SELF-REPRESENTATIONALIST THEORIES OF CONSCIOUSNESS

Dr Tom McClelland

University of Warwick

The Meta-Representationalist Family

To understand Self-Representationalism (SR) you need to understand its family. Self-Representationalism is a branch of the Meta-Representationalist family, and according to theories in this family what distinguishes conscious mental representations from unconscious mental representations is that conscious ones are themselves the target of a mental *meta-*representational state. A mental state M_1 is thus phenomenally conscious in virtue of being suitably represented by some mental state M_2. What distinguishes the Self-Representationalist branch of the family is the claim that M_1 and M_2 must be the same token mental state, so a mental state is phenomenally conscious in virtue of suitably representing itself.[1] This Self-Representationalist branch of the family divides into further branches, giving us specific implementations of the Self-Representationalist approach. But before asking whether we should adopt Self-Representationalism, and in what form, we should reflect on

---

[1] Some Meta-Representationalist theories of consciousness aren't specifically presented as theories of *phenomenal* consciousness (see Block 2011 for discussion of the some of the complications surrounding this fact). Interestingly though, Self-Representationalist theories seem to universally be presented as theories of phenomenal consciousness, so I will continue on the assumption that the target of all the theories under discussion is indeed phenomenal consciousness. To that end, I will generally drop the 'phenomenal' qualifier from here on.

why Meta-Representationalism is an attractive family in the first place. After all, Self-Representationalist theories trade on their family name, claiming to deliver on the promises that drive the Meta-Representationalist approach. The two most important promises of Meta-Representationalism are: a) the promise of capturing the *transitivity* of consciousness and; b) the promise of rendering consciousness *naturalisable*. I discuss each in turn.

What's the difference between conscious mental states and non-conscious mental states? A plausible initial answer is that conscious mental states are those of which you're aware. In other words, a mental state of yours is conscious in virtue of you being conscious *of* it. There are two different senses of 'conscious' in play here that need to be distinguished (see Rosenthal 1986; Kriegel 2009). 'Consciousness' can be used in a *transitive* sense to designate the relation of being conscious *of* something. But the same term can also be used in an *intransitive* sense to designate the non-relational property of being a conscious state. Equipped with this distinction, we can frame the foregoing a little more precisely: the question is what the difference is between mental states that are intransitively conscious and mental states that are not; the plausible initial answer is that intransitively conscious mental states are those mental states of which we are transitively conscious. According to this 'transitivity principle', transitive and intransitive consciousness are two sides of the same coin. Every theory in the Meta-Representationalist family makes the following promise:

The Transitivity Promise: We promise to offer a theory of consciousness that conforms to the principle that intransitively conscious mental states are those mental states of which the subject is aware.[2]

How do Meta-Representationalist theories hope to deliver on that promise? They do so by making the simple claim that when we talk of being aware *of* a mental state, the 'of' is the 'of' of intentionality (Lycan 2001). Just as being a picture *of a house* is a matter of representing a house, so too being aware *of a mental state* is a matter of representing that state. Meta-Representationalists thus vindicate the principle that conscious mental states are those

---

[2] For a useful examination of arguments for and against the transitivity principle, see the entry on Higher-Order Theories of Consciousness in this volume. A number of arguments for the principle are helpfully catalogued in the Appendix to Kriegel's (2009). Thomasson (2006) offers a particularly insightful critique of the main arguments. Many critics of Self-Representationalism object to the transitivity principle (e.g. Gertler 2012; Siewert 2013; Seager 2006; Lyrra 2008). Though these objections deserve to be taken seriously, I will not discuss them here. Whether the transitivity principle is well-motivated is a problem for the whole Meta-Representationalist family. In order to determine the relative merits of Self-Representationalism, it will be best to focus on the issues that distinguish it from its relatives, and that distinguish different versions of Self-Representationalism from each other. As such, I will continue on the assumption that the Transitivity Promise is a promise worth delivering on.

mental states of which we are aware by cashing out 'awareness-of' in representational terms, and claiming that conscious states are those mental states that we suitably represent.[3]

Can phenomenal consciousness be naturalised? A common response is that the answer *ought* to be yes, but that we're not sure how it *could* be. The answer ought to be yes because naturalism is an incredibly successful world view that we should be reluctant to compromise, and because anti-naturalist views face a swathe of problems regarding how consciousness fits into an otherwise naturalistic world. Yet we're not sure how the answer *could* be yes because phenomenal consciousness is peculiarly and recalcitrantly resistant to naturalisation. Consequently, if a theory can improve the prospects for naturalising consciousness it would be a major point in its favour. This leads us to the second promise of Meta-Representationalism.

> **The Naturalisability Promise:** We promise to offer a theory of consciousness
>
> amenable to naturalisation.[4]

---

[3] The 'suitably' clause is designed to accommodate the fact that not just any representation of one's mental state suffices to make that state conscious. The entry on Higher-Order Theories of Consciousness in this volume discusses some of the details of this clause.

[4] Unlike the Transitivity Promise, not every Meta-Representationalist promises to offer a naturalistic theory. Although we will be looking at some non-naturalist theories later on, the point remains that the Naturalisability Promise is a key motivation for the majority of Meta-Representationalists, and Self-Representationalist theories ought to be evaluated with respect to their capacity to deliver on that promise.
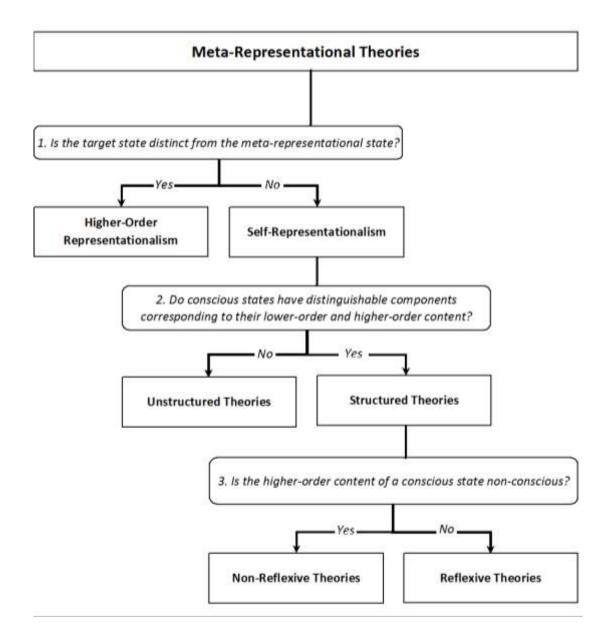
How does Meta-Representationalism hope to deliver on this promise? The key is the naturalisability of mental representation. We have already seen how Meta-Representationalists aim to account for consciousness in representational terms, so if representation can in turn be explained naturalistically we will have a naturalistic account of consciousness. It must be conceded that an adequate naturalistic theory of mental representation has not yet been developed. However, many feel that the prospects for naturalising mental representation are strong, so if Meta-Representationalist theories can explain consciousness in representational terms they will have made good on their promise. They might not have actually explained consciousness in naturalistic terms, but they will have offered an account of consciousness amenable to naturalisation in the long run.[5]

Having familiarised ourselves with what characterises and motivates the Meta-Representationalist family as a whole, we're now ready to start distinguishing the different households that make up that family. Below is the Meta-Representationalist family tree:

---

[5] This move needs to be taken with a hefty pinch of salt (Kidd 2011). Debates around the naturalisation of intentionality have been raging for some time and have no immediate end in sight. It's an open possibility that intentionality will transpire to be unnaturalisable, meaning that representational theories of consciousness will ultimately fail to deliver on their promise of naturalisation. Nevertheless, the prospects of naturalising intentionality are better than the immediate prospects of naturalising consciousness, so explaining consciousness in representational terms at least *improves* the prospects of naturalising consciousness.

