

Self-representationalism and phenomenology

Uriah Kriegel

Published online: 8 February 2008
© Springer Science+Business Media B.V. 2008

Abstract To a first approximation, self-representationalism is the view that a mental state *M* is phenomenally conscious just in case *M* represents itself in the appropriate way. Proponents of self-representationalism seem to think that the phenomenology of ordinary conscious experience is on their side, but opponents seem to think the opposite. In this paper, I consider the phenomenological merits and demerits of self-representationalism. I argue that there is phenomenological evidence *in favor* of self-representationalism, and rather more confidently, that there is no phenomenological evidence *against* self-representationalism.

Keywords Consciousness · Self-representationalism · Peripheral inner awareness · Phenomenology · The transparency of experience · Subjective character

1 Introduction: theories of consciousness and phenomenological argumentation

The philosophical literature features a great variety of theories of consciousness. There are any number of ways one may taxonomize these, but the main fault line is the distinction between *one-level* and *two-level* theories. The latter hold that a mental state's being conscious involves the subject's being aware of it, the former do not.

The main versions of one-level theories are functionalist and representationalist. Perhaps more accurately, the division is between purely functionalist and functionalist-representationalist views. According to purely functionalist theories,

U. Kriegel (✉)
Department of Philosophy, University of Arizona, Tucson, AZ 85721, USA
e-mail: kriegel@email.arizona.edu

U. Kriegel
Department of Philosophy, University of Sydney, Sydney, NSW 2006, Australia

conscious states are conscious in virtue of occupying the right functional role. Typically, the relevant functional role is understood in terms of some kind of *global availability*: a mental state is conscious just in case it is available to a wide variety of consuming systems (Baars 1988; Dennett 1991). According to functionalist-representationalist accounts, conscious states are conscious in virtue of carrying the right representational content (sometimes in addition to occupying the right functional role).¹ Typically, the relevant content is understood in terms of the representation of highly specific properties in the subject's environment and body (Dretske 1995; Tye 2000).

The main two-level theories are higher-order representationalism and self-representationalism. According to higher-order representationalism, conscious states are conscious in virtue of being targeted by suitable higher-order representations, either quasi-perceptual representations (Armstrong 1968; Lycan 1996) or thoughts (Rosenthal 1990, 2005; Gennaro 1996). According to self-representationalism, conscious states are conscious in virtue of representing themselves in a suitable way (Smith 1986; Kriegel 2003, 2006).

Curiously, arguments for and against these theories are infrequently phenomenologically based. Perhaps this is because it is deemed that any phenomenological bases would be inevitably shaky. For to the extent that one's theory of consciousness rests on phenomenological observations that are not entirely pedestrian, it is open to glib rejection. All an opponent would have to do is profess that the unpedestrian phenomenological pronouncements do not resonate with her. Ditto when one's *critique* of a theory of consciousness rests on unpedestrian phenomenological claims.² Nonetheless, some phenomenological argumentation does surface now and again in the relevant literature. In particular, one finds a quite powerful phenomenological argument for representationalism, as well as a phenomenological argument for self-representationalism.³ By contrast, there seem to be no phenomenological arguments for functionalism and higher-order representationalism.

There is a straightforward reason why there cannot be phenomenological arguments for functionalism. Functionalism construes phenomenal consciousness as a functional role property. Such properties are *dispositional*: a mental state occupies the functional role it does in virtue of being disposed to cause (and be caused by) certain things. But dispositions are most certainly not phenomenologically manifest.

¹ Sometimes the representationalist adds functionalist qualifications to this basic form of representational theory. For example, in Tye's PANIC theory the relevant representational content is required to be "poised," where a content is poised if it is carried by a vehicle with the right kind of causal impact on beliefs and desires. The poise requirement is thus a functionalist qualification to the otherwise representationalist account.

² At the background here is a very real and important worry about the rational resolvability of phenomenological disputes, a worry I do not propose to address here. What is interesting for our present purposes is that, in some cases, phenomenological claims that were not previously taken for granted have sometimes been successfully used to mount genuinely compelling arguments for and against certain theories of consciousness.

³ The argument is phenomenological not in the sense that every premise in it is phenomenological, but rather in the looser sense that the key premise in it is phenomenological. Of course, what makes a premise the "key premise" may depend on a variety of less than fully objective factors. In any event, we can say that the argument in question—to be presented in Sect. 3—is *largely* phenomenological.

Just as fragility is not perceivable (though it is thinkable), global availability is not introspectible (though it is thinkable). Only occurrent, categorical properties are introspectible. To the extent that functionalism construes consciousness as a dispositional property, then, it entails that consciousness is non-introspectible. This makes it hard to see how a *phenomenological* argument might be mounted in its favor.⁴

Higher-order representationalism faces a different phenomenological discomfort. The theory construes the higher-order representation it posits (and the positing of which is distinctive of it) as typically non-conscious (we will see why in Sect. 3). Since the crucial higher-order representation is typically non-conscious, it typically cannot be phenomenologically manifest. Thus by the theory's own admission there can be direct phenomenological evidence for its crucial, distinguishing posit only in non-typical cases.⁵ There may be *indirect* phenomenological evidence, in the sense that some phenomenological facts may seem to require the positing of a higher-order representation for their explanation. But there cannot be any *direct* phenomenological evidence for the theory's definitive posit.

Things are different with self-representationalism. Like higher-order representationalists, self-representationalists hold that conscious states are necessarily represented. But unlike higher-order representationalists, they hold that conscious states are always *consciously* represented. Since self-representing conscious states are represented by themselves, and they are conscious, they are represented by states that are conscious. Thus self-representationalism leaves the door open for direct phenomenological evidence for the thesis that conscious states are always represented—something that, as we just saw, is excluded by higher-order representationalism.

And indeed, some self-representationalists have attempted to develop at least partially phenomenological arguments for their view (Kriegel 2003). The phenomenological cornerstone of self-representationalism is the claim that consciousness is intimately connected with a sort of peripheral inner awareness. It is then argued that only self-representationalism can accommodate this fact, wherefore it is singularly phenomenologically adequate among theories of consciousness. An argument of this form will be expounded in Sects. 2 and 3.

However, the most powerful phenomenological argument for a theory of consciousness, at least sociologically speaking, is most certainly an argument in

⁴ Indeed, this consideration could be developed as a positive argument *against* functionalism. Arguably, it is a priori true that phenomenal consciousness is introspectible. If functionalism entails that it is not, then the only way it can avoid being incoherent is by collapsing into eliminativism about phenomenal consciousness.

⁵ The typical cases are cases of non-introspective consciousness, i.e., conscious awareness that is focused primarily on the external world (or the subject's body). The non-typical cases are cases of introspective consciousness, when the subject consciously attends to her ongoing conscious experience. In these cases, according to higher-order theory, the subject enters a third-order state that represents the second-order state in such a way that both the second- and first-order states are conscious. In such a scenario, the higher-order state that makes the first-order state conscious is phenomenologically manifest. Nonetheless, the third-order state that makes the second-order state conscious is not, since it itself is not phenomenologically manifest. Thus the theory allows for direct phenomenological evidence of some conscious-making higher-order states, but only some and only on the rare occasions of introspective consciousness.

favor of representationalism—the celebrated argument from transparency. According to the transparency thesis, conscious experience is transparent, in the sense that when one examines introspectively one's concurrent conscious experience, one only becomes aware of what the experience is *of*, that is, of what it represents (Harman 1990). The only introspectible aspect of a conscious state, then, is its representational content. The transparency thesis is then claimed to lead, when conjoined with highly plausible assumptions, to representationalism (Tye 2000).⁶

There may also be a temptation to interpret the transparency thesis as constituting evidence against self-representationalism. The thought might be that the phenomenological reality of peripheral inner awareness is belied by the transparency of experience, leaving self-representationalism either false or unmotivated. I will consider this phenomenological argument against self-representationalism in Sect. 4.

Thus the present paper examines the relationship between phenomenology and self-representationalism. It is, more specifically, a phenomenological defense of self-representationalism. In Sect. 2, I explicate and motivate the notion of peripheral inner awareness. In Sect. 3, I develop the phenomenological argument from peripheral inner awareness to self-representationalism. In Sect. 4, I argue that the transparency argument against self-representationalism is unsuccessful. I close, in Sect. 5, with a discussion of recent phenomenological objections to self-representationalism due to Gennaro (forthcoming).

