

RECEPTIVITY AND PHENOMENAL SELF-KNOWLEDGE¹

Dr Tom McClelland

In this paper I argue that an epistemic question about knowledge of our own phenomenal states encourages a certain metaphysical picture of consciousness according to which phenomenal states are reflexive mental representations. Section 1 describes and motivates the thesis that phenomenal self-knowledge is 'receptive': that is, the view that a subject has knowledge of their phenomenal states only insofar as they are inwardly *affected* by those states. In Sections 2 and 3 I argue that this model of phenomenal self-knowledge is unable to accommodate knowledge of our own phenomenology or knowledge of our own awareness. In Section 4 I seek a non-receptive model of phenomenal self-knowledge. I argue that Kriegel's (2009) Self-Representationalist theory of consciousness is uniquely equipped to show how phenomenal self-knowledge is possible.

1. RECEPTIVE SELF-KNOWLEDGE

I take it that self-knowledge is any knowledge a subject has of their own states – states of themselves. The knowledge a subject has of their own *phenomenal* states is a species of such self-knowledge. We can divide phenomenal self-knowledge into two kinds: reflective and pre-reflective. *Reflective* phenomenal self-knowledge is the kind of knowledge we gain by explicitly attending to our own experiences. Consider, for example, the sophisticated knowledge a phenomenologist might gain when they carefully introspect a perceptual experience. *Pre-reflective* phenomenal self-knowledge is the kind of knowledge we have without such explicit reflection. For instance, when the phenomenologist is *not* introspecting they are still aware of their own experience. This is the kind of knowledge that unsophisticated conscious subjects, such as infants and animals, have of their experience. Such pre-reflective self-knowledge is not *propositional* and does not require sophisticated *concepts* (or perhaps *any* concepts). Nevertheless, there remains a clear sense in which these subjects 'know' about a state of themselves; they are in touch with their own experience.²

I am concerned with just one of the many interesting questions raised by knowledge of our own experiences: must phenomenal self-knowledge be causally mediated? Put another way, should we assent to the following thesis:

¹ Thanks to everyone at the NIP early-career conference for a stimulating discussion and to two anonymous referees for their probing comments. Special thanks are due to David Barnett for his insightful response to the paper. This article was completed with the support of ERC Grant 313552: "The Architecture of Consciousness".

² Some might wish to reserve the label 'knowledge' for propositional knowledge. The important claim for our purposes is that to be pre-reflectively aware of a thing is to stand in an epistemic relation to that thing. Russell's (1917) notion of 'knowledge by acquaintance' goes some way toward capturing this. However, describing this relation as 'knowledge' is inessential to my argument.

Receptivity of Phenomenal Self-Knowledge (RPSK): A subject has knowledge of their own phenomenal states only insofar as they are affected by those phenomenal states.

The best way to unpack and appreciate RPSK is to consider two key arguments that might be put forward in its favour. The first starts from a general epistemological thesis – the Receptivity principle – then applies it to the specific case of phenomenal self-knowledge. The second starts from a metaphysical account of phenomenal consciousness – Higher Order Representationalism – then extrapolates its epistemological implications for phenomenal self-knowledge.

Plausibly, a subject can only gain knowledge of a concrete entity if there is a chain of influence between that entity and the subject. Subjects must be *receptive* to objects of knowledge, for ‘...how should our capacity for knowledge be awakened into action, if objects did not affect our senses, and partly of themselves produce representations...?’ (Kant, B1) We can capture this general principle as follows:

Receptivity: A subject can have knowledge of concrete entities (objects, properties, states) only insofar as they are *affected* by those entities.³

The basic idea behind Receptivity is that we are only in contact with reality if our representations of the world are appropriately responsive to how things actually are. Without such responsiveness our mental lives would be, in McDowell’s words, a ‘frictionless spinning in the void’ (1996, p. 11). Receptivity complements (though is not equivalent to) a counter-factual condition on knowledge. When a subject *S* accurately believes that *p*, this can only qualify as knowledge if, had *p* not been the case, *S* would *not* have believed that *p*. Similarly, an accurate perceptual representation of an object *O* can only provide perceptual knowledge of *O* if, had *O* been absent, *S* would *not* have represented *O*’s presence. How do we achieve this counter-factual sensitivity? Advocates of Receptivity claim that a *causal* connection between object and subject must be what does the job. This is by no means a distinctively Kantian thought – many contemporary epistemologists would gladly accept some version of this principle. It is useful to distinguish two commitments of Receptivity.