Meta-Representational Theories

1. Is the target state distinct from the meta-representational state?

— Yes → Higher-Order Representationalism

— No → Self-Representationalism

2. Do conscious states have distinguishable components corresponding to their lower-order and higher-order content?

— No → Unstructured Theories

— Yes → Structured Theories

3. Is the higher-order content of a conscious state non-conscious?

— Yes → Non-Reflexive Theories

— No → Reflexive Theories

As you can see, the Meta-Representationalist family tree is structured around three choice-points. The first choice-point is where Self-Representationalism differentiates itself from its well-established relatives: the Higher-Order Representationalist (HOR) theories of consciousness. The second and third choice-points allow us to differentiate the various households in the Self-Representationalist branch of the family. As we work our way down the family tree, we'll introduce the key members of each household and identify the problems that plague each of them. Although different theories face different problems, the

overarching theme of the paper is that no existing version of Meta-Representationalism can satisfy all three of the following desiderata:

i)   To deliver on the Transitivitiy Promise

ii)   To deliver on the Naturalisability Promise

iii)   To respect the 'intimacy' of consciousness

After exploring the main existing forms of SR, I make some suggestions about how Self-Representationalists might develop a version of the theory that satisfies all three desiderata.

The First Choice-Point: Is the Target State Distinct from the Meta-Representational State?

Take a subject S who is enjoying a visual experience of the sunset. Meta-Representationalists explain the subject's experience in terms of her having a mental state M_1 that perceptually represents the sunset, and a mental state M_2 that represents her as being in M_1. The first choice-point for Meta-Representationalism concerns whether M_1 and M_2 are distinct, or whether they're one and the same token mental state. Higher-Order Representationalism chooses the former: it explains S's experience in terms of two mental states, one with the lower-order content and the other with the higher-order content. Self-Representationalism chooses the latter: it explains S's experience in terms of her having a single mental state that represents both the sunset *and* itself.

Option 1: Higher-Order Representationalism (HOR)

HOR comes in many different flavours, and different versions tell different stories about the format of the higher-order representation required for consciousness. Some claim that the

higher-order state is a *thought* about the lower-order state (e.g. Rosenthal 1986; Carruthers 2004). Others claim it is more akin to an inner *perception* of the lower-order state (e.g. Armstrong 1981; Lycan 2004). A case can also be made for regarding the higher-order state as having a status somewhere between that of ordinary thoughts and ordinary perceptual states. HOR theorists also tell different stories about the exact relationship between M_1 and M_2 required for consciousness. What each of these views has in common is the claim that M_1 and M_2 are distinct.[6]

Why do Higher-Order Representationalists claim that M_1 and M_2 are distinct states? Some explicit arguments can be found in the literature, but the main reason seems to be that this is the *default* way of formulating Meta-Representationalism. If M_1 is conscious in virtue of being targeted by a mental state M_2, it is natural to assume that M_2 will be a distinct mental state. Meta-Representationalists should only make the bolder claim that a mental state is only conscious when it is represented *by itself* if there are specific reasons to do. Of course, Self-Representationalists claim to have just such reasons for taking the other branch of this choice-point. Although a number of problems have been raised against the view that M_1 and M_2 are distinct, the most serious is that HOR theorists have trouble with intimacy.[7] When we are in a conscious state we have a distinctively intimate relationship with that state: our conscious

---

[6] For a more detailed survey of the varieties of position available, see the entry on Higher-Order Theories of Consciousness in this volume.

[7] I borrow this phrase from Weisberg (2008). The intimacy problem for Higher-Order Representationalism has been put forward (in varying forms) by: Neander (1998); Levine (2003; 2010); Gennaro (2006); Kriegel (2009); Kidd (2011); Picciuto (2011); Block (2011); Van Gulick (2012); Coleman (2015) and others.

state seems to be immediately disclosed to us. One way of putting this is that there can be no gap between how a conscious state *seems* to us and how it *really is*.[8] The apparent intimacy of consciousness has two aspects.

First, consciousness is *qualitatively* intimate insofar as the qualities that we experience our conscious state as having are necessarily qualities it really has. For instance, if we experience our conscious state as having a painful quality, then it must actually have that quality. Consciousness doesn't allow for a mismatch between the qualities that we experience our conscious state as having and the qualities it really has. Second, consciousness is *existentially* intimate insofar as experiencing ourselves as being in a conscious state guarantees that we really are in a conscious state. Experiencing a conscious state guarantees the existence of that state. Consciousness doesn't allow for cases where we experience ourselves as being in a conscious state when no such state exists.

The problem for HOR theorists is that the relation they posit between the distinct states M_1 and M_2 falls short of the intimacy of consciousness. Where a representation is distinct from what it represents, there is the possibility of error. It might be that the object represented lacks the properties attributed to it, like when a perceptual state represents a pencil in water as bent. Or it might be that the object represented doesn't exist at all, like when Macbeth's

---

[8] Note, the claim here is that there is no gap between how a conscious state *phenomenally* seems to us and how it really is. This is quite consistent with the possibility of our *judgements* about our conscious states being in error: a state might phenomenally seem to be an itch and yet cognitively seem to be a pain. The intimacy of consciousness should not be confused with claims about the infallibility of phenomenal judgements.

perceptual state represents a dagger before him. Since M_1 and M_2 are distinct, there is nothing to preclude M_2 from misrepresenting M_1 in either of these ways.

First, M_2 might misrepresent the qualities of M_1. Perhaps M_1 is a tactile representation of itchiness, but M_2 misrepresents it as a tactile representation of pain. In such a case, the pain quality that we represent our conscious state as having diverges from the itchy quality that it actually has. Second, M_2 might misrepresent the very existence of M_1. Perhaps M_2 is a targetless state that represents the subject as being in pain, but fails to represent any actual lower-order state of the subject. The intimacy of consciousness seems to preclude these kinds of error, yet HOR is committed to the possibility of such errors occurring. Note, it won't help to insist that the processes responsible for higher-order representation are reliable so such misrepresentations will never actually occur. The point is that the apparent intimacy of consciousness seems to make such errors *impossible*, but if M_2 and M_1 are distinct states then it is at least possible for M_2 to misrepresent M_1. The representation relation described by HOR is consistent with such error, but the transitive consciousness relation is not, so HOR has not successfully captured the transitive consciousness relation.

HOR theorists have responded to the problem of intimacy by proposing a constitutive connection between how a subject's higher-order state represents their mental life as being and the phenomenology they actually undergo (e.g. Rosenthal 2011; Weisberg 2011b). On this view, if you *represent* yourself as having a pain experience then you thereby *really have* a pain experience. If the lower-order state you represent is actually a representation of an itch, your experience will still be painful because what matters is how the lower-order state *seems* to you. It doesn't even matter if your higher-order state fails to target any actual lower-order state, you'll still have a pain experience because it appears to you that you are in pain.

There is a sense in which this constitutive view collapses the distinction between how a conscious state seems and how a conscious state is: how your higher-order state represents things as being cannot diverge from the experience you actually have.[9]

This response goes at least some way to accommodating the apparent intimacy of consciousness, but at what cost? By proposing a constitutive link between the content of a subject's higher-order state and her phenomenology, the HOR theorist risks reneging on the Transitivity Promise. Remember, HOR theory promises to capture the principle that conscious states are those mental states of which we are aware. But on the current picture, the state of which we are aware drops out as irrelevant. A subject's pain experience can't plausibly be identified with her represented lower-order state in either scenario: in the first scenario the represented state is an itch not a pain, and in the second scenario the represented state doesn't even exist. If the pain experience cannot be identified with a lower-order state, perhaps it should be identified with the higher-order state? After all, the HOR theorist is claiming that the subject's higher-order state is wholly responsible for her pain experience. The problem with this move is that this higher-order state isn't represented. This leaves the HOR theorist with a choice of reneging on the claim that being in a conscious state requires one to be aware of that conscious state, or reneging on the claim that being aware of our conscious state requires us to suitably represent it. Either way, HOR will have gone back on the Transitivity Promise.