2 Peripheral inner awareness

The phenomenological starting point of self-representationalism is the claim that consciousness involves a kind of peripheral inner awareness. To come to grips with the notion of peripheral inner awareness, we would do well to start by clarifying the two distinctions underlying it: between outer and inner awareness and between focal and peripheral awareness.

Outer awareness is the awareness we have of external objects, events, and states of affairs (in our environment and body), e.g., awareness of a milk truck driving by. Inner awareness is awareness of inner, mental states and events, e.g., awareness of a growing nervousness about an afternoon appointment.

The distinction between focal and peripheral awareness is primarily a distinction regarding the distribution of attention: focal awareness of something is highly attentive awareness of it, peripheral awareness of it is less attentive awareness.⁷ Thus, in my current visual experience I am focally aware of the laptop before me and only peripherally aware of a bonsai tree to my left; which is just to say that I am very visually attentive to the laptop and not so much to the bonsai. Likewise, when

⁶ For example, one might argue that since all phenomenal properties are in principle introspectible, if all introspectible properties of conscious states are representational properties, then all the phenomenal properties of conscious states are representational properties. Alternatively, one might argue that since conscious experiences seem to be nothing but representations, they probably are, on pain of massive error.

⁷ Since attentiveness is a matter of degree, focal and peripheral awareness are two poles of a single spectrum.

one listens to a piano concerto, one's auditory experience is liable to involve focal awareness of the piano and merely peripheral awareness of the cellos, the violas, and for that matter the sound of raindrops on the roof. Ditto for awareness in other perceptual modalities: every form of perceptual awareness comes in focal and peripheral varieties.

A less pedestrian claim is that the focal/peripheral distinction applies also to non-perceptual awareness. Arguably, there is a felt difference between thinking of an unpaid bill in a focused and attentive way and doing so in a vague and fleeting manner. More crucially for self-representationalists, the same holds for inner awareness and self-awareness. I can deliberately turn inward and introspectively attend to my ongoing nervousness, scrutinizing it and dwelling on it in an attentive and focused way. But I can also be aware of my ongoing nervousness in a non-attentive, dim, unimposing, non-consuming, indeed peripheral manner.

The claim at the base of the phenomenological argument for self-representationalism is that consciousness involves crucially an element of peripheral inner awareness. There are in fact three separate theses of relevance here. The first thesis is that peripheral inner awareness is *occasionally* involved in consciousness, the second is that peripheral inner awareness is *universally* involved in consciousness, and the third that peripheral inner awareness is *constitutively* involved in consciousness.⁸

We may formulate them as follows:

- (1) For some conscious state C of a subject S, S has peripheral inner awareness of C.
- (2) For any conscious state C of a subject S, S has peripheral inner awareness of C.
- (3) Necessarily, for any conscious state C of a subject S, S has peripheral inner awareness of C.

Self-representationalists are committed to the strongest of those theses, (3).⁹ On their view, my conscious experience of the laptop involves not only focal visual awareness of the laptop and peripheral visual awareness of the bonsai, but also peripheral inner awareness of that very visual experience. The same holds for the auditory experience of the piano concerto, the cognitive experience of the unpaid bill, and any other conscious experience.

Why believe in a universal, let alone constitutive, peripheral inner awareness? The answer lies in a certain natural conception of phenomenal character. When I look at the white wall to my right, I have a conscious experience of the wall, and there is a whitish way it is like for me to have that experience. This “whitish way it is like for me” constitutes the experience's phenomenal character. Plausibly, there are two discernable components to this phenomenal character, this “whitish way it is like for me.” One is the “whitish” component, the other is the “for me” component. We may call the former *qualitative character* and the latter *subjective character*. On this conception of phenomenal character—expounded in Levine 2001

⁸ Note well: the claim that consciousness involves peripheral inner awareness constitutively is neutral on the issue of the order of explanation or constitution between the two. That is, it is neutral on the issue of whether a conscious state is conscious in virtue of being the object of peripheral inner awareness or is the object of peripheral inner awareness in virtue of being conscious.

⁹ Although see footnote 32, where a possible modification of this claim is suggested.

and Kriegel 2005—phenomenal character is rightly thought of as the compresence of qualitative and subjective character.¹⁰

The qualitative character of a conscious experience is ordinarily captured by the various sensuous qualities it involves: whitish qualities, bitterish qualities, loudish qualities, etc.¹¹ But what captures the subjective character of an experience? What grounds its for-me-ness, its subjective significance or givenness to the subject? The thought that peripheral inner awareness is a constitutive element of conscious experience is motivated by such questions. The idea is that an experience's for-me-ness or subjective character is constituted by the subject's concurrent peripheral inner awareness of the experience.

To this it might be objected, from the left, that there is no such thing as a phenomenologically manifest for-me-ness in experience, and from the right, that this for-me-ness is a more special and elusive phenomenon than peripheral inner awareness. We will consider the left-wing objection in Sects. 4 and 5, so let me say a few quick words on the right-wing one here.

On the view under consideration, subjective character has more to it, or is more inherently peculiar, than is suggested by anything like the notion of peripheral inner awareness. On some views one finds in the phenomenological tradition, for example, subjective character (or “pre-reflective self-consciousness,” as the phenomenon is often called within that tradition) is an altogether unstructured feature, a simple intrinsic glow that cannot be accounted for in terms of ordinary awareness-of.¹²

The intrinsic-glow view may represent something of an overreaction. It may well be that, by the end of inquiry, we would be forced to embrace a *sui generis* conception of subjective character along the above lines. But it would be unwise to do so at the *beginning* of inquiry. There is, after all, an element of hopelessness about the proposal. To call a phenomenon *sui generis* is not quite the same as to

¹⁰ I intend this as a substantive claim, not as a stipulation about how to use “phenomenal character.” For the claim to be substantive, we need an independent way to fix on phenomenal character. There are any number of ways to do that, but my preferred one is to fix the reference of “phenomenal character” via to the following rigidified definite description: “that property of conscious state for which in the actual world there is an appearance of an explanatory gap.” (See Block (1995) for this method of fixing on specifically phenomenal consciousness.) The substantive claim before us, then, is that the property for which there is (actually) an appearance of an explanatory gap is the compresence of qualitative and subjective character.

¹¹ In describing qualitative character this way—through appeal to sensuous qualities such as whitishness and bitterishness—I do not intend to prejudge the debate between representationalist and anti-representationalist (“phenomenalists”) accounts of qualitative character. According to the former, qualitative character is nothing but representational content; according to the latter, it is not. Accordingly, according to a representationalist account, an experience has a whitish sensuous quality just in case it represents a white surface, volume, or film; according to an anti-representationalist account, there is more to an experience's having whitish sensuous quality than its representing anything white. My description of qualitative character in the terms I do is supposed to leave both views an open possibility. I describe the qualitative character in those terms to distinguish it from the corresponding qualities of worldly objects—the whiteness of a wall or the bitterness of a lemon (see Levine (2001) for the same practice).

¹² In this vein, Husserl (1928) suggested that subjective character involves a *sui generis* form of intentional directedness, which he called *act-intentionality* and distinguished from *object-intentionality* (see also Zahavi 2004; Drummond 2006). This *sui generis* form of intentionality is nowhere else to be found in nature and is categorically different from regular awareness-of, whether focal or peripheral. This sort of intrinsic-glow view has more recently been given an adverbial treatment by some modern exponents, notably Thomasson (2000). See also Thomas (2003) and Zahavi (2004).

account for it. More deeply, the very suggestion that subjective character is simple and unstructured is rather implausible. It is natural to hold that, in virtue of this subjective character, the subject of a conscious experience is *aware of* her experience. It would be quite odd indeed to maintain that an experience is *for* the subject but the subject is nonetheless completely unaware of it. Since this awareness is awareness-*of*, it involves an of-ness relation to the experience. To that extent, it cannot be an unstructured intrinsic glow, because it enables a structure involving a relation between the subject and her experience.¹³

There is no question that our awareness of our conscious experiences is *special*. But this specialness can be done justice to as a form of peripheral awareness. According to self-representationalism, peripheral inner awareness has some unusual properties (Kriegel 2004): it is ubiquitous (all conscious experiences involve it, so since we have conscious experiences throughout our waking life, we have peripheral inner awareness throughout our waking life); involuntary (we cannot shut it down by will, since we cannot shut down consciousness by will); effortless (like seeing, and unlike calculating, this way of being aware is not so much something that we do as something that happens to us), etc.¹⁴ These peculiarities of peripheral inner awareness may account for the specialness and elusiveness of the subjective character.

