

Special Attention to the Self: A Mechanistic Model of Patient RB's Lost Feeling of Ownership¹

Hunter R. Gentry

hgentry@wisc.edu

University of Wisconsin-Madison

[PREPRINT] Forthcoming in *Review of Philosophy & Psychology*; accepted 07/21/2021

Abstract: Patient RB has a peculiar memory impairment wherein he experiences his memories in rich contextual detail, but claims to not own them. His memories do not feel as if they happened to him. In this paper, I provide an explanatory model of RB's phenomenology, the self-attentional model. I draw upon recent work in neuroscience on self-attentional processing and global workspace models of conscious recollection to show that RB has a self-attentional deficit that inhibits self-bias processes in broadcasting the contents of episodic memories to the global workspace. Typically, self-related contents enjoy a higher salience level than other-related contents. Elimination of bias toward self-related contents diminishes the salience of those contents to the level of other-related contents. Because the typical high salience of self-related content is necessary for the feeling of ownership, RB lacks the feeling of ownership. I also discuss potential applications of the self-attentional model to other psychopathological cases.

0. Introduction

Philosophers and cognitive scientists have recently been engaged in a debate about the function and nature of episodic memory (Allen & Fortin 2013; De Brigard 2014; Klein 2016; Michaelian 2011a; Robins 2017; Tulving 1972, 2002). Tulving (1972) originally distinguished between the two memory systems and this distinction was grounded in the nature of the stored content.² Episodic content is characterized by autobiographical detail, temporally dated episodes, and the temporal-spatial relations between the episodes. Semantic content, on the other hand, is necessary for language use, characterized by verbal symbols, meaning, and reference. Episodic memory is further distinguished from semantic memory by its phenomenology (Tulving 1983, 1985, 1993, 2002). Autonoetic consciousness, the feeling of mentally travelling back in time, is thought to be *essential* to episodic memory. Whereas semantic memory lacks autonoetic consciousness. Tulving (1993), commenting on the famous amnesiac patient KC, says that “for a rememberer to [episodically] remember something means that he or she is autonoetically aware of a past happening in which he or she has participated” (pp. 69-70).³ Importantly,

¹ Special thank yous go to Cameron Buckner, Hayley Clatterbuck, Larry Shapiro, Farid Masrour, and Elizabeth Schechter.

² Tulving notes other dimensions where the two systems are supposed to differ, but content will suffice for present purposes. For a review see Tulving (2002).

³ See also Tulving (1985).

autonoetic consciousness is not intrinsic to the content of an episodic memory. For Tulving, auto-noesis is the way that episodic content is experienced. However, episodic memory is not merely a particular kind of mental experience, nor is it merely a particular kind of content. Rather, it is a complex state comprised of both (Tulving 2002, 2005).

Stanley Klein (2016a) argues against Tulving's view. For Klein, episodic and semantic memory are not distinguished by content; indeed, memory content is stored system-neutrally. What makes a memory episodic is whether auto-noesis binds to content at retrieval. Klein uses the case of patient RB to illustrate that he lacks some, "[m]echanisms that enables auto-noetic consciousness to bond with retrieved content, making possible recollective experience" (Klein 2016a, pg. 391). Klein thinks this allows him to establish that episodic memory is a kind of mental experience-- that is, "it is the act of conjoining of retrieved content with a particular mode of temporal subjectivity that makes experienced content 'memory content'" (pg.387). Thus, the distinction between episodic and semantic memory consists in the mode of experience that is determined by whether auto-noetic consciousness attaches to neutral content upon retrieval. RB reports that his memories do not feel like his own-- that they lack the feeling of "mineness". For Klein, the feeling of ownership over one's memories is entailed by auto-noetic consciousness by being a constituent part (see Klein 2016b, pg. 490). So, if RB lacks the feeling of ownership, then he lacks auto-noetic consciousness. In agreement with Tulving, Klein thinks that auto-noesis is essential to episodic memory. Contrary to Tulving, Klein thinks that episodic memory is just a kind of mental experience, and is not distinguished from semantic memory by its content.

As far as I know, no mechanistic explanation for RB's condition has been given in the literature. That is, the "binding mechanism" that Klein gestures at, has not been identified. In this paper, I hope to make progress toward identifying this mechanism. While I will not be identifying the binding mechanism for *auto-noesis*, I will endeavor to mechanistically explain the feeling of ownership. 'Auto-noesis' is at best an umbrella term for many different kinds of abilities and phenomenal experiences. At worst, it is a conceptually muddled category. I intend to explain the lost feeling of ownership without taking a stance on what auto-noesis is or what it entails.