The Entity Constraint: A subject can only have knowledge of *X* if they are causally affected by *X*.

The Powers Constraint: A subject can only have knowledge of *X* in terms of its powers to bring about certain effects.

The first constraint places restrictions on *what* a subject can have knowledge of, while the second places restrictions on what a subject can know *about* that entity. The second claim is stronger than the first, but follows naturally from Receptivity. Receptive knowledge can reveal the powers that entities have. Knowing that an entity has some power means knowing that it will produce certain effects in certain circumstances; for instance, knowing *X* is a poison means knowing that it will cause harm if ingested. Madden & Hare explain that ‘[t]o ascribe a power to a thing asserts only that it can do what it does in virtue of its nature, whatever that is. It leaves open the question of the exact specification of the nature or constitution in virtue of which [it] has the power.’ (quoted Allais,

³ Since we cannot stand in causal relations to abstracta, Receptivity only credibly applies to knowledge of concrete entities.

2006 p. 155) Hence knowing that X is a poison does not reveal the *ground* of X's power to do harm. Receptive knowledge is thus inherently limited; it can reveal how entities are causally related to *other* entities but leaves us ignorant of what those entities are like *in themselves*. Advocates of Receptivity should therefore follow Kant in concluding that '[w]e have no insight whatsoever into the intrinsic nature of things.' (A277/B333). There are a number of competing formulations of this bold thesis of epistemic humility.⁴ For our purposes, the important lesson is that if our knowledge of X reveals more about X than the powers it bears, then our knowledge of X is not receptive.

What does Receptivity have to say about *self*-knowledge? Presumably what goes for knowledge of *external* entities also goes for knowledge of *internal* entities. In other words, self-knowledge must be receptive too. Kant captures this view neatly:

If ... we admit that we know objects only in so far as we are externally affected, we must also recognise, as regards inner sense, that by means of it we intuit ourselves only as we are inwardly affected by ourselves. (B156)

Applying this to the specific case of *phenomenal* self-knowledge, we get the conclusion that a subject's knowledge of their own phenomenal states is causally mediated. That is, we get RPSK. The argument here is that since phenomenal self-knowledge has to be receptive, we must know our own experiences through a self-monitoring mechanism. The second argument for RPSK is a mirror-image of this. Advocates of Higher Order Representationalism (HOR theory) hold, for independent reasons, that we know our experiences through a self-monitoring mechanism, and so from this can infer that phenomenal self-knowledge is receptive.

HOR theory is an account (or family of accounts) of what makes a mental state conscious. A mental state is conscious just when its subject is *aware* of being in it. Awareness is a species of representation, therefore a mental state is conscious only if it is appropriately represented by another of the subject's mental states. Different HOR theorists tell different stories about the specific relationship between the higher- and lower-order mental states: the higher-order state might be a thought directed at the lower (Rosenthal 1986), or might instead be a quasi-perceptual representation (Lycan 1996). Either way though, we know our experiences *via* higher-order representations that are responsive to those experiences. In other words, consciousness yields a *causally mediated* knowledge of our own phenomenal states. As such, HOR theorists of all stripes have reason to advocate RPSK.

These two arguments in favour of RPSK are far from conclusive. They will only convince those already attracted to Receptivity or HOR theory (or both). Receptivity is not without its problems: knowledge of the future, knowledge of our own actions and knowledge of general truths are all difficult to account for in causal terms. HOR theory faces difficulties too: the targetless higher-order representation problem, the generality problem and the self-intimation problem are all well-documented (Kriegel 2009, pp. 129-155). My motivation for sketching the arguments is two-fold. First, they show why one might find RSPK attractive before I go on to attack its credibility in Sections 2 and 3. Second, they show that my rejection of RSPK will commit me to a rejection of Receptivity and HOR theory. This sets up the challenge I address in Section 4 of providing a non-receptive model of phenomenal self-knowledge.

⁴ See Russell (1927), Langton (1998), Allais (2006), Lewis (2009) and Pereboom (2011).