I conclude that we are better off adopting the view that subjective character is a form of awareness-of, rather than an intrinsic, unstructured feature that involves no epistemic relation between the subject and her experience. What I now want to argue is that once we admit that conscious experiences involve a subjective character, and that the latter is constituted by peripheral inner awareness, a line of argument opens up in favor of self-representationalism.

3 The argument from peripheral inner awareness

Self-representationalists take (2) and (3) above as springboards for argumentation in favor of their theory. This is because they believe something like the following bridge principle:

(B) S's having peripheral awareness of C consists in S being in C and C's representing itself.

Or perhaps more accurately:

(*) For any conscious state C of a subject S, if S has peripheral awareness of C, then C represents itself.

¹³ On one interpretation, the Husserlian view is compatible with the idea that in virtue of our conscious states' subjective character we are aware of these states. The thought is that act-intentionality is different from familiar awareness-of in that it is a *non-objectifying* awareness. It *is* an awareness-of, just not an awareness-of-object. However, so interpreted the Husserlian view strikes me as incoherent, inasmuch as it strikes me as conceptually true that, in the relevant sense of "object," awareness-of is always awareness-of-object. In essence, I am merely reporting here that I cannot really wrap my mind around the notion of non-objectifying awareness.

¹⁴ Moreover, in Sect. 3, I will argue that peripheral inner awareness, despite being phenomenologically manifest, is introspectively inaccessible. As we will see, this may well account for some of the paradoxical air surrounding pre-reflective self-awareness.

What this bridge principle has to recommend it is something we will return to momentarily. For now, note that, conjoined with (*), propositions (1)–(3) above entail, respectively:

- (1*) For some conscious state C of a subject S, C represents itself.
- (2*) For any conscious state C of a subject S, C represents itself.
- (3*) Necessarily, for any conscious state C of a subject S, C represents itself.

Strictly speaking, only (3*) may qualify as self-representationalism in its philosophical sense, claiming a *constitutive* connection between consciousness and self-representation.¹⁵ But (2*) may be taken as a form of contingent, empirical self-representationalism. Furthermore, (2*) may be taken as a stepping stone in a non-demonstrative argument for (3*). Thus one may develop a two-step argument for (3*): in a first step, the conjunction of (2) and (*) is shown to entail (2*), and in a second step, (2*) is shown to support (3*) by something like an inference to the best explanation.

It is necessary to take this more circuitous road in arguing for (3*), because (3) is probably *not* a purely phenomenological claim. That a certain feature is sometimes or always present in consciousness may be phenomenologically manifest to us, say via introspection; but that the feature is *constitutively* present in consciousness arguably cannot be established by introspection, and is not a phenomenological fact. For presumably constitutive connections are not phenomenologically manifest, in more or less the sense in which Hume noted that the necessary connexion between cause and effect is not perceivable (we can have no “impression” of it). If so, to the extent that there is a genuinely phenomenological argument for self-representationalism, it must take (2) as its starting point.

Let us turn now to the way in which the bridging of (2) and (2*) proceeds—which is, again, in two steps. The first step is to show that a state of which the subject is peripherally aware is represented by the subject. The second step is to show that the best account of represented conscious states construes them as represented by themselves, i.e., as self-representing.

The first step is based on the simple idea that awareness requires representation. For me to be aware of a tree, I must have a mental representation of it. Likewise for inner awareness: subject S is aware of a conscious state C only if S has a representation of C. More specifically:

- (a) For any conscious state C of a subject S, if S is peripherally aware of C, then C is represented by S.

¹⁵ I say that (3*) *may* qualify as self-representationalism in its philosophical sense because it may also be thought that, on one natural reading, (3*) falls short of self-representationalism in its philosophical sense. The latter requires not only that conscious state be constitutively and necessarily self-representing, but also that they be conscious precisely *because*, or in virtue of the fact that, they are self-representing. That is, self-representationalism might naturally be thought to include an “in virtue of” claim, or direction-of-explanation claim, that does not follow from (3*) strictly speaking. However, one legitimate attitude we could take here is that there are two important theses in the area, both of which deserve the title “self-representationalism,” though perhaps to different degrees: one is (3*), the other is the conjunction of (3*) and the direction-of-explanation claim. In what follows, I allow the former thesis to go by self-representationalism. So in essence we can distinguish two grades of self-representationalism—a purely modal claim and a claim that also incorporates an “in virtue of” clause—but here I focus on the former.

From this and (2), it already follows that:

(2a) For any conscious state C of a subject S , C is represented by S .

Since subjects represent things by harboring mental representations of them, we could add that:

(+) For any x and any subject S , if x is represented by S , then there is a mental state M , such that S is in M and M represents x .

It follows from (2a) and (+) that:

(2a+) For any conscious state C of a subject S , there is a mental state M , such that S is in M and M represents C .

This concludes the first step of the bridging of (2) and (2*). It brings us from (2) to (2a+). The second step will take us from (2a+) to (2*).

Moving to this second step, we may note that either (i) every represented conscious state is represented by itself, or (ii) every represented conscious state is represented by a mental state numerically distinct from it, or (iii) some represented conscious states are self-represented and some are other-represented. Combining each of these options with (2a+), we would obtain:

- (2a + i) For any conscious state C of a subject S , there is a mental state M , such that S is in M , M represents C , and $M = C$.
- (2a + ii) For any conscious state C of a subject S , there is a mental state M , such that S is in M , M represents C , and $M \neq C$.
- (2a + iii) For any conscious state C of a subject S , there is a mental state M , such that S is in M and M represents C ; for some conscious state C of a subject S , there is a mental state M , such that S is in M , M represents C , and $M \neq C$; and for some conscious state C of a subject S , there is a mental state M , such that S is in M , M represents C , and $M = C$.

The second step of the argument proceeds by establishing the superiority of (2a + i), which is equivalent to (2*), over its competitors. Let us see, then, in what ways (2a + ii) and (2a + iii) are unsatisfactory.

(2a + ii) is a form of higher-order representationalism. We noted in Sect. 1 that higher-order representationalism construes the higher-order representation, M , as ordinarily non-conscious. This is because otherwise it would fall into a vicious regress. If *all* higher-order representations were conscious, every one of them would have to be itself represented by a yet higher-order representation, and we would be off on an infinite regress. If just *most* higher-order representations were conscious, the regress would be vicious in a less straightforward manner. It would implicate the higher-order theorist in positing a great many representations, of ever ascending hierarchies, whenever a subject underwent a single conscious experience. Such a hierarchy, while not *logically* problematic, is nonetheless *empirically* incredible. To avoid such an empirically vicious regress, the higher-order theorist must claim that the great majority of higher-order representations are non-conscious. Typically, higher-order representationalists claim that ordinary conscious states, those that are not explicitly introspected, are represented by *unconscious* higher-order states,

while introspected conscious states are represented by *conscious* ones (which are themselves unconsciously represented) (Rosenthal 1990).

If this was the case, however, higher-order representations would not show up in the phenomenology of ordinary, unintrospected conscious states. Far from being regular, universal, and indeed constitutive elements in the phenomenology, they would be most of the time altogether external to the phenomenology (Kriegel 2003). This could be the basis for an immediate phenomenological argument against (2a + ii): that it is incompatible with the fact that (2) is supposed to be a *phenomenological* claim. The only way (2) could be a phenomenological claim is if the peripheral inner awareness of each conscious state is internal to the phenomenology of that conscious state.¹⁶ But the higher-order representationalist cannot accommodate the idea that the peripheral inner awareness is internal to the phenomenology of *every* conscious state, since in the case of most conscious states she claims that the peripheral inner awareness is implemented by a non-conscious representation, and non-conscious representations, being non-conscious, do not show up in the phenomenology.

But the argument can be pressed beyond this immediate point. What the immediate point shows is that higher-order representationalism ascribes to ordinary, non-introspected conscious experiences the same phenomenology one-level theories do. It portrays the phenomenology of ordinary, non-introspected conscious experiences as featuring exclusively floor-level elements, precisely as (first-order) representationalism does. However, once it is admitted that the phenomenology of ordinary conscious states is entirely at the floor level, it is difficult to see what is supposed to motivate the postulation of a higher-order representation in the first place. In the absence of direct phenomenological evidence for a higher-order representation, the evidence for it would have to be either (a) *indirect* phenomenological evidence, (b) *a posteriori* experimental evidence, or (c) evidence from something like *a priori* conceptual analysis. Yet there is no sign of either type of evidence.

