oxford University Press.
- Williams, Paul. 1998. *The Reflexive Nature of Awareness: A Tibetan Madhyamaka Defence*. Delhi: Motilal Banarsidass.
- Wu, Wayne. 2014. *Attention*. New York: Routledge.
- . 2023. "We Know What Attention Is!" *Trends in Cognitive Sciences*, December.
- Zahavi, Dan. 2005. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: MIT Press.