I draw upon recent work in neuroscience on self-attentional processing (Humphreys and Sui 2016) and global workspace models of conscious recollection (De Brigard 2012) to show that RB has a self-attentional deficit that inhibits self-bias processing in broadcasting the contents of episodic memories to the global workspace. It is important to note at the start that the self-attentional model I provide is tentative. The only data on RB that we have are excerpts from an interview with Klein. There is no neuropsychological profile of RB, so my tentative explanation is based almost entirely on these excerpts. Indeed, RB has been pretty widely discussed in philosophy journals as if he is a case like KC or HM. The reality is that RB is not like these more well known cases precisely because there is no neuropsychological profile. However, the goal is to provide an explanatory model that can be empirically tested, opening an avenue for future research. Moreover, based on the explanation I provide, I argue for a necessary condition on the feeling of ownership in episodic recollection.⁴ If successful, the self-attentional model of ownership might go some way toward identifying the neural correlates of auto-noesis, as well as having wider applicability to other kinds of self-disturbances.⁵

The strategy for the paper is as follows. In section 1, I describe the case of RB and how it is supposed to establish Klein's position. In section 2, I review evidence for the necessity of attention in conscious episodic recollection. In section 3, I review evidence for a separate self-attentional network (SAN) in self-bias processing. In section 4, I review additional evidence for the self-attentional network's involvement in episodic retrieval processes. In section 5, I tie together the evidence in sections 2, 3, and 4 for a mechanistic explanation of RB's condition. Additionally, I argue for a necessary condition on the feeling of ownership based on the self-attentional model. Finally, in section 6, I discuss predictions that the self-attentional model makes about RB and gesture at potential applications of the model to depression and schizophrenia.

1. The Case of RB

⁴ This condition, to be elaborated on later, is a certain level of salience of self-related content. By "necessary", I mean that the feeling of ownership cannot come about without such salience. If one thinks that the feeling of ownership supervenes on content, then the salience of self-related content is part of the supervenience base.

⁵ I say "might go some way toward identifying the neural correlates of auto-noesis", because I wish to remain neutral about whether ownership is a constituent of auto-noesis.

In this section, I will introduce the reader to patient RB, show how RB is supposed to establish Klein's position, and then briefly discuss how I am thinking about the feeling of ownership.

RB was in a traumatic accident that initially resulted in retrograde and anterograde amnesia. RB recovered from the amnesias, but reported that his memories did not feel like they were his own-- as if these memories were not things he originally experienced. Interestingly, his reports suggest that this is the only disruption in his recollective experiences. He reports vivid scene construction, self-referential details, and continuity between episodes. Consider a few notable quotes from RB:

I remember eating pizza at XXX in Isla Vista about a month before [his accident], but the memory *belongs to someone else*. But knowing I like pizza in the present . . . now . . . is owned by me . . . when I recall memories from my past I intellectually know they are about me. It just does not feel like it . . . when I remember scenes from before [the accident] *they do not feel as if they happened to me*—though intellectually I know they did.

I can picture the scene perfectly clearly ... studying with my friends in our study lounge. I can 'relive' it in the sense of *re-running the experience* of being there. But it has the feeling of imagining, [as if] re-running an experience that my parents described from their college days. It did not feel like it was something that really had been a part of my life. Intellectually I suppose I never doubted that it was a part of my life. Perhaps because there was such continuity of memories that fit a pattern that lead up to the present time. But that in itself did not help change the feeling of *ownership* (Klein and Nichols 2012, pg.686; emphasis added).

It seems clear that RB's semantic memory is intact because, for example, he knows that he likes pizza. Indeed, it's likely that RB uses his semantic memory to make inferences about what actually happened to him despite it not feeling that way. And as RB states in the second quoted passage, the continuity of the remembered events seem to play a role in his judging that these events happened to him.⁶ Moreover, the issue does not seem to be a lack of content, but rather how that content is presented to consciousness. RB seems to be a participant in his memories-- there is self-referential content-- but they are not experienced as such.

Another interesting aspect of RB's pathology is that the lack of feeling of ownership seems only to be for memories prior to the accident (premorbid). Consider this exchange between RB and Klein:

⁶ In fact, RB apparently now claims ownership over all of his memories (pre- and post-morbid). He does not say that he *feels* ownership over all of his memories, however. So his coming to judge that he owns all of his memories might very well be explained by inferences made based on his semantic memory.

S.B.K.: You say you own your memories from after the accident. What do you mean by “owning” your memories?

R.B.: They feel like things that happened to me. I can see them (in my head). ... They feel like part of my life ... my past ... things that happened to me. I was there. I guess that’s what I mean by owning them(memory). They’re mine, my past ... not something I just know about. ... I was there ... lived through it.

S.B.K.: And memories from before the accident—you don’t feel like you own them?

R.B.: Right. I see them play in my head. Like watching a movie. I watch them, they fit (with what I know about my past), but I don’t own them. ... I don’t feel like I was there (when the events happened). I know they must be mine, must have happened to me, but I feel like I’m watching a movie about me that I didn’t make (Parentheses added for clarification—based on discussions with R.B.) (2019, pg.492).