Indirect phenomenological evidence would be a matter of phenomenological facts the best explanation of which would involve the postulation of higher-order representations.¹⁷ The only consideration I can think of that takes this form is in Rosenthal's (1990) early work. Rosenthal notes that a wine connoisseur and a novice can have very different experiences upon degusting a sophisticated wine, even if the strictly sensory aspect of their experiences is identical. The best

¹⁶ And indeed this is what we would expect if (2) is motivated by the idea that peripheral inner awareness captures a for-me-ness that is a component of phenomenal character—as we claimed it does in Sect. 2.

¹⁷ By “indirect” phenomenological evidence, I mean that what we have phenomenological evidence for is not the very same thing whose existence we are trying to establish. Consider an analogy from perceptual evidence. Suppose we are seeking perceptual evidence for the existence of parrots in the neighborhood. One kind of perceptual evidence is that we saw parrot nests in a number of trees in the neighborhood. A very different kind of perceptual evidence is that we saw *parrots* in the neighborhood a few times. The former is *indirect* perceptual evidence for the existence of parrots in the neighborhood, the latter is *direct* perceptual evidence. By the same token, direct phenomenological evidence for the existence of a mental representation R would be constituted by the fact that R itself is phenomenologically manifest; indirect phenomenological evidence for the existence of R would be constituted by a phenomenological fact P, such that P is not identical to, but makes probable, the fact that R exists.

explanation for this, he claims, is that the connoisseur possesses second-order concepts the novice lacks, and can therefore form more acute second-order representations of her wine experience, which then modulate the experience's overall phenomenology. This argument has the right form, but is probably unsound. As (first-order) representationalists have long noted, the phenomenological difference between the connoisseur's experience and the novice's can be explained purely in terms of first-order concepts, or at least first-order discriminative abilities: the connoisseur can discriminate properties of the wine that the novice cannot.¹⁸ It is not clear to me that either of these two explanations is better positioned than its rival, but what that suggests is that there is no phenomenological argument from the best explanation to be had here for higher-order representationalism.¹⁹

Another form of argument based on indirect phenomenological evidence might advert to the fact that, according to higher-order representationalism, introspected conscious states are consciously represented, inferring from this that *all* conscious states are represented (either consciously or unconsciously). In one version, the inference would be inductive, reasoning that since all the conscious states one has actively introspected have been represented states, it is plausible that all conscious states are represented, even those that have not been actively introspected. The obvious objection to this reasoning, however, is that the inductive sample is wildly biased. *Of course* all the conscious states in the sample would be represented if what makes them belong to the sample is that they are introspected.²⁰ In another version, the inference would be abductive, reasoning that the best explanation for there being a phenomenology of introspective consciousness is the higher-order representationalist one. Here the problem is that, as I will argue in Sect. 4, the self-representationalist account of the phenomenology of introspective consciousness is actually superior.²¹

Moving on to the second alternative source of evidence, what sort of *a posteriori* experimental evidence might require the postulation of higher-order representations involved in consciousness? Perhaps neuro-imaging studies might reveal that, whenever a subject appears to be conscious, the part of the brain usually responsible for producing higher-order representations (probably a prefrontal area) is activated. To date, we do not have solid evidence for this proposition from neuro-imaging research, although scraps of evidence are starting

¹⁸ I follow here the common practice of distinguishing flavors from tastes and construing flavors as properties belonging properly to external objects such as wine, not to mental states such as wine experiences.

¹⁹ I am assuming here that Rosenthal's relevant argument is the only one that takes the form of an indirect phenomenological argument. As I said, I am not familiar with any other argument from this literature that takes this form.

²⁰ The deep problem here is that the introspecting itself constitutes the representing, so the very criterion for selecting the sample adverts to representedness. Furthermore, if the inductive inference were warranted for the conclusion that all conscious states are represented, it would also be warranted for the conclusion that they are all *consciously* represented, since all the instances in the sample are consciously represented. And that would land the higher-order representationalist with the problem of infinite regress.

²¹ See third paragraph to last in Sect. 4, including Footnote 37.

to come in.²² It is telling, in any case, that no higher-order representationalist has ever made the case for her view on such experimental grounds. The thought is rather that there is something intuitive, hence (presumably) altogether pre-theoretic, about the idea that, as Rosenthal (1990) puts it, “conscious states are states we are conscious of.” This idea would not be *intuitive* if the reason to believe it were purely experimental.

Perhaps the motivation for positing higher-order representations, and the reason it is intuitive that conscious states are states we are conscious of, are broadly *verbal*, or *conceptual*. Perhaps the thought is that we are dealing here with a conceptual truth. Lycan (1996), for example, explicitly claims that it is just a contingent fact about the English word “conscious” that it applies appropriately only to states of which one is aware (and hence represents). However, this response would render higher-order representationalism quite insubstantial, making it a theory of consciousness in the mundane sense of the word, whatever that turns out to be, rather than in the sense of *phenomenal* consciousness. Since it is phenomenal consciousness that generates the explanatory gap, and hence the philosophical conundrum about consciousness, this move would strip higher-order representationalism of any relevance to said conundrum. The source of the philosophical anxiety surrounding consciousness is phenomenal consciousness, not consciousness in the mundane sense of the term. To cast higher-order representationalism as a theory of the latter is effectively to give up on it as a theory that addresses the philosophical anxiety surrounding consciousness.

It might be claimed that consciousness in the mundane sense of the word *just is* phenomenal consciousness. But presumably we would have to be convinced of that. On the face of it, phenomenal consciousness is whatever property generates the explanatory gap, or the mystery of consciousness, *not* (at least not necessarily) the property of being denoted by the English word “consciousness” and its cognates. A theory of the latter property would also be a theory of the former only on the outside chance that there is a necessary coextension between the two, something which no higher-order representationalist has argued and which is almost certainly false.

Relatedly, it is a common charge against higher-order representationalism that it is, in Block’s (1995) terms, a theory of *access* consciousness, not *phenomenal* consciousness. Higher-order representationalists routinely reject the charge (Rosenthal 2002), but the reason the charge sticks is that they have failed to cast their posited higher-order representations as in any way relevant to the phenomenology. As we just saw, there are principled reasons why they *cannot* do better: the threat of regress forces them to construe the higher-order representations as by and large non-conscious and therefore external to the phenomenology. If a phenomenological story were told about our awareness of our conscious states, the force of Block’s charge would dissolve. But once such a phenomenological story is told, one

²² This topic is too vast and too complex to go into in much detail here, and anyway it is irrelevant to the main thrust of the present discussion. For these reasons, I will not go into this matter in much detail. Suffice it to say that although consensus in the scientific community is starting to form that executive function and other “higher cognition” brain area are implicated in the neural correlates of consciousness, the consensus (inasmuch as it exists) is driven primarily by theoretical, not experimental considerations. This is not to say that experimental evidence is entirely absent (see Kriegel 2007), but the evidence is nowhere near overwhelming at this point.

must construe the awareness of our conscious states as conscious and internal to the phenomenology, which entails relinquishing higher-order representationalism (on pain of infinite regress).

These difficulties beset (2a + ii) because it construes the conscious state C and its higher-order representation M as numerically distinct, which creates the threat of vicious regress. The threat can be averted only by construing M as ordinarily non-conscious, but that in turn (i) clashes with the idea that (2) is a phenomenological claim and (ii) makes higher-order representationalism impossible to motivate. These difficulties are immediately avoided if C and M are construed as numerically identical, as they indeed are in (2a + i). Once they are construed as numerically identical, M can be construed as conscious without threat of regress.

It might be objected that (2a + i) involves its own vicious regress: although it does not lead to an infinite regress of *states*, it does lead to an infinite regress of *contents*. More generally, it is sometimes claimed that such a regress of contents bedevils the self-representational theory of consciousness (Zahavi 1998; Drummond 2006). I will now argue, however, that the objection is misguided.²³

For starters, compare and contrast the following two theses:

- (S) Necessarily, for any mental state S, S is conscious iff S represents itself.
 (C) Necessarily, for any mental content C, C is conscious iff C represents itself.