---

[9] This move is discussed in more detail the Higher-Order Theories of Consciousness entry in this volume.

Can the HOR theorist take a different tack and simply deny that we have the proposed intimate relation to our conscious states? Some say that the intimacy of consciousness is guaranteed *conceptually*: that it is part of our concept of phenomenal consciousness, and of phenomenal qualities, that the way a conscious state appears to us is the way it is (e.g. Kidd 2011). I recommend allowing that how best to conceptualise consciousness is still up for grabs (after all, we don't want to accidentally define features into consciousness that it might not really have). A better response to the HOR theorist who denies the intimacy of consciousness is that such a denial is *phenomenologically* implausible. Reflection on our experience strongly suggests that we have an intimate relationship with our conscious states that precludes the kinds of error countenanced by HOR. There is an open possibility that our phenomenology is misleading, but the burden is on the HOR theorist to explain why consciousness has this misleading appearance. So as things stand, there is a pervasive phenomenological feature of experience – its apparent intimacy – for which HOR has difficulty accounting.

Overall, the claim that M_1 and M_2 are distinct is plausibly undermined by the intimacy problem.[10] Consequently, we should go down the other branch of this choice-point to see whether matters are any better if M_1 and M_2 are held to be token identical.

---

[10] Although other objections to the distinctness route have been offered, they have less bite than the intimacy objection. For instance, some have suggested (Dretske 1995; Van Gulick 2004; Kriegel 2009) that if M_1 is made conscious by being represented by a distinct state, then by parity of reasoning non-mental entities should also become conscious when represented, yet consciousness is clearly restricted to mental entities. Rosenthal (unpublished) rightly responds that this objection rests on a misconstrual of what Higher-Order

Option 2: Self-Representationalism (SR)

As we will see, there are a number of different households on the SR branch of the family tree. What unites these households is the claim that a mental state is conscious in virtue of suitably representing itself.[11] What motivates the claim that the state of which we are aware is token identical with the state in virtue of which we are aware of it: that $M_1 = M_2$? The key motivation is the promise of capturing the intimacy of consciousness. HOR theory had trouble

---

Representationalism says about consciousness: consciousness is how your mental life appears to you, so if non-mental entities appear some way to you then that doesn't make them conscious (this response is explored in the entry on Higher-Order Theories of Consciousness in this volume). Another objection (Kriegel 2009: 139) is that if M_1 is made conscious by a distinct mental state then M_1's causal powers would remain the same, yet becoming conscious clearly changes a mental state's causal powers. However, this problem is plausibly dealt with by noting other cases in which relations confer new causal powers on an object (say, Theresa May entering into a new relation to the British state).

[11] This should not be confused with the claim that a conscious mental state is conscious in virtue of suitably representing *the* self i.e. the subject of the mental state. Some SR theories propose that the self is at least implicitly represented in consciousness (e.g. Ford 2009; Sebastian 2012), but this is by no means characteristic of the theory. One can be in a state that represents itself without being in a state that represents you as its bearer. For the purposes of this paper, I bracket questions about what role (if any) the self plays in experience, though see my (forthcoming).

with intimacy because M_2 was independent of M_1 so could misrepresent it. By proposing that a conscious state is just a single state *M** SR theorists hope to give the higher-order content of that state the requisite dependence on its lower-order content. Regarding qualitative intimacy, the hope is that since M* is a single state there is no room for a mismatch between the lower-order content that M* represents itself as having and the lower-order content it actually has. Regarding existential intimacy, the hope is that since the target of M* is itself there's no possibility of M* being targetless. SR thus adds a third promise to its list of motivations:

> **The Intimacy Promise:** We promise to accommodate the apparent intimacy of the relationship between a conscious state and the subject's awareness of that conscious state.

With these considerations in place, we can present the *Master Argument* for Self-Representationalism:

> MA1) A mental state is intransitively conscious in virtue of being suitably represented.

> MA2) It is not the case that a mental state is intransitively conscious in virtue of being represented by a numerically distinct state.

> MA3) Therefore, a mental state is intransitively conscious in virtue of suitably represents itself.[12]

The motivations for the first premise are the motivations for Meta-Representationalism discussed earlier viz. the promise of accommodating the transitivity principle and giving a

---

[12] Adapted from Kriegel (2009: 15-6).

naturalisable account of consciousness. The key motivation for the second premise is that if consciousness involves two distinct states then you run into the troubles with intimacy considered above. From these two premises, the conclusion follows that conscious states are self-representing. As things stand, this is only an argument built on promises. Whether SR can actually deliver on those promises depends on how exactly it's cashed out, and to understand the different ways it might be cashed out we need to work our way down the family tree to the second and third choice-points.

Second Choice-Point: Do conscious states have distinguishable components corresponding to their lower- and higher-order content?

The self-representing states posited by SR are multifaceted – they have both higher-order and lower-order content. Self-Representationalists explain a subject's perceptual experience of the sunset in terms of a single state that represents both the sunset and itself. Structured theories propose that conscious states have an internal structure, and that within this structure one can isolate the components responsible for the lower-order and higher-order content of the state. So within the conscious state that constitutes the sunset-experience, one can isolate a component that represents the sunset and a component that represents that very state. Unstructured theories deny that conscious states have this internal structure. Although they hold that the subject's conscious state does indeed represent both the sunset and itself, they deny that this state is divisible into one component that represents the sunset and another component that represents the state. On this view, the conscious state is analogous to the sentence "this very sentence is about the sunset". The sentence represents a sunset and represents itself but it cannot be divided into a sentential representation of the sunset and a sentential representation of itself. Rather, it is an indivisible representation with

complex contents. Unstructured theories claim that conscious states are analogously indivisible.

Option 1: Unstructured Self-Representationalism

Some accounts of consciousness clearly qualify as Unstructured Self-Representationalist accounts. Brook & Raymont, for example, claim that conscious states are self-representing but explicitly deny that they have a composite structure (2006: 9). Williford (2006) and Block (2011) can also be put under this umbrella. A strong case can be made for reading Brentano this way too, although complex issues of interpretation abound (Zahavi 2004; 2005; 2006; Textor 2006; 2015; Kriegel 2016). However, there are a large set of views according to which: a) conscious states do not have a composite structure; b) those states confer awareness of themselves, but c) the awareness conferred is non-representational. Claims 'a' and 'b' certainly put such views in the neighbourhood of Unstructured SR, but claim 'c' looks like a deal-breaker: if you deny that conscious states represent themselves you can't qualify as a Self-Representationalist. One way of putting this position is that conscious states are self-*presenting* but not self-*representing*. This position is influentially articulated by Zahavi (2004; 2005; 2006) who denies that our awareness of our own conscious states is 'intentional'. In this he claims to have the support of just about every major phenomenologist besides Brentano, including Husserl, Scheler, Heidegger, Sartre, Merleau-Ponty, Henry and Ricoeur (2004: 82).[13] Going further back in history, Caston (2002) attributes this kind of view to

---

[13] Lyyra (2008) offers an interesting alternative interpretation of these phenomenologists according to which what is special about conscious states is not that we are aware of them,

Aristotle. Or more recently, Levine (2006; 2015) and Janzen (2006) offer accounts in line with this outlook.