2. RECEPTIVITY AND KNOWING WHAT IT'S LIKE

There is *something it's like* to be in a phenomenal state for the subject of that state. When you are in a phenomenal state you know *what it is like to be* in that phenomenal state. When we are aware of our own phenomenology – of how things seem to us, of our phenomenal qualities, of the qualitative character of our experience – we enjoy a kind of self-knowledge. The existence of this self-knowledge is hard to deny, but I will present two arguments that show why RPSK is committed to the impossibility of such knowledge.

Argument A: Not Knowing What It's Like and the Entity Constraint

- A1) If a subject is aware of some object, property or state X at some time *t*, then they have knowledge of X at *t*.
- A2) Therefore, if a subject is aware of their phenomenology at *t* then that subject has knowledge of their phenomenology at *t*. [A1]
- A3) A subject only has knowledge of X if they are causally affected by X.
- A4) If Y is causally affected by X, then X and Y are distinct existences.⁵
- A5) A subject at *t* and their phenomenology at *t* are not distinct existences.
- A6) Therefore, a subject at *t* cannot be causally influenced by their phenomenology at *t*. [A4&A5]
- A7) Therefore, a subject at *t* cannot have knowledge of their own phenomenology at *t*. [A3&A6]
- A8) Therefore, a subject at *t* cannot be aware of their own phenomenology at *t*. [A2&A7]

The first premise simply notes that general epistemic principles must apply to the awareness relation as a species of knowledge relation. It also supposes that awareness requires simultaneity: it is credible that where one might have a belief about an event after the fact, one can only be *aware* of an event while it is happening. A2 slots the qualitative character of experience into the schema described by A1. The third premise is Receptivity; or specifically the Entity Constraint entailed by Receptivity. Note, HOR theorists might not adopt this as a *general* principle, but they will say it holds in the specific case of phenomenal self-knowledge. A4 is a credible claim about the metaphysics of causation. A5 proposes that a subject's phenomenology is a *modification* of that subject rather than some distinct state to which the subject is related; a conscious state and what that conscious state is like cannot be ontologically distinct.

Steps A6 and A7 follow from the five premises, and we are left with the conclusion that a subject has no knowledge of their own phenomenology. Clearly this is not a conclusion we can accept. Introspection provides knowledge of what it is like to be us right now. Furthermore, even when we are *not* explicitly attending to our experience, it is credible that we still have a pre-reflective knowledge of our phenomenology. To be in a phenomenal state *just is* for things to seem

⁵ A possible objection here is that *parts* can affect *wholes* e.g. an insult can be the cause of a fight and be a part of it (i.e. not distinct from it). I will exploit such mereological relations in Section 4, though I have reservations about regarding this as *causation* in the relevant sense.

some way to you. A phenomenology of which no subject is aware is simply not a phenomenology, so perhaps *every* conscious experience is a counter-example to A8.

In order to avoid this problematic conclusion, one might reject A5. HOR theory suggests that your state of awareness and the qualitative character of your awareness are constituted by distinct mental states. We have a lower-order state with a certain non-conscious qualitative character. When a higher-order state represents that lower-order state, we become aware of that qualitative character. It remains the case that there can be no awareness without a qualitative character, but on this view the character is provided by what you are aware *of* rather than by the state through which you are aware of it. Since the phenomenology and the awareness are distinct, it is possible for them to stand in the causal relation that Receptivity demands for knowledge.

A number of serious doubts can be raised against this 'division of phenomenal labour' model (Neander, 1998). However, I will make the stronger claim that even if the model is *true*, RPSK would *still* be unable to accommodate the kind of knowledge we have of our own phenomenology.