(S) is an account of what makes mental *states* conscious, (C) an account of what makes mental *contents* conscious. I grant that (C) leads to infinite regress of contents. But self-representationalism as such—and certainly as construed here, i.e., along the lines of (3*)—is not committed to (C), only to (S). This does mean that the self-representationalist still owes us an account of what makes a mental *content* conscious, but to my ear, what it means to say that a mental content is conscious is just to say that it is the content of a conscious state.²⁴ Thus I am inclined to embrace the following thesis about conscious *contents*:

- (C–) Necessarily, for any mental content C, C is conscious iff there is a mental state S, such that (i) C is the content of S and (ii) S is conscious.

The conjunction of (S) and (C–) does *not* lead to infinite regress. What it leads to is the claim that necessarily, for any mental content C, C is conscious iff there is a mental state S, such that (i) C is the content of S and (ii) S represents itself. This claim does not by itself produce any regress.

There would still be some threat of regress if self-representationalism required conscious states not just to represent themselves, but to represent their own representational properties—all their representational properties. Thus the following indeed leads to an infinite regress:

²³ I will discuss the regress-of-contents argument against self-representationalism itself, but what I say should apply *mutatis mutandis* to (2a + i) more specifically. For more detailed responses, see Kriegel (2003) and especially Williford (2006).

²⁴ It follows from this analysis that a conscious state cannot carry a non-conscious content. But I do not find this implication to be independently implausible.

(S+) Necessarily, for any mental state S and mental content C, such that S has C, S is conscious iff S represents itself to have C.

But self-representationalism is not committed to (S+) either. Self-representationalism requires that conscious states represent themselves, but it does not require that they represent themselves to have their self-representational properties (Williford 2006). Perhaps they represent themselves to have their qualitative properties; perhaps they represent themselves to have some of their representational properties, just not the self-representational ones; perhaps they represent themselves to have *some* properties, no matter which; perhaps they represent themselves without representing *any* of their properties (e.g., by representing themselves *indexically*). All these ways of self-representing would not lead to infinite regress.²⁵ It is clear, in any case, that the regress-of-contents charge is successful only if self-representationalism is committed either to (C) or to (S+), but that in fact it is committed to neither.

I conclude that (2a + i) is greatly superior to (2a + ii).²⁶ As for (2a + iii)—according to which conscious states are sometimes represented by themselves and sometimes by states other than themselves—this type of (poorly motivated) view fares only marginally better than higher-order representationalism with respect to the regress problem, while at the same time accruing further liabilities.

For higher-order representationalism, the problem of regress presents itself for every conscious state. For the type of mixed theory under consideration, it presents itself only for a subset of those, namely, the non-self-represented ones. But this means that, within the framework of the mixed view, a subset of conscious states generate regresses that are, if not infinite, otherwise vicious. That only a subset of conscious states do so is certainly an advantage, but ultimately vicious regress is implausible for even a single conscious state.

In addition, the mixed view introduces an untoward heterogeneity in the underlying nature of consciousness. It is rather implausible, on the face of it, that some conscious states are conscious in virtue of being represented one way, while others are conscious in virtue of being represented another way. One would expect a certain underlying unity, at a reasonable degree of abstraction, among conscious states. Indeed, if conscious states form a natural kind, as they probably do, they should boast a relatively concrete “underlying nature” common to all of them. The mixed view threatens to deny them that.

Thus there are good reasons to reject (2a + ii) and (2a + iii), higher-order representationalism and the mixed view. Since these are the only alternatives to (2a + i), the (contingent) self-representational view, those are effectively reasons to adopt the latter. As noted, (2a + i) is equivalent to (2*), so we have here an argument for (2*). Insofar as we can then use (2*) to argue abductively for (3*), the

²⁵ My own preference is for the view that mental states are conscious in virtue of representing themselves to have their qualitative properties. This accounts elegantly for the way in which qualitative character and subjective character are cohesively unified in the phenomenal character of conscious states.

²⁶ The argument was in large part phenomenological, though not entirely so. The immediate point against (2a + ii) is purely phenomenological, but the further pressure on higher-order representationalism, as impossible to motivate in any other way, was only partially phenomenological.

argument would lead us to self-representationalism in its most proper form. Thus we have before us a largely phenomenological argument for self-representationalism: an argument leaning heavily (though perhaps not exclusively) on a phenomenological premise (i.e., thesis (2)).

4 The transparency argument against self-representationalism

The idea commonly referred to as the “transparency of experience” was probably first introduced by G. E. Moore (1903), and has been defended again by Armstrong (1968), Harman (1990), and others.²⁷ Conscious experience is transparent, according to Moore (as commonly understood at least), in that whenever we try to introspect one of our experiences, we can only become acquainted with what it is an experience *of*—not with the experiencing itself. We thus cannot help but see the world right through the experience of it, as though the experience was, in itself, purely transparent.

There are many different ways to formulate the thesis of transparency—or perhaps many different interpretations of what the thesis actually is. In the present paper, I will work with the following formulation of the thesis:

(TE) The only introspectively accessible aspect of a phenomenal experience is its world-directed representational content.

TE carries intuitive conviction. But it seems on the face of it to conflict with (2), the thesis of the universality of peripheral inner awareness. (2) claims that there is a phenomenological element in all conscious experiences that does not have to do with their world-directed representational content, namely, peripheral inner awareness. (TE) appears to deny this.

More precisely, we may construct the following argument from transparency to the phenomenological unreality of peripheral inner awareness: the only introspectively accessible aspect of a phenomenal experience is its world-directed representational content; peripheral inner awareness involves no world-directed representational content; therefore, peripheral inner awareness is not an introspectively accessible aspect of phenomenal experience. Call this the *transparency argument against self-representationalism*.

To defend (2), the universality of peripheral inner awareness, one must argue either that (TE) is false or that, despite initial appearances, (TE) is not in conflict with (2). In this paper, I pursue the latter strategy. Although a number of philosophers have argued against the transparency thesis (e.g., Block 1996), and

²⁷ Moore (1903, p. 20, 25) writes: “[T]he moment we try to fix our attention upon consciousness and to see what, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as it were diaphanous... [T]he other element which I have called “consciousness”... seems to escape us: it seems, if I may use the metaphor, to be transparent—we look through it and see nothing but the blue.” (This quotation is stitched together from two different passages in Moore’s paper, but it involves no misrepresentation of the text.)

although I have my own reservations about it,²⁸ the thesis does appear to be more popular among philosophers of mind than its negation. What I want to do in this section is argue that, despite initial appearances to the contrary, self-representationalism is *compatible* with the transparency thesis. More specifically, I will argue that, combined with a natural account of introspection, self-representationalism actually *predicts* that peripheral inner awareness is not introspectible.

Note, to begin with, that (TE) does not say that there *are* no elements in the phenomenology that go beyond world-directed representational contents. Rather, it says that no such elements are *introspectively accessible*. This leaves open the possibility that there are introspectively *inaccessible* elements in the phenomenology that go beyond the world-directed representational content. In fact, (TE) does not even say that there are no *phenomenologically manifest* elements in the phenomenology beyond world-directed representational contents; merely that if there are such phenomenologically manifest elements, it is not through introspection that they are manifest.²⁹

For most putative elements in the phenomenology, it would be odd to claim that, though they are phenomenologically manifest, they are not introspectively accessible.³⁰ But what I want to argue is that the case of peripheral inner awareness is an exception: peripheral inner awareness is phenomenologically manifest, yet not introspectible.

To see how this could be, consider what is involved in introspecting one's concurrent conscious experience. According to one model, when one introspects one's current experience, one enters a new representational state, a state of introspective awareness, that takes one's current experience as its object. But for the self-representationalist, a more natural model of introspecting suggests itself. On this model, introspecting one's current experience does not involve entering a completely new representational state. Rather, it involves reorganizing the center/periphery structure of one's overall experience, by transforming one's peripheral inner awareness of one's current experience into a focal one. Before the introspecting, my visual awareness of the laptop is focal, whereas my visual awareness of the bonsai and my inner awareness of my overall experience are peripheral. When I introspect, what happens is that the inner awareness displaces the visual awareness of the laptop in the focal center. Now the inner awareness is

²⁸ My own reservation is that, if the transparency thesis was true, then the content of a perceptual experience and the content of an introspecting of that perceptual experience would be identical, which seems false. Suppose I have a perceptual experience of a white wall that represents whiteness and rectangularity. According to the transparency thesis (at least in one natural construal of it), when I introspect my experience the only features I am aware of are whiteness and rectangularity. If this was the case, then the experience and its introspecting would represent the same properties. The only place in the literature that I know of where something like this worry is aired is in Janzen (2006).