One could take these self-presentational theorists at their word that they don't think conscious states represent themselves. However, I suggest that the disagreement might be dismissed as merely verbal. Terms like 'representation' and 'intentionality' are notoriously unclear, so it is an open possibility that when the proper sense of 'representation' is specified these theorists *do* think conscious states are self-representing. For instance, when Zahavi claims that conscious states aren't self-representing, his driving claim seems to be that our conscious state isn't an *object* for us in the way that perceived objects are.[14] This seems quite consistent with claiming that conscious states represent themselves, just not in the same way that they represent worldly objects. Although these interpretative issues are beyond the scope of the paper, it is worth noting that the Unstructured SR household plausibly extends beyond its card-carrying members.

What difficulties does Unstructured SR face? Rosenthal introduces a number of general objections to SR that seem to have particular bite against the unstructured form of the view. First, he suggests SR is unable to explain how an unconscious mental state becomes conscious. On HOR, a state becomes conscious when it enters into an extrinsic relation to a

---

but that we 'live through' them: a notion that certainly can't be understood in self-representational terms.

[14] Zahavi himself (2004: 74) seems to acknowledge there's a terminological issue here. Williford (2006: 3) the same issue and concludes that Zahavi is Self-Representationalist in every way that matters.

distinct mental state. But on SR, consciousness is intrinsic to a state so entering into different relations cannot explain how that state becomes conscious. Second, he suggests that SR is at odds with empirical evidence indicating a time-lag between the occurrence of a mental event and a subject's consciousness of that event. Libet's experiments (e.g. Libet 1985; Haggard & Libet 2001), found that subjects only became conscious of a volition a short time after the on-set of that volition. If the state of which we are conscious, and the state through which we are conscious of it, are one and the same then such a time-lag would be impossible. Third, he suggests that SR's claim that a conscious state is a single state with both lower and higher-order content cannot be reconciled with the fact that the lower- and higher-order content might involve different propositional attitudes. Consider a case in which we are conscious of our doubt. HOR explains this in terms of our having a lower-order state with an attitude of doubt toward some content, and a higher-order state with an assertoric attitude to the content that one is undergoing such a doubt. SR is forced into the potentially paradoxical position that we have a single state that has both a propositional attitude of doubt *and* an attitude of assertion.

Although these objections deserve to be taken seriously, the most important objection to Unstructured SR is that it fails to deliver on the Naturalisability Promise. Indeed, the majority of the theorists identified above as advocates of Unstructured SR have no commitment to naturalism. The problem for Unstructured SR is that naturalistic accounts of representation revolve around causation. Even if the details of natural representation are yet to be worked out, a plausible core principle is that a state represents an object in virtue of being suitably causally responsive to that object. However, causation is an irreflexive relation: nothing is a

cause of itself. This seems to preclude mental states from representing themselves.[15] Structured SR attempts to overcome this problem by claiming that conscious states have composite parts that stand in the requisite causal relations. However, by denying that conscious states have such a compositional structure Unstructured SR is unable to avail itself of this response.

Unstructured SR delivers on the Transitivity Promise by claiming that conscious states confer an awareness of that very state. And it delivers on the Intimacy Promise by claiming that a pain state represents both the bodily pain and itself. This seems to preclude any mismatch between the qualities you're aware of a conscious state as having and the qualities it really has. It also seems to preclude the possibility of being aware of oneself as being in a state that doesn't exist. However, by denying that conscious states have internal structure Unstructured SR makes it naturalistically inexplicable how conscious states could represent themselves. As such, it fails to deliver on the Naturalisability Promise.[16]

---

[15] Interestingly, Buras (2009) draws the opposite conclusion from this situation. He claims that some mental states represent themselves despite not being causes of themselves, and infers that causal theories of mental content must therefore be mistaken. However, for this move to overcome worries about naturalisability, we would need reason to believe that a non-causal naturalistic theory of representation is available. But the prospects of finding such a theory are poor.

[16] A further objection I have not considered is the threat of regress. Unstructured SR seems to be committed to all of the content of a conscious state being represented by that state. The problem is that representing all of that content is also part of the content of the state, so

Option 2: Structured Self-Representationalism

According to Structured SR, a conscious state is a complex state with both a lower-order and higher-order components that stand in some integration relation in virtue of which they constitute a single mental state. When a lower-order state M_1 such as a perceptual representation of a sunset is suitably related to a higher-order state M_2 that represents M_1, the two components form a single state M* that represents both the sunset and itself. Although M* cannot be a cause of itself as such, parts of M* can be causes of other parts of M*. As such, there is no obstacle to M_1 standing in the causal relation to M_2 required for natural representation. Structured SR thus shows far greater promise of delivering on the Naturalisability Promise than its unstructured counterpart.

Structured SR also promises to avoid some of the objections to SR broached by Rosenthal. Advocates of Structured SR can agree with Rosenthal that a state's becoming conscious must be explained in terms of it entering into new relations to other mental states. They only diverge from Rosenthal by proposing that these relations integrate the original state into a new complex state that is intrinsically conscious. Similarly, they can agree with Rosenthal that there might be a time-lag between a volition occurring and a subject becoming conscious of it, but diverge from Rosenthal by explaining this transition in terms of the volition coming to enter a complex self-representing state. Finally, they can accommodate Rosenthal's case of

---

must itself be represented. But then that content too must be represented and so on. This issue goes back at least to Brentano and is insightfully discussed by e.g. Williford (2006); Siewert (2013); Schear (2009) and Textor (2006).

the incompatible propositional attitudes by holding that the doubting attitude belongs to M_1 while the assertoric attitude depends on M_2. M_1 and M_2 constitute a single representation, but the constituent components of that representation can still exemplify different propositional attitudes. So it seems that SR can dodge Rosenthal's worries by attributing conscious states a compositional structure.

The foregoing provides some preliminary reasons to prefer Structured SR over Unstructured SR. However, to properly evaluate Structured SR we need to distinguish the different versions of the view. To that end, we should move down the family tree to our third and final choice-point.

The Third Choice-Point: Is the higher-order content of a conscious state non-conscious?

According to SR, by being in a suitably self-representing state M* we are conscious of that very state. According to Structured SR, this is to be explained in terms of a higher-order component of M* that represents the lower-order component of M* thereby making us conscious of that state's lower-order content. All forms of Structured SR thus agree that when we are in a conscious state we are aware of the lower-order content of that state. What they disagree on is whether we are ever also aware of the higher-order content of that state. Non-Reflexive Theories (as I will call them) claim that we are never aware of the higher-order content of our conscious states. Reflexive Theories claim that we are, at least sometimes, aware of the higher-order content of our conscious states.

Option 1: Non-Reflexive Self-Representationalism

According to HOR, in ordinary cases of consciousness we are aware of a first-order mental state but are not aware of the second-order state that bestows this awareness. In order to be

aware of our second-order state, we would need to have a third-order state that represents it. Such third-order representation is quite possible – HOR theorists claim that this is exactly what happens when we introspect – but even then the higher-order state in virtue of which we are conscious would not figure in our introspective awareness. Perhaps we could climb up to a fourth-order state to represent this, but it will always be the case that the *highest*-order state is outside of our awareness. Through our highest-order state, we can be aware of lower-order states but the highest-order state itself remains off the phenomenological stage.