Argument B: Not Knowing What It's Like and the Powers Constraint

- B1) A subject can be aware of some object, property or state X only in terms of its powers to bring about certain effects.
- B2) If a subject is aware of X only in terms of its powers to bring about certain effects, then they are not aware of X as it is in itself.
- B3) Therefore, a subject can never be aware of an object, property or state X as it is in itself.
[B1&B2]
- B4) Therefore, a subject can never be aware of the phenomenal qualities of their experience as they are in themselves. [B3]

The first premise deploys Receptivity again but focuses on the Powers Constraint. B2 and B3 establish that this constraint leaves us ignorant of the underlying nature of entities; of what they are like beyond their dispositions to do certain things in certain circumstances. B4 simply slots phenomenal qualities into this schema. If our awareness of phenomenal qualities – the properties that constitute our phenomenology – is receptive then we can only be aware of those qualities as bearers of certain powers. This clearly fails to do justice to the intimacy of our awareness. When I am aware of a pain quality I am aware not just of its power to make me wince but of the underlying *nature* in virtue of which it has this power – the *painfulness* of pain. As Blackburn puts it, '...there is nothing dispositional, for the subject, in the onset of a pain or a flash in the visual field' (1990, p. 65). Consider the knowledge Mary has of qualitative redness before leaving her famous monochromatic prison (Jackson 1982). She might have a comprehensive understanding of the distinctive causal relations in which that quality stands, but there would still be something *we* know about redness that she does not. We know what it's like to be in a state with a reddish phenomenology; we know qualitative redness *as it is in itself* rather than as it is in relation to other things. But B4 denies that such knowledge is possible. So even if we charitably allow (in response to Argument A) that a subject's awareness can be distinct from the qualitative character of that awareness, the Receptivity model is still unable to accommodate the kind of knowledge we have of our own phenomenology.

3. RECEPTIVITY AND KNOWING YOU'RE AWARE

In conscious experience we are aware of what that experience is like for us but we can also be aware of our awareness. Just as we can reflect on the qualitative character of our awareness, we can reflect on the awareness itself – we can attend to our consciousness. We are, as Russell suggests, acquainted with the acquaintance relation (1917). One might even make the stronger claim that *whenever* we are aware, we are in some sense pre-reflectively aware that we are aware. As Strawson puts it, ‘...all awareness comports awareness of itself’ (2011, p. 282). On this view, states *in* which we are aware must also be states *of* which we are aware. The awareness we have of our own awareness, whether it be intermittent or persistent, is an interesting variety of phenomenal self-knowledge. We can show that RPSK is unable to accommodate such knowledge by deploying arguments parallel to A and B.

Argument C: Being Unaware of Awareness and the Entity Constraint

- C1) If a subject is aware of some object, property or state X at some time *t*, then they have knowledge of X at *t*.
- C2) Therefore, if a subject is aware at *t* of their own awareness at *t*, they have knowledge of their own awareness at *t*. [C1]
- C3) A subject only has knowledge of X if they are causally affected by X.
- C4) If Y is causally affected by X, then X and Y are distinct existences.
- C5) A subject at *t* and their awareness at *t* are not distinct existences.
- C6) Therefore, a subject at *t* cannot be causally influenced by their own awareness at *t*. [C4&C5]
- C7) Therefore, a subject at *t* cannot have knowledge of their own awareness at *t*. [C3&C6]
- C8) Therefore, a subject at *t* cannot be aware of their own awareness at *t*. [C2&C7]

It seems that advocates of RPSK must regard being aware of your awareness as analogous to jumping on your own shadow. Might some equivalent of the ‘division of phenomenal labour’ account help us make sense of self-awareness in a way that respects RSPK? One move would be to deny C5 by claiming that there is always a time-lag between a state of awareness, and the awareness of that state. Strawson (2011, p. 281) picks up on Ryle’s notion of ‘swift retrospective heed’. The idea is to scrap present-moment self-awareness, and have it that a subject at *t+1* is aware of their awareness at *t*. This is not ruled out by C8. The old and the new awareness would be distinct existences, so can stand in the requisite affective relation. On this picture, our self-awareness is a continuous stream of retrospection (perhaps with a persistent illusion of simultaneity that accounts for any intuitions to the contrary).

The retrospective model will not satisfy those who hold that *all* states of awareness are states of which the subject is aware. If *every* state of awareness is retrospected by some later state of awareness then there would have to be infinite chains of retrospection. On the other hand, if these chains terminate in states of awareness that are not retrospected, then there would be states of awareness of which the subject is unaware (Strawson 2011, p. 281). Putting this issue to one side

though, there is a deeper reason for thinking that RPSK is unable to accommodate awareness-of-awareness, regardless of whether or not that awareness is retrospective. This argument runs parallel to B, and deploys the Powers Constraint.