²⁹ I bring up the notion of phenomenological manifestness to accommodate the possibility that some aspects of the phenomenology are completely opaque to the subject, who bears no epistemic relation to them whatsoever. This view strikes me as extremely implausible, but it appears to be coherent. What I want to argue, in any case, is that peripheral inner awareness is such that, although we cannot introspect it, we do bear some epistemic relation to it, in virtue of which it is phenomenologically manifest to us.

³⁰ They may not be introspectively accessed most of the time, but to claim that they are not accessible in principle may appear to stretch credulity.

focal, whereas both the visual awareness of the laptop and the visual awareness of the bonsai are peripheral (though the latter may be *more* peripheral than the former).³¹

This model of introspecting is more natural for the self-representationalist, since she holds that all conscious experiences—and not only the explicitly introspected ones—are objects of awareness. It also has the added advantage of illuminating the fact that introspecting does not feel, phenomenologically, like performing a “dramatic” mental act, an act that creates an altogether new representation (as, say, visualizing a cat does). Instead, it feels more like shifting around one’s attention, and attending more carefully to contents that were already previously there.

My claim is that this model predicts that peripheral inner awareness is not introspectible. On this model, introspection cannot *reveal* peripheral inner awareness, because it *supplants* it. Introspecting one’s current experience is a matter of having focal inner awareness of it. But once one enters the state of focal inner awareness, the state of *peripheral* inner awareness that existed prior to the introspecting has gone out of existence; or rather, it has “graduated” to focal awareness. This is because, on the present model, it is one and the same awareness that is peripheral before the act of introspecting is performed and focal thereafter. Thus introspecting cannot reveal peripheral inner awareness because it *annihilates* it (by supplanting it). All this is to say that, for someone who believes in the phenomenology of peripheral inner awareness, and embraces the model of introspecting under consideration, the right—indeed, only—prediction to make is that peripheral inner awareness is not introspectible.

The apparent tension between self-representationalism and (TE) is based on an attachment to a model of introspection as involving the production of a new awareness of that which is being introspected. That model is perhaps natural for someone who holds that there are conscious experiences of which one is unaware. But the self-representationalist denies this, and therefore that model of introspection is unnatural for her. The natural model for her is the attention-shift model, according to which no new awareness is produced, but rather attention is shifted from one preexisting awareness to another. Once we adopt the attention-shift model, we see that self-representationalism predicts that peripheral inner awareness is non-introspectible. This makes self-representationalism compatible with (TE). If the transparency theorist is right that all the other elements in a conscious experience’s phenomenology are presented to introspection as world-directed representational contents, then it would follow that (TE) is true. Self-representationalism is in itself

³¹ It is also worth noting that, in the case of visual awareness, there may be two focal/peripheral distinctions that could be confused. One lines up with the (cognitively based) attentive/non-attentive distinction, but the other lines up with the (physiologically based) foveal/non-foveal distinction. Thus one can attend to something in one’s visual field that one is not foveating. In that case, one would be focally visually aware of that thing in one sense but not in the other. The distinction that is relevant to our present purposes is between attentive/inattentive visual awareness. So in the case described in the text, I am peripherally aware of both the laptop and the bonsai, though the laptop I am also foveating, whereas the bonsai I am not.

silent on the truth of the antecedent, of course, but this is precisely why we can say that it is *compatible* with (TE).³²

Let us dwell on this point a little longer. If the transparency theorist is right on the nature of all the aspects of phenomenology other than peripheral inner awareness, then according to self-representationalism, before I introspect my experience of the laptop and bonsai, all the phenomenological components of my experience *but one* are representations of the external world (or at least present themselves this way to introspection). Once I introspect the experience, the introspecting destroys that sole item, leaving only the representations of the external world standing.³³ What I become aware of when I introspect is thus that the experience represents the world

³² The attention-shift model of introspecting may, however, raise the objection that on this very account, there are conscious experiences that do not involve peripheral inner awareness, namely, explicitly introspected conscious experiences. Since in those experiences the peripheral inner awareness has been destroyed, they must lack peripheral inner awareness. There are two ways to respond to this objection. As for the first point, we may claim that what is destroyed in an introspected visual experience is only the peripheral inner awareness of that visual experience, not *every* peripheral inner awareness. In fact, as soon as one consciously introspects the visual experience, a peripheral inner awareness is formed, namely, the peripheral inner awareness of the conscious introspecting. Thus the overall conscious experience of the subject still features peripheral inner awareness, it is just not peripheral inner awareness of the same thing as before the introspecting. The second response is to modify the original phenomenological observation from Sect. 2. The modified claim would be not that every conscious experience involves peripheral inner awareness, but that every conscious experience involves inner awareness. Ordinary, non-introspective conscious experiences involve peripheral inner awareness, whereas introspective conscious experiences involve focal inner awareness. But they all involve inner awareness.

³³ It might be objected that this has the untoward consequence that one and the same conscious state cannot be unintrospected at one time and introspected at another. The objector's reasoning might be this: if a token experience E1 at t1 and a token experience E2 at t2 have different phenomenologies, then E1 and E2 are numerically distinct experiences; self-representationalism implies that introspected and unintrospected experiences always have different phenomenologies (since the latter does, but the former does not, involve peripheral inner awareness); therefore, self-representationalism implies that introspected and unintrospected experiences are always numerically distinct. There are two responses to this objection: first, it is unclear that the first premise in this reasoning is true across the board; second, there are reasons to deny that introspected and unintrospected experiences have the same phenomenology that are independent of self-representationalism. First, it is plausible that one and the same token experience can have different phenomenologies at different times. This may not be plausible if we take mental states to be property exemplifications, but is plausible if we take them to be bare particulars. For bare particulars can persist even when their properties change. Scenarios involving phenomenal sorites support the bare-particular view: one and the same token experience can change, by a series of indiscriminable changes from being yellowish to being reddish. It is true that on this view one and the same phenomenal token may fall under different phenomenal types at different times, but although this is certainly unusual, it should not be deemed impossible. Thus, it is plausible to maintain that one and the same concrete particular can be a caterpillar at one time and a butterfly at another, thus falling (quite essentially) under two different kinds at different times. Such scenarios are rare, but not impossible. As for the second point, Hill (1988) notes, very plausibly, that introspection appears to function as a "volume control" of phenomenology: an introspected toothache hurts more than an unintrospected one, and does so *because* it is introspected. Thus there is always a difference in phenomenal intensity between introspected and unintrospected experiences. So if phenomenal difference between experiences rules out their token-identity across time, then this is a consequence for all views of phenomenology, not only self-representationalism. For my part, I think it is more plausible to maintain that the introspected toothache is numerically identical to the unintrospected one despite their phenomenological difference. Thus it cannot be true across the board that phenomenological difference entails token-difference, and the fact that an experience E1 involves peripheral inner awareness while E2 involves focal inner awareness does not rule out their token-identity across time.

to be so and thus—and of nothing more. This is just what the transparency thesis claims. Thus the transparency theorist ought to have no qualms about embracing self-representationalism.³⁴

If peripheral inner awareness is not introspectively accessible, in what sense is it phenomenologically manifest? More generally, how can any mental phenomenon be phenomenologically manifest but not through introspection? This is a good question, and one for which I do not have a general answer. But note that a parallel quandary could be raised for peripheral *perceptual* awareness. As you listen to a piano concerto and are (auditorily) focally aware of the piano and peripherally aware of the cellos, you can decide to attend to the cellos rather than the piano. What you cannot do, however, is attend to the cellos *qua* peripherally experienced. For that would amount to attending to the cellos *qua* unattended to, or become focally aware of cellos *qua* objects of non-focal awareness—a manifest contradiction.³⁵ Yet few would infer that the sound of cellos is something you are unaware of. For despite the fact that we cannot attend to the periphery *qua* periphery, we have a general impression of this periphery from our ordinary everyday phenomenology. Just how we have such an impression, and what its nature is, are important questions that ultimately should be addressed; but there is certainly such an impression. The same situation, I maintain, applies to peripheral *inner* awareness. We cannot attend to peripheral inner awareness *qua* peripheral, because that would amount to attending to it *qua* unattended to, which is impossible. But we do have a general impression of peripheral inner awareness from our ordinary, non-introspective consciousness.