Non-Reflexive Theories inherit the principle that the higher-order representation in virtue of which we are conscious inevitably falls outside of our awareness. One might think that the core tenets of SR preclude such an inheritance from HOR. According to SR, the state of which we are conscious and the state in virtue of which we are conscious are one and the same, so we cannot fail to be conscious of the state in virtue of which we are conscious. However, when we zoom in on the internal components of M^* posited by Structured SR we find a situation akin to that of HOR. M_1 is the component of M^* of which we are aware, and M_2 is the component in virtue of which we are aware. So even though M_2 is a constituent of M^* it remains off-stage: by being in M^* we are aware of the lower-order content of M^* but not its higher-order content. We might become aware of the higher-order content of a state when we introspect, but even then the *highest*-order component of our conscious state will inevitably be unconscious.

The foregoing captures what unifies the theories in the Non-Reflexive household. Now what differentiates them? Non-Reflexive theories differ mainly with respect to the account they give of what unifies a lower- and higher-order representation into a self-representing state.

Let's start with Gennaro's (2004; 2006) Wide Intrinsicality View (WIV), which he describes as follows:

> On the WIV, we have two parts of a single conscious state with one part
> directed at ('aware of') the other. In short, there is a complex conscious mental
> state with an inner, intrinsic relation between parts (2004: 60-1).

What makes the lower-order and higher-order components part of a single complex? Gennaro suggests that we have higher-order concepts that *synthesise* the content of the lower-order component, and thereby constitute a single self-representing complex. Moreover, he speculates that these complex representations are realised by neural feedback loops in the brain that display the kind of cross-level integration one might expect from such a complex.

Van Gulick's (2004; 2006) Higher-Order Global State (HOGS) theory holds that the relevant self-representing states are constituted by the integration of a lower-order state into a subject's wider mental economy. A subject is aware of a lower-order mental state insofar as the contents of that state are appropriately linked to that subject's other mental states. Van Gulick's theory displays something of the spirit of the global workspace theory, which equates a state's consciousness with its availability to higher mental processes. His proposal is that this mental network – which has the lower-order state as an integrated part – constitutes a single state that implicitly represents its own lower-order component. He explains:

> The transformation from unconscious to conscious state is not a matter of
> merely directing a separate and distinct meta-state onto the lower-order state
> but of "recruiting" it into the globally integrated state that is the momentary
> realization of the agent's shifting transient conscious awareness. (2004: 74-5)

Like Gennaro, Van Gulick has a story to tell about how such states are realised in the brain. He suggests that non-conscious states are realised in relatively localised areas of the brain, but when they become conscious they are integrated into a complex neural network that is globally distributed.

Both WIV and HOGS face serious difficulties that may well be endemic to the Non-Reflexive household. First, there is a worry that neither theory offers an account in which the lower- and higher-order components are genuinely unified into a single mental state.[17] When a lower-order state is 'synthesised' by higher-order concepts in the manner proposed by WIV, it's not clear why we shouldn't regard the resulting scenario as an interaction between two distinct states. Similarly, when a lower-order state is integrated with a wider mental network, it's not clear why we shouldn't regard it as an individual state that has entered into new relations to other distinct mental states. Appealing to the neurological processes that underwrite these states won't help: it is one thing for the neural processes that underwrite a pair of representations to be integrated, but quite another for the representations themselves to form a single representational state. Van Gulick and Gennaro might be tempted to resort to the claim that the integration relation is a kind of primitive property that joins mental representations. However, both are committed to the Naturalisability Promise, so cannot appeal to an unexplained integration relation (Kidd 2011).

There is also a worry that WIV and HOGS lose sight of the Transitivity Promise. A driving principle of Meta-Representationalism is that conscious states are represented, but doubts

---

[17] This objection is pushed by Weisberg (2008; 2011a) Kriegel (2009); Kidd (2011) and Picciuto (2011).

can be raised over whether either theory respects this principle. Is Gennaro's conceptual synthesis a species of representation? If not, then we no longer have a meta-representational theory and the transitivity of consciousness is left unaccounted for. If it is a kind of representation, it is unclear how such conceptual synthesis differs from the conceptual higher-order representations posited by higher-order thought theory. Does being integrated into a network of mental states constitute higher-order representation of the integrated state? Van Gulick has a teleosemantic story to tell in which the integrated state alters the dispositions of a subject's wider mental network, and the network thereby implicitly represents that state (2004; 2006). However, objectors have pointed out that such a dispositional account of implicit higher-order representation is at odds with awareness being an *occurent* and *explicit* phenomenon (e.g. Weisberg 2008).

More could be said about whether WIV and HOGS manage to successfully implement the Non-Reflexive approach. At the very least, the foregoing shows that implementing this approach is far from straightforward. However, the most dialectically salient objection to Non-Reflexive SR pertains not to how it is implemented but to its essential commitments. Remember, the characteristic feature of Non-Reflexive SR is the claim that the higher-order content of a conscious state is always unconscious. Kriegel suggests that any theory with this commitment will be unable to explain a crucial feature of our phenomenology that he describes as *inner awareness*. Inner awareness is the awareness we have *of our own awareness*:

> …I cannot envisage what it would be like to have a phenomenology lacking the
>
> kind of inner awareness that constitutes for-me-ness. Even the simplest visual
>
> experience – say, a homogenously bluish experience – folds within it an outer

awareness of blue and an inner awareness of that very outer awareness.

(Kriegel 2009: 175)

Here Kriegel is making the strong claim that consciousness is *always* characterised by inner awareness. However, even the weaker claim that consciousness is *sometimes* characterised by inner awareness is enough to cause trouble for Non-Reflexive SR. According to Non-Reflexive SR, a conscious state never represents its own higher-order component. Because awareness is to be understood representationally, we are therefore never aware of that higher-order component. But since that higher-order component is what constitutes our awareness at that time, we are therefore never aware of our concurrent awareness. So by making the higher-order content of a conscious state non-conscious, Non-Reflexive theories are committed to the impossibility of inner awareness. If Kriegel is right that there are at least some cases in which our awareness figures in our phenomenology, then Non-Reflexive SR is false.

Can Non-Reflexive theorists simply reject Kriegel's phenomenological claim? Many have stated that reflection on their own phenomenology fails to reveal anything akin to Kriegel's inner awareness. Others, however, have agreed whole-heartedly with Kriegel's phenomenological report.[18] Attempts have been made by each side of this dispute to explain away the phenomenological reports of their opponents, but none of these attempts stand up to scrutiny (McClelland 2015). A more promising route is to cite theoretical considerations in favour of the thesis that we are at least sometimes aware of awareness. However, the

---

[18] Detractors include Gennaro (2008); Schear (2009); Mehta (2013); Gertler (2012); Coleman (2015). Supporters include Zahavi (2004; 2005; 2006) and Strawson (2011).

theoretical assumptions that drive these arguments are no less contentious than the putative phenomenological datum itself, so this strategy too has proved inconclusive (McClelland 2015).

One argument for inner awareness that deserves special consideration is Kriegel's 'epistemic argument' (2009: 115-129). In line with the transitivity principle, everyone in the Meta-Representational family agrees that all conscious states are represented. But *how* do we know this about conscious states? Kriegel suggests that we know this precisely because it is manifest to us in our experience: when we are in a conscious state, we are aware of that very state being represented. He goes on to claim that this direct phenomenal justification is the *only* justification we could have for the transitivity principle. There are four other potential sources of evidence for the principle - indirect phenomenological evidence (including evidence from introspection), a posteriori experimental evidence, a priori conceptual analysis and general philosophical principles – but Kriegel argues that none of them can provide what is needed. If Kriegel is right about this, then it would be extremely difficult for Meta-Representationalists to deny that consciousness is characterised by inner awareness. Meta-Representationalists are motivated by the transitivity principle, so by denying that consciousness is characterised by an inner awareness that directly justifies the principle, they inadvertently undermine their own position. Critics have responded to Kriegel by saying that he underestimates introspection as a source of evidence (Levine 2010; Van Gulick 2011). This gets us into difficult issues about the warrant of generalisations based on introspective judgement, and it again seems that the theoretical assumptions of the disputants are no less contentious than the phenomenological datum they are contesting. Kriegel's epistemic argument thus captures one reason we might have for concluding that consciousness is characterised by inner awareness, but it does not settle the matter entirely.