Argument D: Being Unaware of Awareness and the Powers Constraint

D1) A subject can be aware of some object, property or state X only in terms of its powers to bring about certain effects.

D2) If a subject is aware of X only in terms of its powers to bring about certain effects, then they are not aware of X as it is in itself.

D3) Therefore, a subject can never be aware of an object, property or state X as it is in itself. [D1&D2]

D4) Therefore, a subject can never be aware of awareness as it is in itself. [D3]

Notice that D1 has no time-index, so simultaneity is not a factor. D2 and D3 extrapolate the restrictions that Receptivity places on the kind of knowledge we can have of an entity. D4 slots awareness into the schema and concludes that our awareness-of-awareness can reveal the causal powers of awareness, but cannot disclose the underlying nature in virtue of which it has those powers. This clearly does not do justice to our knowledge of awareness. Consider the knowledge a zombie scientist could have of awareness: they could be causally affected by states of awareness and build a comprehensive picture of its causal powers. But this knowledge would be dramatically incomplete – the zombie scientist would be unacquainted with *what awareness is*. On the receptive model, our epistemic situation as regards our *own* awareness would be analogous to the epistemic situation of the zombie scientist. We would be aware of what our awareness *does* but the deeper non-functionalizable nature of awareness would be hidden from us. The knowledge we have of our own awareness is clearly more intimate than this, but RSPK is unable to accommodate this knowledge.

4. CONSCIOUSNESS AND REFLEXIVE REPRESENTATION

It is clear that RPSK is entirely unable to accommodate the distinctive self-knowledge afforded by phenomenal consciousness. How then *can* we accommodate such self-knowledge? I propose that the notion of reflexive representation holds the key. In reflexive states, the *representing* and the *represented* are one and the same. This puts the subject of that representational state in a distinctive epistemic position. I will sketch Kriegel's Self-Representationalist theory of consciousness before showing how it can overcome the epistemic problems faced by RPSK. I will also show how Self-Representationalism can go some way to appeasing those who find Receptivity and/or HOR theory attractive.

Self-Representationalism (SR) proposes that '...a state has subjective character in virtue of suitably representing itself...' (Kriegel 2009, p. 2).⁶ That is, a mental state is a state of awareness – a

⁶ Kriegel fleshes out 'suitably' so that subjectivity does not come too cheap (2009, pp. 155-164). Having a self-representing belief, for instance, is not sufficient for subjectivity.

conscious state – just when it appropriately represents its own occurrence. Like HOR theory, SR is driven by the familiar insight that to be aware of a mental state is to represent it in some way. HOR theorists assume that the representing and represented states are *distinct*. In contrast, SR respects the need for higher-order representation but claims that the higher- and lower-order content belong to the very same mental state: a state that represents itself. Adopting a ‘one-state’ view allows SR to overcome the key objections to HOR theory. For instance, difficulties surrounding targetless higher-order representations are avoided by SR, because ‘...it is incoherent to suppose that a mental state may represent itself to exist when in reality it does not exist.’ (Kriegel, 2009, p. 136)

I will not digress into an examination of the arguments for SR. For my purposes, it is SR’s implications for the epistemology of phenomenal self-knowledge that are important. Kriegel emphasises that according to SR ‘...knowledge of consciousness, and of consciousness alone, does not require causal contact with the known.’ (2009, p. 295) It could well be the case that a subject can know things *distinct from* itself only insofar as it is affected by them. SR can thus respect the driving thought behind the Receptivity model. However, according to SR a subject can have a kind of *non-causal* way of knowing states of itself. This does not open the floodgates to counter-examples to Receptivity. Rather, the claim is that self-knowledge is a *principled exception* to the model. We can make sense of this reflexive knowledge in terms of those all-important counter-factual conditionals. Part of what it is for a subject to have knowledge of X is for states of the subject to ‘track’ X. Causation is one way of providing this tracking feature, but *constitution* is another. If X is partially constitutive of Y, then changes to X can constitute changes to Y. In self-representation, the represented is partially constitutive of the representation. Any changes to the object of representation are thus inevitably accompanied by changes to the representing state.