To be sure, this talk of a “general impression” is, as it stands, somewhat mysterious. What is this general impression? How come we have one, given that we cannot attend to that of which we allegedly have a general impression? Is the general impression grounded in a higher-order representation or not, and how is all this supposed to work? These too are good questions, but I think we would do well to treat them here as material for future investigation. As the case of peripheral

³⁴ To say this is not, however, to say that the transparency theorist is right. One could reject the transparency thesis and hold that some features of experience are not presented to introspection as world-directed representational contents. For example, one may hold that the whitish qualitative character of my perceptual experience of the white wall involves more than just the worldly-white-involving representational content. If one held this view, then in embracing self-representationalism, one would hold that before I introspect my experience of the white wall, all the phenomenological components of my experience *but two* (at least) are representations of the external world (or at least are presented to be so to introspection), but that once I introspect the experience, the introspecting destroys one of these two items, leaving only representations of the external world and the extra-representational whitish quality standing.

³⁵ In the case of visual awareness, the duality of focal/peripheral distinctions noted in footnote 31 may tempt us to think that I can attend to the visual periphery *qua* periphery after all—say, because I can attend to a bonsai without foveating it. But this would be merely attending to the foveal periphery *qua* foveal periphery. It would not be attending to the attentional periphery *qua* attentional periphery. It is the latter kind we are concerned with, however, since the relevant focal/peripheral distinction, as drawn in Sect. 2, is drawn in terms of attention distribution.

auditory awareness just discussed shows, there clearly *is* some way this all works. The fact that it is not clear *how* should not blind us to this fact.³⁶

All this is to claim that the transparency thesis does not by itself pose a threat to the phenomenological reality of peripheral inner awareness. There may also be, in addition, a more positive piece of evidence for that phenomenological reality. As I noted above, introspecting feels more like a phenomenologically light shifting around of attention than like a dramatic mental act that produces a completely new awareness. Indeed, it feels very much like shifting one's attention from piano to cellos, rather than creating a previously non-existent awareness of cellos. This may be taken to constitute phenomenological evidence that prior to the introspecting, there was already inner awareness of the conscious experience, albeit peripheral—just as it suggests that prior to attending to the cellos, there was already an auditory awareness of them in place, albeit peripheral. If the inner awareness was altogether outside the phenomenology, introspecting would be phenomenologically more dramatic than it actually is: it would feel like the introduction of a new item into the phenomenology, not like a reorganization of the center/periphery structure over already existing items.³⁷

I conclude that there is no conflict between (2) and (TE)—the two are perfectly compatible against the background of the attention-shift model of introspecting. Returning to the transparency argument against self-representationalism, the response is that self-representationalism can accept the argument's conclusion, namely that peripheral inner awareness is not introspectible. What it rejects is the implication that peripheral inner awareness is not phenomenologically manifest (let alone that it is not phenomenologically real). This depends, of course, on finding daylight between introspective accessibility and phenomenological manifestness. The above discussion, in particular the adopted model of introspecting, is supposed to show that there is this daylight. The combination of the attention-shift model and self-representationalism allows us to see how there could be a peripheral inner awareness that is both phenomenologically manifest and non-introspectible. An opponent of self-representationalism would thus have to do more than just assert (and defend) the transparency of experience; she would also have to argue either against the attention-shift model of introspecting or against the claim that it portrays peripheral inner awareness as both phenomenologically manifest and non-introspectible.

³⁶ Moreover, my purpose here is only to show that the transparency of experience as such is not in tension with (2), the thesis of the universality of peripheral inner awareness. And the above discussion should suffice to establish that. The discussion suggests that at most (2) is in tension with the conjunction of the transparency thesis and the further claim that there can be no general impressions of things to which we cannot attend. But first, it is not in tension with the transparency thesis *by itself*, and second, it is surely false that there can be no general impression of things to which we cannot attend, as the case of peripheral auditory awareness shows.

³⁷ This consideration suggests that the self-representational account of the phenomenology of introspective consciousness is more accurate than the higher-order representational account, which requires the appearance in the phenomenology of a representation that was previously not there. This is why the abductive version of the second indirect phenomenological evidence proposed for higher-order representationalism in Sect. 3 fails.

5 Gennaro's objections

In his recent critique of the phenomenological underpinnings of self-representationalism, Gennaro (forthcoming) comes close to denying not only (2), but (1)—the claim that peripheral inner awareness is *occasionally* involved in consciousness. According to Gennaro, peripheral inner awareness sometimes accompanies focal *inner* awareness, but never focal *outer* awareness. More generally, he allows combinations of outer focal and peripheral awareness, inner focal and peripheral awareness, and inner focal and outer peripheral awareness, but disallows the combination of outer focal and inner peripheral awareness. This means that as soon as our conscious awareness is focused outward, there can be no peripheral inner awareness accompanying it. Thus peripheral inner awareness is *never* involved in non-introspective consciousness, and only sometimes involved in introspective consciousness. A fortiori, then, peripheral inner awareness cannot be always involved in—be a universal feature of—consciousness.

Why reject the existence of any peripheral inner awareness in the absence of focal inner awareness? Gennaro (forthcoming, section 3.4) cites three considerations. The first is that there are no uncontroversial examples of peripheral inner awareness, in the way there are of peripheral outer awareness. The second is that, upon consideration of his own phenomenology, he cannot find any peripheral inner awareness. The third is that the professions of those who claim they do find such peripheral inner awareness in their phenomenology can be readily explained as resting on the following illusion: when examining their phenomenology, they enter a state of reflective, explicit, indeed *focal* inner awareness, and so they can never catch themselves with a conscious experience that does not involve inner awareness. Let me address each of these considerations.

The first claim is that there are no uncontroversial cases of peripheral inner awareness. What counts as “uncontroversial” is of course not a straightforward matter, but the claim that there are many uncontroversial cases of peripheral inner awareness seems far from absurd. Thus, just as I am often focally aware of a nervousness or apprehension before delivering a colloquium presentation, so I am also peripherally aware of such nervousness/apprehension.

Moreover, Gennaro seems to assume that if there exists peripheral inner awareness, there *should* be uncontroversial instances of it—perhaps because there are uncontroversial cases for peripheral outer awareness, e.g., peripheral vision. It is worth noting, however, that the case of visual awareness may set the bar too high, inasmuch as visual awareness is notoriously more phenomenologically overwhelming than just about any other form of awareness (perhaps with the exception of somatic awareness). Thus suppose philosopher P goes eliminativist with respect to peripheral olfactory awareness. According to P, there exist both focal and peripheral awareness in all perceptual modalities other than olfaction, while olfactory awareness comes only in the focal variety. Moreover, P challenges, there are no uncontroversial cases of peripheral olfactory awareness. The reason we are not overly impressed with P's challenge is that (i) we think that the standard of uncontroversiality is too demanding for olfactory awareness and (ii) there is something arbitrary about the idea that olfactory awareness, alone among all forms

of awareness, does not come in a peripheral variety. I maintain that our reaction to Gennaro's first consideration should be the same: as long as there are cases that are *plausibly* described as involving peripheral inner awareness, then even if they are not *uncontroversially* so described, we should accept the description. All the more so given that there is something arbitrary about denying that the focal/peripheral distinction applies to inner awareness, given that it applies to outer awareness.

Interestingly, Gennaro himself admits that there are cases plausibly described as peripheral inner awareness, since he admits that peripheral inner awareness may accompany focal inner awareness. His nice example is of becoming *focally* aware of wanting to eat a chocolate cake and at the same time *peripherally* aware of an attendant guilt about doing so. The volitional experience of desire is the focus of inner awareness, the emotional experience of guilt its periphery. To my mind, this is a plausible enough description of a phenomenology involving peripheral inner awareness. Granted, Gennaro's view is that while there are plausible cases of peripheral inner awareness accompanied by focal inner awareness, there are none of peripheral inner awareness not so accompanied. But this is where the consideration of arbitrariness ought to be raised. There is something odd about the idea that peripheral inner awareness can be phenomenologically manifest, but only when accompanied by focal inner awareness. All the more so since (again by Gennaro's own admission) inner and outer awareness can co-exist with a different distribution of attention, namely, when one has focal inner awareness and peripheral outer awareness. Here Gennaro's nice example is of catching oneself engaged in introspective self-scrutiny, say attending to a growing anxiety over a looming doctor appointment, while watching television. There is certainly peripheral visual awareness of the figures on the screen accompanying the focal inner awareness of one's anxiety. It does seem arbitrary, however, to suppose that the converse cannot happen. Indeed, it seems plain to me that, in some circumstances, one might be engaged in the very same activity of both watching television and innerly attending to a growing anxiety, but with a different distribution of attention: one might be attentively, focally aware of the little screen while at the same time being inattentively, peripherally aware of one's anxiety.