Given the recalcitrance of this phenomenological disagreement, it would be premature to say that Non-Reflexive SR faces a fatal objection here. Insofar as you find it plausible that your concurrent awareness sometimes figures in your phenomenology, you should reject Non-Reflexive SR. If, on the other hand, you find this phenomenological claim implausible then Non-Reflexive SR will survive Kriegel's objection. However, we will see shortly that a problem with intimacy faced by all Structured SR theories might undermine the Non-Reflexive household. In the meantime, we should explore whether the other branch of this choice-point better enables SR to accommodate inner awareness.

Option 2: Reflexive Self-Representationalism

Reflexive SR claims that the higher-order content of a conscious state can itself be conscious. Reflexive SR is driven by the following promise:

> **The Inner Awareness Promise:** We promise to accommodate the fact that
> inner awareness characterises at least some phenomenological episodes.

Reflexive SR thus accumulates a fourth promise on top of the three already accrued: the Transitivity Promise; the Naturalisability Promise and the Intimacy Promise. An influential Reflexive Self-Representational theory has been developed in detail by Kriegel. According to Kriegel's Cross-Order Integration (COI) theory, a conscious self-representing state is formed by the integration of a lower- and higher-order state. Kriegel proposes that this integration is achieved in the same way that different perceptual contents are integrated in perceptual binding. The early visual system has discrete representations of, say, an object's red colour and its circular shape. These signals must then be bound in order to form a representation of a red circle. Our best understanding of how these neural signals are bound is that they

synchronise the frequency of their oscillations. Kriegel suggests that neural synchrony is also what binds a lower- and higher-order representation into a single self-representing state.

How does Kriegel hope to accommodate inner awareness on this model? The challenge for Kriegel is to offer an account that is naturalisable. We have already seen that on a naturalistic understanding of representation it is likely to be an irreflexive relation. Although this was a problem for Unstructured SR, Structured SR gets around this problem by proposing that a conscious state has components M_1 and M_2 that are causally related. Although this might explain how M_2 represents M_1, it can't straightforwardly explain how M_2 could represent M_2 as this component of the conscious state cannot be a cause of itself. Kriegel's strategy is to appeal to indirect representation. He notes that a painting may represent a whole house in virtue of representing part of that house, and that a perceptual state may represent a whole apple in virtue of representing the facing surface of that apple. In other words, by directly representing an integrated part of a thing, one can thereby indirectly represent the whole of which it is a part.[19] According to COI, M_1 is an integrated part of M*. Since M_2 *directly* represents M_1, it thereby *indirectly* represents M*. Since M_2 is part of M*, there is thus a sense in which M* represents its own higher-order component. By representing its own higher-order component, M* thus makes one aware of one's awareness. COI thus entails that

---

[19] Here parthood has to be understood as a substantive relation. One cannot simply stipulate that there is a complex entity Z consisting of X and Y. X and Y are only parts of a single whole if there is a contingent relation between X and Y that is necessary for the existence of Z but not for the existence of X or Y (Kriegel 2009).

we can be aware of our awareness. In fact, it yields the stronger conclusion that *whenever* we are aware we are aware of our awareness.

Kriegel's proposal faces a number of objections. The first objection is that the binding relation cited by COI does not bind distinct representations into a single representational state. In vision, the neural synchrony of a red-signal and a circle-signal binds them to a single object viz. the red circle. But binding of contents to a single object is not equivalent to the joining of representations into a single representation.[20] Like WIV and HOGS, worries can thus be raised about whether COI really yields the result that consciousness is underwritten by a single self-representing state rather than a pair of distinct states as in HOR.

COI's appeal to neural synchrony raises further problems (see e.g. Sebastian 2012 and Coleman 2015). However, a deeper problem for Kriegel is that even if we grant his account of what binds the components of a conscious state, COI is still unable to deliver on the Inner Awareness Promise. Kriegel relies on a number of claims about indirect representation that can be called into doubt.

First, some doubt that there is even such a phenomenon as indirect representation. When we represent the surface of an apple we also represent the whole apple, but why can't this be understood in terms of a direct representation of the surface bringing about a *distinct* direct representation of the whole apple? The notion of indirect representation may not be needed at all (Levine 2010; Phillips 2013).

---

[20] See Van Gulick (2012); Levine (2010); Kidd (2011); Weisberg (2008; 2011a) and Rosenthal (unpublished).

Second, even if we grant that there is such a thing as indirect representation, it is doubtful that direct representation of a part suffices for indirect representation of the whole. Phillips (2013) offers an example of seeing a lemon that, unbeknownst to you, is an integral part of a lemon battery. It is implausible that by seeing the lemon you thereby see the lemon battery. But this raises doubts about why representing M_1 should suffice for representing M^*.[21] Perhaps Kriegel can respond by claiming that if there are further conditions that must be met for indirect representation, they are met in the case of M^*. But even if that is granted, a third problem remains.

To accommodate inner awareness our representation of M^* must be *phenomenal* representation: Kriegel's promise is, after all, to explain our *awareness* of our own awareness. However, indirect representation is most plausibly *non*-phenomenal. When you phenomenally represent the surface of an apple, you might thereby represent the whole apple but you do not thereby represent the whole apple phenomenally (Levine 2010; Sebastian 2012). Your visual experience seems to be exhausted by the appearance of the apple's surface, so even if there is a sense in which the whole apple is indirectly represented, it is not a sense that has any bearing on your phenomenology. So even if M^* does indirectly represent itself, the phenomenology of M^* ought to be exhausted by what it directly represents i.e. by its represented lower-order content. Overall then, COI is unable to convincingly deliver on the Inner Awareness Promise.

---

[21] It is also worth noting that being part of a complex isn't plausibly a *necessary* condition of indirect representation either. Coleman (2015) and Weisberg (2011) each provide cases in which we represent an object via something other than an integrated part of that object.

I noted earlier that the existence of inner awareness is a contentious phenomenological datum. Might Structured SR fare better if it denied that datum and scrapped the Inner Awareness Promise? I suggest not. A promise with which no version of SR can dispense is the Intimacy Promise, but there are reasons to think that Structured SR is unable even to deliver on that promise. Structured SR sidesteps the naturalisability worries faced by Unstructured SR by proposing that the higher- and lower-order content of a conscious state are underwritten by distinct components. However, as soon as this move is made the possibility emerges that the higher-order component misrepresents the lower-order component. And if such misrepresentation is possible, Structured SR can ensure neither qualitative intimacy nor existential intimacy.

Let us consider qualitative intimacy first. If the lower-order component of the self-representing state is an itch, there is nothing to prevent the higher-order component from representing it as a pain. This is evident in each of the Structured SR theories we've examined. On WIV nothing precludes the higher-order component from conceptualising a representation of an itch as a representation of a pain. On HOGS, the global network of mental states that are responsive to the lower-order state might *misconsume* that state as if it were a pain. On COI, by Kriegel's own admission a lower-order representation of an itch might be neurally synchronised with a higher-order state that represents it as a pain. The Structured SR theorist might respond by insisting that on a proper understanding of the state-integrating relation, one would see that misrepresentation is impossible. However, the relations actually sketched by the three theories discussed do nothing to guarantee accuracy, and it is difficult to see how any naturalisable relation *could* yield such infallibility.