Here, RB suggests that the lack of feeling of ownership is not associated with memories formed post-morbid, only premorbid. More curious still is that RB reports that he was able to plan for the future, albeit with some difficulty:

I was able to plan for the future. Although my working memory impairment. . . made it challenging. When I slowly returned to work, it was hard to plan a complex strategy. I had to think of useful things to do and then do them. The best compensation I found was to separate the planning of the strategy from the execution. It worked best if I made a list of “Things To Do” (Klein 2016a, pg. 393).

Here, and elsewhere, it seems that RB has *not* lost the ability to self-represent. Indeed, he explicitly states that he does not have an identity problem (Klein and Nichols 2012, pg.687). However, it is an open question whether the lack of feeling ownership extends to other self-related states besides memory, e.g., dreaming, perception, imagination.⁷ Indeed, the difficulty with planning for his personal future might suggest that there are disruptions beyond memory.⁸

In the introduction, I said that Klein uses this case of RB to argue that memory content is stored system-neutrally and becomes episodic only if auto-noesis binds to it at retrieval. Because the feeling of ownership is a constituent of auto-noesis, the lack of the feeling of ownership, is a lack of auto-noesis. Hence, RB’s memories are not episodic. I will not be engaging with Klein’s view, however. I intend only

⁷ Personal correspondence with Klein.

⁸ I discuss this at length in section 6.1.

to supply a mechanistic explanation of RB's lack of feeling of ownership. Which, if successful, makes progress on identifying the autooiesis binding mechanism that Klein thinks is malfunctioning in RB.

So what is the feeling of ownership?⁹ It will help to say what it is not. RB is aware of the fact that he is instantiating a memory-- that he is in a particular mental state. This is minimal ownership. Minimal ownership, following Bradley (2019), is mere awareness that one is in a particular mental state. This kind of awareness arises out of the *de se* content of that mental state. But there is a more interesting sense of ownership that RB lacks-- the feeling of "mineness"-- or as William James (1890) put it, the feeling of warmth and intimacy. RB does not feel as if his memories happened to him. Jordi Fernandez (2019) analyzes the feeling of ownership, with respect to a memory, as the feeling that the memory matches the past. Without getting into the details of Fernandez's view, intuitively this is the feeling that RB lacks. RB lacks the non-conceptual, pre-reflective sense that his memories are his (Klein 2018). And this is the sense in which I will proceed.¹⁰¹¹ In the next section, I will lay out the first tool on the workbench for building my explanatory model of RB's condition.

2. The Necessity of Attention in Conscious Recollection

In the next two sections, I will be laying two tools on the workbench for explaining RB's curious lack of ownership over his memories: the necessity of attention in conscious recollection and the self-attentional network model of self-bias processing. But before getting into the details of these tools, I

⁹ I am aware that there is a distinction between judgements of ownership and feelings of ownership (Bayne and Pacherie 2007). Roughly, one judges that they own X, when they have a belief that they own X. In contrast, the feeling of ownership is characterized as a kind of non-conceptual awareness (de Vignemont 2007; Klein 2018). This latter notion is what I will assume.

¹⁰ De Vignemont (2011) distinguishes between three accounts of the relation between ownership and disownership: independent account theory, unified account theory, and discovery theory. The independent account states that the feeling of ownership is independent of the feeling of disownership, i.e., lack of ownership does not entail disownership. The unified account states that lack of ownership entails disownership and vice versa. Finally, the discovery account, which de Vignemont endorses, states that lack of ownership does not entail disownership, but lack of ownership and awareness of this fact does. I agree with de Vignemont that the independent account is mistaken. But I make no claims about whether the unified or discovery account is correct. However, my view is compatible with both.

¹¹ Some philosophers think that "mineness" should be understood in intentionalist terms, that is, that mineness is identical to or supervenes on content (Fernandez 2019). Some think that it should be understood in non-intentionalist terms. For example, that mineness is a feature of the mode of presentation of content (Klein 2018), or mineness is a metacognitive epistemic feeling (Perrin et al. 2020). Still others think that mineness is primitive (Billon 2017). Given that I think the feeling of ownership supervenes on the salience of self-related contents, my account assumes intentionalism.

want to provide an analogy to the kind of explanation that I will be giving. This will, hopefully, frame the rest of the discussion in a useful way for the reader.