The second consideration raised by Gennaro is that, personally, he can never find peripheral inner awareness in his phenomenology. To this we must respond primarily by recalling our discussion of transparency from Sect. 4. If one expects to find peripheral inner awareness via introspection, one is bound to be disappointed, since one's introspective attempt would supplant, and thus destroy, the peripheral inner awareness one is looking for.

Nonetheless, I have claimed in Sect. 4 that we do have a "general impression" of peripheral inner awareness (in virtue of which the latter is phenomenologically manifest); and Gennaro may profess to having no such impression. Ultimately, of course, we might simply accuse Gennaro of phenomenological carelessness. But perhaps more interesting and more substantive remarks can be made by way of explaining why the allegedly universal peripheral inner awareness of our ongoing conscious experience is so elusive. In Sect. 2, we noted two interesting features of this peripheral inner awareness: its ubiquity and involuntariness. These two characteristics may partially explain the fact that peripheral inner awareness does

not lend itself to easy noticing.³⁸ It is in general difficult to notice even stimuli that are constant for a relatively short period, such as the hum of the refrigerator pump. If peripheral inner awareness is indeed ubiquitous, its absolute constancy throughout our waking life would account for the fact that it is so phenomenologically elusive. And the fact that it is involuntary means that we cannot control the conditions of its presence and “compare,” as it were, a situation in which it is present with one in which it is absent, in a way that would bring its presence into sharp relief.³⁹ I put forward that these features of peripheral inner awareness make it singularly elusive among mental phenomena, but that still a general impression of it can be had and that considerations designed to bring one’s overall conception of phenomenology into reflective equilibrium (such as the considerations of arbitrariness appealed to above) recommend its admission.

Gennaro’s third consideration against peripheral inner awareness is that there is a ready diagnosis of the error allegedly committed by proponents of peripheral inner awareness, namely, that they enter a state of inner awareness whenever they seek inner awareness. Here my response is twofold. First, the fact that there is an explanation of why one always finds x when one seeks x that does not appeal to the presence of x does not by itself rule out the alternative explanation, that one always finds x when one seeks x because there always is x . Indeed, the latter explanation is the simpler one, so that some reason would have to be given why the former is actually the better explanation all things considered. Second, it is unclear that the diagnosis offered by Gennaro is all that compelling. For on the view defended in Sect. 4 we have a general impression of peripheral inner awareness, just as we have of peripheral visual awareness, even when we are *not* engaged in seeking this peripheral inner awareness.⁴⁰

I conclude that Gennaro’s three considerations, while they bring into sharp relief the web of issues surrounding the phenomenological dimension of self-representationalism, fail to make a case against the existence of peripheral inner awareness unaccompanied by focal inner awareness.

6 Conclusion: the phenomenological case for self-representationalism

If the argumentation in Sects. 4 and 5 is correct, self-representationalism faces no serious phenomenological challenge: it can accommodate the transparency of

³⁸ I am using the term “noticing” so as to cover both introspective evidence and the “general impression.” Thus one may notice something in one’s mental life by introspecting it or by having a general impression of it.

³⁹ These considerations are developed more fully in Kriegel (2004).

⁴⁰ I would like to agree with Gennaro, however, that there is a danger of doing phenomenology in a naïve way. But then again, it is always possible to construct bad arguments for good conclusions! I should hope that the discussion over the past two sections comes across as sensitive to the potential pitfalls of naïve phenomenologizing. There may be a more general warning in Gennaro’s claim here, namely, that introspecting a slice of experiential life can change the phenomenological reality we are attempting to uncover. But I certainly concur: my main objection to the transparency argument against self-representationalism was precisely that introspection changes the introspected experience by destroying its phenomenologically manifest component of peripheral inner awareness.

experience and overcomes Gennaro's objections. If the argumentation in Sects. 2 and 3 is correct, there is also a positive phenomenological argument *for* self-representationalism. Very roughly, the argument takes the following form: the phenomenology of conscious experience involves peripheral inner awareness; only self-representationalism can accommodate this; therefore, self-representationalism is true.⁴¹

References

- Armstrong, D. M. (1968). *A materialist theory of the mind*. New York: Humanities Press.
- Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Block, N. J. (1995). On a confusion about the function of consciousness. *Behavioral and Brain Sciences*, 18, 227–247.
- Block, N. J. (1996). Mental paint and mental latex. *Philosophical Issues*, 7, 19–49.
- Dennett, D. C. (1991). *Consciousness explained*. Cambridge, MA: MIT Press.
- Dretske, F. I. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Drummond, J. J. (2006). The case(s) of (self-)awareness. In Kriegel and Williford.
- Gennaro, R. J. (1996). *Consciousness and self-consciousness*. Philadelphia and Amsterdam: John Benjamin Publishing Co.
- Gennaro, R. J. (forthcoming). Representationalism, peripheral awareness, and the transparency of experience. *Philosophical Studies*.
- Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives*, 4, 31–52.
- Hill, C. (1988). Introspective awareness of sensations. *Topoi*, 7, 11–24.
- Husserl, E. (1928). *Phenomenology of internal time-consciousness*, In Heidegger M. (Ed.), (trans. J. S. Churchill). Bloomington IN: Indiana University Press, (1964).
- Janzen, G. (2006). The representational theory of phenomenal character: A phenomenological critique. *Phenomenology and the Cognitive Sciences*, 5, 321–339.
- Kriegel, U. (2003). Consciousness as intransitive self-consciousness: Two views and an argument. *Canadian Journal of Philosophy*, 33, 103–132.
- Kriegel, U. (2004). Consciousness and self-consciousness. *The Monist*, 87, 185–209.
- Kriegel, U. (2005). Naturalizing subjective character. *Philosophy and Phenomenological Research*, 71, 23–57.
- Kriegel, U. (2006). The same-order monitoring theory of consciousness. In Kriegel and Williford.
- Kriegel, U. (2007). A cross-order integration hypothesis for the neural correlate of consciousness. *Consciousness and Cognition*, 16, 897–912.
- Kriegel U., & Williford K. W. (Eds.). (2006). *Self-representational approaches to consciousness*. Cambridge, MA: MIT Press.
- Levine, J. (2001). *Purple haze*. Oxford and New York: Oxford University Press.
- Lycan, W. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
- Moore, G. E. (1903). The refutation of idealism. In *Philosophical papers*. London: Routledge and Kegan Paul.
- Rosenthal, D. M. (1990). A theory of consciousness. ZiF Technical Report 40, Bielfeld, Germany. Reprinted in N. J. Block, O. Flanagan, & G. Guzeldere (Eds.), *The nature of consciousness: Philosophical debates*. Cambridge, MA: MIT Press, 1997.
- Rosenthal, D. M. (2002). How many kinds of consciousness? *Consciousness and Cognition*, 11, 653–665.
- Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Oxford University Press.
- Smith, D. W. (1986). The structure of (self-)consciousness. *Topoi*, 5, 149–156.

⁴¹ For comments on a previous draft, I would like to thank Rocco Gennaro, Kristie Miller, Dan Zahavi, and an anonymous referee. I have also benefited from presenting this paper to audiences at CUNY, the University of Arizona, and the Australasian Association of Philosophy annual meeting; I would like to thank the audiences there, in particular Stephen Biggs, Jared Blank, Richard Brown, Brian Fiala, Benjamin Kozuch, David Rosenthal, and Josh Weisberg.

- Thomas, A. P. (2003). An adverbial theory of consciousness. *Phenomenology and the Cognitive Sciences*, 2, 161–185.
- Thomasson, A. L. (2000). After Brentano: A one-level theory of consciousness. *European Journal of Philosophy*, 8, 190–209.
- Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.
- Williford, K. (2006). The self-representational structure of consciousness. In Kriegel and Williford.
- Zahavi, D. (1998). Brentano and Husserl on self-awareness. *Études Phénoménologiques*, 27–28, 127–169.
- Zahavi, D. (2004). Back to Brentano? *Journal of Consciousness Studies*, 11, 66–87.