Perhaps Structured SR can respond to this problem by claiming that there is a constitutive link between the qualities that a conscious state *represents* itself to have and the qualities that it actually has. Indeed, Kriegel makes this move explicitly and claims this COI thereby accommodates qualitative intimacy (2009: 110)*.* Gennaro could make a parallel move by saying that our phenomenology is fixed by how the higher-order component of a state represents itself as being. Van Gulick would have to say that what fixes our phenomenology is how your wider mental economy takes your lower-order state to be. The problem with such an account of qualitative intimacy is that it is indistinguishable from the account offered by HOR theorists: an account on which the qualities of the lower-order state drop out as irrelevant. We saw reasons to regard that account as unsatisfactory, but even it was to be deemed satisfactory this would justify the adoption of HOR theory, and render the digression into SR entirely pointless.

A further problem with this constitutive response is that it compromises Structured SR's ability to accommodate existential intimacy. It seems that the higher-order component of a self-representing state can exist in the absence of any lower-order component.[22] Advocates of Structured SR say that a state is conscious only if it is a composite that has an actual lower-order component, but are they entitled to this claim? Remember, on the constitutive view

---

[22] On WIV, the lone higher-order state would be a conceptual representation to the effect that one is in some lower-order state. For HOGS, the lone higher-order state would be some global state of one's mental network such that the network is disposed to take it to be the case that one is in some lower-order state. For COI, self-representation involves a higher-order state being bound through neural synchrony to a lower-order state, so a lone higher-order state would be whatever that state is prior to binding.

our phenomenology is fixed entirely by how the higher-order component of the complex represents things as being: if it represents us as being in a lower-order state of pain, then we thereby have a pain experience. But if our pain experience is constituted by the higher-order component of our conscious state, why would that component need to be integrated with a lower-order state before constituting a pain experience? Remember, the lower-order state is explanatorily idle on the constitutive approach. The Structured SR theorist would have to say that an unintegrated state that represents us as being in a pain state would not yield a pain experience, but that if this state was then integrated with a lower-order representation of an itch, we would thereby have a pain experience. This seems at best *ad hoc* and at worst absurd. Advocates of Structured SR might respond by just *stipulating* that consciousness requires the higher-order component to be integrated with some lower-order component. But if such a stipulation is permissible, the HOR theorist can simply stipulate that higher-order states only yield consciousness when they target an actual lower-order state.[23] Again, Structured SR fails to yield any advantage over the HOR branch of the family.

---

[23] Kriegel hopes to have an advantage over his competitors here because being bound to a lower-order state is a necessary condition of the indirect self-representation constitutive of conscious, so the higher-order component of a conscious state can't fail to have a target. However, Weisberg (2008) points out that nothing in Kriegel's account precludes a higher-order state from being bound with some lower-order state other than the one it represents. So even if the theory entails that it has to be bound to some real state, it doesn't entail that this real state must be its target. Kriegel might respond that if the higher-order component referred to a non-existent state, it would fail to directly refer to any part of the conscious state of which it is a component, and so fail to achieve the indirect self-representation

Moving Forward

Having worked our way down the Meta-Representationalist family tree, we can now see that no existing version of SR is adequate. Along the way we have accrued three promises on which any adequate version of SR must deliver:

1) The Transitivity Promise

2) The Naturalisability Promise

3) The Intimacy Promise

HOR theory delivered on 1 and 2 but not 3. Unstructured SR delivers on 1 and 3 but not 2. Structured SR appeals to a natural representation relation between the components of a conscious state, so promises to deliver on 1 and 2. However, we have just seen that it is ill-equipped to deliver on 3. Perhaps the specific character of the relation that unifies the two components into a single state can put some distance between Structured SR and HOR and allow it to deliver on the Intimacy Promise. However, by moving too far away from an ordinary representation relation Structured SR risks violating 1 by positing a relation that doesn't amount to representation, or violating 2 by positing a representation relation that cannot be accounted for naturalistically.

We also discussed a fourth promise: the Inner Awareness Promise. The putative phenomenological datum that drives this promise is contentious, so a case could be made for not factoring it into our assessment. It is worth noting however that neither HOR nor

---

necessary to consciousness. However, Weisberg (2008) proposes that the requisite indirect self-representation can be achieved through reference to a non-existent target state, meaning that COI is vulnerable to the possibility of targetless conscious states after all.

Structured SR show any real potential to accommodate inner awareness. Unstructured SR fares better: there's no reason why an unstructured conscious state's *sui generis* representation of itself couldn't yield inner awareness. However, it remains the case that this kind of self-representation is unnaturalisable.

How should we deal with the fact that no existing theory can deliver on the three main promises? The pessimistic strategy would be to concede that Meta-Representationalism is unable to yield all three results, and then to a pick a theory that delivers on the promises that we value most. We might dispute the apparent intimacy of consciousness and endorse HOR, or we might scrap any aspiration to naturalisability and endorse Unstructured SR. Different versions of SR would deliver on different combinations of promise, but none will deliver on all. This pessimistic strategy shouldn't be dismissed out of hand. It is, after all, an open possibility that the assumptions that drive the three main promises are in error. That said, I would recommend a more optimistic approach.

There is nothing formally inconsistent about the three main promises, so we shouldn't shy away from the ambition of delivering on every one of them. Structured SR fell down on the relations it posited between the higher- and lower-order components of a conscious state. To the extent that they posit an ordinary representation relation, they fail to gain any advantage over HOR. But to the extent that they posit a more distinctive close connection between the two components, they move away from giving an account of awareness that is both meta-representational and naturalisable. This suggests that what we need to do is to find a relation that, like ordinary representation, is representational and naturalisable but, unlike ordinary representation, is able to accommodate the intimacy of consciousness.

To do this, I suggest that we need to reconceive the higher-order component of a conscious state as *disclosing* the lower-order component of the state as it is in itself rather than as attributing the lower-order component properties in a way that might misfire. To account for such a relation in natural representational terms, one might follow Kidd (2011) in appealing to *indexicals*. Indexical reference is normally taken to have a kind of context-relative infallibility, and this might help account for the intimacy of consciousness. Furthermore, there is little threat of indexical reference being unnaturalisable. Alternatively, one might follow Coleman (2015) or Picciuto (2011) in appealing to *quotation*. When we quote something that very thing stands for itself in a way that precludes at least some kinds of misrepresentation. And again, quotation seems open to naturalisation.[24]

Such a self-representational account would face many challenges. It would have to explain why indexical or quotational reference to a state would yield consciousness of that state. It would also have to reassure skeptics that such a relation can indeed be naturalised. And since such an account shows little hope of delivering on the Inner Awareness Promise, it would have to undermine the claim that our phenomenology is ever characterised by inner awareness. I can see no immediate reason to doubt that such challenges can be met. Meta-Representationalism is an attractive family, and the Self-Representationalist branch of that family holds the most appeal. The members of the family to which we have been introduced

---

[24] Appeals to acquaintance (Williford 2006; Hellie 2007), or to a direct awareness relation akin to that posited by naïve realists (Kidd 2015), might also fit the bill. However, in these cases one might worry: a) that the relation is no longer representational and; b) that the relation has poor prospects of naturalisation. Again though, it would be premature to rule these options out.

might not be what we wanted, but there remains reason to hope that we might find another member of the family that provides exactly what we need.

References

Armstrong, D. M. (1981), 'What is Consciousness', from his *The Nature of Mind* (University of Queensland Press).

Block, N. (2011), 'The Higher Order Approach to Consciousness is Defunct', in *Analysis* 71/3: 419-31.