How can SR accommodate the pre-reflective knowledge we have of the qualitative character of our experience? On this view we know the phenomenology of a state by *being in* that state, not by accessing that phenomenology from some distinct vantage point. Qualitative character *stands for itself* rather than being represented by some causal proxy. Phenomenal states represent *their own* qualitative character, so can give a subject unmediated access to their own phenomenology. SR allows us to qualify the claims of Receptivity in arguments A and B. A3 should be rejected because not all knowledge requires causal contact, so the ‘knowing’ state and the ‘known’ state being indistinct need not preclude a knowledge relation between them. B1 should be rejected because in reflexive awareness we are not limited to knowing our own states in terms of their powers. Nothing precludes us from accessing the underlying nature of our phenomenal qualities – from knowing our phenomenology as it is in itself. Importantly, the ‘tracking’ condition on knowledge is still satisfied by this account: say you are reflexively representing the reddish quality of your phenomenology. If your experience changes to one with a bluish phenomenology, there will be a corresponding change to your representational state. This is because the represented is a constituent of the representation.

SR can also accommodate a subject’s awareness of their own awareness. Again, the crucial epistemic premises in arguments C and D are qualified. In reflexive representation, awareness and awareness-of-awareness can be provided by one and the same self-representing state. Furthermore, the knowledge that a subject has of their own awareness need not be limited to its powers. By being in a state of awareness, we can know awareness *as it is in itself* rather than merely knowing it as it relates to other entities.

Overall, RPSK is unable to accommodate phenomenal self-knowledge. In phenomenal consciousness the subject of experience can have (and perhaps *always* has) a special knowledge of both the qualitative character of their experience and of their awareness of that qualitative character. I have argued that SR offers a credible account of how such knowledge is possible. It denies Receptivity, but respects its spirit by showing that consciousness is a principled exception to the rule that knowledge requires causal contact with the known. It denies HOR theory, but respects its spirit by maintaining that higher-order content is integral to what makes a mental state conscious. SR may not provide us with a detailed account of the mechanisms behind phenomenal self-knowledge, but it does show how such self-knowledge is *possible* in the face of the difficulties explored in this paper. We are thus encouraged toward the view that phenomenal self-knowledge is not receptive, and that phenomenal states are reflexive mental representations that can provide their subject with causally unmediated knowledge of that very state.

BIBLIOGRAPHY

- Allais, L. (2006) 'Intrinsic Natures: A Critique of Langton on Kant', *Philosophy and Phenomenological Research*, 23(1), 143-169
- Blackburn, S. (1990) 'Filling in Space', *Analysis*, 50, 62-65
- Jackson, F. (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32, 127-136
- Kant, I. (1781A/1787B) *The Critique of Pure Reason*. Pluhar, W.S. trans. (1996), Indianapolis: Hackett
- Kriegel, U. (2009) *Subjective Consciousness: A Self-Representational Theory*. Oxford: OUP
- Langton, R. (1998) *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Clarendon Press
- Lewis, D. (2009) 'Ramseyan Humility', in Nola, R. & Braddon-Mitchell, D. (eds.) *Conceptual Analysis and Philosophical Naturalism*. Cambridge, Mass: MIT Press, pp. 203-222
- Lycan, W.G. (1996) *Consciousness and Experience*. Cambridge, MA: MIT Press
- McDowell, J. (1996) *Mind and World* 2nd edition. Cambridge, Mass: Harvard University Press
- Neander, K. (1998) 'The Division of Phenomenal Labour: A Problem for Representational Theories of Consciousness', *Noûs*, 32(12), 411-434
- Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*. Oxford: OUP
- Rosenthal, D. (1986) 'Two Concepts of Consciousness' *Philosophical Studies*, 49, 329-359
- Russell, B. (1917) 'Knowledge by acquaintance and knowledge by description' in his *Mysticism and Logic*. London: George Allen & Unwin Ltd.
- Russell, B. (1927) *The Analysis of Matter*. London: George Allen & Unwin Ltd.
- Strawson, G. (2011) 'Radical Self-Awareness' in Sidertis, M., Thompson, E. and Zahavi, D. (eds.) *Self, No Self?: Perspectives from Analytical, Phenomenological and Indian Traditions*. Oxford: OUP, pp. 274-307