Brogaard, B. (2012), 'Are Conscious States Conscious in Virtue of Representing Themselves?', in *Philosophical Studies* 159: 467-74.

Brook, A. and Raymont, P. (2006), 'The Representational Base of Consciousness', in *Psyche* 12/2.

Buras, T. (2009), 'An Argument Against Causal Theories of Mental Content', in *American Philosophical Quarterly* 46/2: 117-29.

Carruthers, P. (2004), 'HOP over FOR, HOT Theory', in R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness* (John Benjamins), 115-35.

Caston, V. (2002), 'Aristotle on Consciousness', in *Mind* 111/444: 751-815.

Coleman, S. (2015), 'Quotational Higher Order Thought Theory', in *Philosophical Studies* 172/10: 2705-33.

Dretske, F. (1995), *Naturalizing the Mind* (MIT Press).

Ford, J. (2009), 'Review of Self-Representational Approaches to Consciousness', in *Minds & Machines* 19: 283-7.

Gennaro, R. J. (2004), 'Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine', in R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness* (John Benjamins), 45-56.

Gennaro, R. J. (2006), 'Between Pure Self-Referentialism and the (extrinsic) HOT theory of consciousness', in U. Kriegel and K. Williford (ed.), *Consciousness and Self-Reference* (MIT Press), 221-48.

Gennaro, R. J. (2008), 'Representationalism, peripheral awareness, and the transparency of experience', in *Philosophical Studies* 139: 39-56.

Gertler, B. (2012), 'Conscious states as objects of awareness: on Uriah Kriegel, Subjective consciousness: a self-representational theory', in *Philosophical Studies* 159: 447-55.

Haggard, P. and Libet, B. (2001), 'Conscious intention and brain activity', in *Journal of Consciousness Studies* 8/11: 47-64.

Hellie, B. (2007), 'Higher-Order Intentionality and Higher-Order Acquaintance', in *Philosophical Studies* 134: 289-324.

Hossack, K. (2002), 'Self-Knowledge and Consciousness', in *Proceedings of the Aristotelian Society* 102: 163-81.

Janzen, G. (2006), 'Phenomenal Character as Implicit Self-Awareness', in *Journal of Consciousness Studies* 13/12: 44-73.

Kidd, C. (2011), 'Phenomenal Consciousness with Infallible Self-Representation', *Philosophical Studies* 152: 361-83.

Kidd, C. (2015), 'The Idols of Inner Sense', *Philosophical Studies* 172: 1759-82.

Kriegel, U. (2009), *Subjective Consciousness: A Self-Representational Theory* (OUP).

Kriegel, U. (2016), 'Brentano's Dual-Framing Theory of Consciousness', in *Philosophy and Phenomenological Research* doi:10.1111/phpr.12327.

Kriegel, U. and Williford, K. (2006), 'Introduction to Self-Representational Approaches to Consciousness', in U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness* (MIT Press), 1-8.

Levine, J. (2003), 'Experience and Representation', in Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives* (OUP), 58-76.

Levine, J. (2006), 'Conscious Awareness and (Self-)Representation', in U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness* (MIT Press), 173-98.

Levine, J. (2010), 'Review of Subjective Consciousness: A Self-Representational Theory', in *Notre Dame Philosophical Review*.

Levine, J. (2015), 'A "Quasi-Sartrean" Theory of Subjective Awareness', in S. Miguens, G Preyer and C. B. Morando (eds.), *Pre-Reflective Consciousness: Sartre and Contemporary Philosophy of Mind* (Taylor & Francis), 242-62.

Libet, B. (1985), 'Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action', in *The Behavioral and Brain Sciences* 8: 529-66.

Lycan, W. G. (2001), 'A simple argument for a higher-order representation theory of consciousness', in *Analysis* 61/269: 3-4.

Lycan, W. G. (2004), 'The Superiority of HOP to HOT', R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness* (John Benjamins), 93-113.

Lyyra, P. (2008), 'Two Senses for "givenness of consciousness"', in *Phenomenology and Cognitive Science* 8: 67-87.

McClelland, T. (2015), 'Affording Introspection: An Alternative Model of Inner Awareness', in *Philosophical Studies* 172/9: 2469-92.

McClelland, T. (forthcoming), 'Do We Experience Our Conscious States as Belonging to Us?' in M. Guillot and M. García-Carpintero (eds.) *Consciousness and the Sense of Mineness* (OUP).

Mehta, N. (2013), 'Is there a phenomenological argument for higher-order representationalism?' in *Philosophical Studies* 164: 357-70.

Neander, K. (1998), 'The Division of Phenomenal Labour: A Problem for Representational Theories of Consciousness', in *Noûs* 32/12: 411-34.

Phillips, B. (2013), 'Indirect representation and the self-representational theory of consciousness', in *Philosophical Studies* doi:10.1007/s11098-012-0087-1

Picciuto, V. (2011), 'Addressing Higher-Order Misrepresentation with Quotational Thought', in *Journal of Consciousness Studies* 18/3-4), 109-36.

Rosenthal, D. (1986), 'Two concepts of consciousness', in *Philosophical Studies* 49: 329-59.

Rosenthal, D. (2011), 'Exaggerated reports: reply to Block' in *Analysis* 71/3: 431-437.

Rosenthal, D. (unpublished), 'Intrinsicalism and its Discontents' https://sites.google.com/site/davidrosenthal12/ last accessed 07/10/16.

Schear, J. K. (2009), 'Experience and self-consciousness', in *Philosophical Studies* 144: 95-105.

Seager, W. (2006), 'Is Self-Representation Necessary for Consciousness?', in *Psyche*, 12/2.

Sebastian, M. A. (2012), 'Review of Uriah Kriegel, Subjective Consciousness', in *Disputatio* 32: 413-7.

Siewert, C. (2013), 'Phenomenality and Self-Consciousness', in U. Kriegel (ed.), *Phenomenal Intentionality* (OUP), 235-60.

Strawson, G. (2011), 'Radical Self-Awareness', in M. Sidertis, E. Thompson and D. Zahavi (eds.), *Self, No Self?: Perspectives from Analytical, Phenomenological and Indian Traditions* (OUP), 274–307.

Textor, M. (2006), 'Brentano (and Some Neo-Brentanians) on Inner Consciousness', in *Dialectica*, 60/4: 411-32.

Thomasson, A. L. (2006), 'Self-Awareness and Self-Knowledge', in *Psyche*, 12/2.

Van Gulick, R. (2004), 'Higher Order Global States (HOGS): An Alternative Higher-Order Model of Consciousness', in R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness* (John Benjamins), 67-92.

Van Gulick, R. (2006), 'Mirror, mirror - is that all?', in U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness* (MIT Press), 11-40.

Van Gulick, R. (2012), 'Subjective consciousness and self-representation' in *Philosophical Studies* 159: 457.

Weisberg, J. (2008), 'Same old, same old: the same-order representation theory of consciousness and the division of phenomenal labour', in *Synthese* 160: 161-81.

Weisberg, J. (2011b), 'Abusing the notion of what-it's-likeness: A response to Block', in *Analysis* 71/3: 438-43.

Weisberg, J. (2011a), 'Review of Subjective Consciousness: A Self-Representational Theory',
in *Mind* 120/478: 538-542.

Williford, K. (2006), 'Zahavi versus Brentano: A Rejoinder', in *Psyche* 12/2.

Zahavi, D. (2004), 'Back to Brentano', in *Journal of Consciousness Studies* 11/10-11: 66-87.

Zahavi, D. (2005), 'Intentionality and Experience', *Synthesis Philosophica* 40: 299-318.

Zahavi, D. (2006), 'Two Takes on a One-Level Account of Consciousness', *Psyche* 12/2.