2.1 An Analogy Explanation: Capgras' Syndrome

Capgras syndrome is a rare condition in which patients report that their loved ones are imposters. For example, patient DS claimed that his parents were imposters when he saw them. Hirstein and Ramachandran (1997) provide an elegant explanation. They note that when a person is considered a loved one, one's perception of them contains a kind of affective "warmth" or "glow". These researchers measured skin conductance responses to photographs of familiar people against responses to non-familiar people in patients with Capgras syndrome. They found that there was no difference in the patient's responses, but controls showed significant differences. The researchers argue that when patients view their loved ones and the affective glow is not present (as it was prior to the onset of the syndrome), patients post-hoc rationalize that the loved one is an imposter. There is something like an expectation that the affective component will be present in their perception, and when it is not, the patient tries to explain the discrepancy. I intend to provide a similar explanation for RB. Namely, that prior to the accident, self-bias processing played a crucial role in the recollective experience-- it raised the saliency of self-referential contents higher than other contents-- but because of the accident, self-bias is no longer playing that role. There is a kind of expectation that the saliency of self-referential content will be the same as before the accident, and when it's not, RB feels that his memories are not his own.¹² I now turn to a discussion of the necessity of attention for conscious recollection.

2.2 The Role of Attention in Episodic Recollection

Felipe de Brigard (2012) argues that attention is necessary, but not sufficient for conscious episodic recollection. In particular, he argues that *internal* attention (as opposed to external attention) is necessary. He follows Chun et al.'s (2011) taxonomy of attention according to which there are two basic attentional systems distinguished by the type of information attention operates over. External attention has as its objects external, environmental information, whereas internal attention operates over mental

¹² I do not mean that there is an explicit representation of such an expectation.

representations (Lau and colleagues 2004, as well as Chun et al. 2011, provide behavioral and neuroimaging evidence for such a view). Even though these systems are distinct in virtue of the information they operate over, there are shared features computationally as well as neuroanatomically. Chun et al. claim that both systems filter information, they are selective about that information, and they modulate other systems' performance.

Neuroanatomically, these systems are both implicated in prefrontal regions as well as anterior cingulate cortices. But for all they share, there are differences: a study by Nobre et al. (2004) showed that the right inferior parietal cortex extending to posterior angular gyrus is preferentially involved in external attention tasks (across modalities). Whereas the bilateral intraparietal sulcus, right ventral and bilateral dorsolateral prefrontal cortices are preferentially involved in internal attention tasks.

The dominant view on attention's role in episodic memory is that it is crucial for encoding, but not for retrieval. Parietal lesion patients with attentional deficits show no impairment during memory (recall) tasks, but severe impairment in encoding (see Baddeley et al. 1984). De Brigard notes however that these experiments were not using a *divided* attention paradigm.¹³ When this paradigm is implemented, patients with attentional deficits show severe impairment in recall tasks. The impairment is especially evident on the "remember"/ "know" task-- patients are typical for "know" judgements, but are impaired for "remember" judgements. This suggests that attention is especially implicated for episodic recollection. Other evidence comes from Berryhill et al. (2007). Patients with bilateral ventral parietal lesions (implicated in internal attention) showed reduced levels of free-recall during autobiographical memory tasks and in particular decreased levels of vividness and detail (see also Bisiach and Luzzatti 1978).

De Brigard hopes that at this point he has convinced the reader that internal attention is at least necessary for episodic recollection. But as he notes, there is still the question of whether attention is required for *conscious* recollection. Drawing on the global workspace model of consciousness (Dehaene

¹³ Standard divided attention paradigms have subjects engage in tasks that divide their attention between the two tasks. This paradigm is typically used to measure whether attention is implicated in the target task.

and Changeux 2000), he suggests that internal attention enhances (or increases) gamma-band responses above threshold (30-100Hz) in local processing networks which renders them available for broadcast. In other words, attention is the gatekeeper to consciousness for episodic contents (although as we will see, the gatekeeper is necessary, but not sufficient).

De Brigard notes that there are many regions that contribute to episodic recollection. Here's a recap:

- 1) Prefrontal cortex (PFC): initiation, monitoring, and maintenance (esp. Ventrolateral (vlPFC) and dorsolateral (dlPFC) regions)
- 2) Medial temporal lobe (MTL): required for successful binding and accessing of relational information from neocortex
- 3) Parietal cortex: filtering and selection of information distributed across sensory cortices

Perhaps not so coincidentally the prefrontal/medial-temporal/parietal network of activation is associated with information processing that gives rise to conscious awareness. This local network can globally broadcast its contents when the gamma-band responses reach threshold levels (between 30 and 100Hz). It turns out that attentional modulation of those contents spike the gamma frequency. So, attentional processing on those contents can increase the gamma-band responses above threshold.

If parietal regions modulate the availability of representations into the global workspace, then patients with parietal damage should (*ex hypothesi*) experience a loss of auto-noesis. The Berryhill et al. (2007) experiments (mentioned above) confirm this. Fewer representations to broadcast diminishes the subjective experience of recollection. Another reason is worth quoting:

The fact that richer retrieval cues increase the likelihood of successful retrieval is at the heart of the notion of retrieval support, but it also suggests that these richer cues work precisely because they have a better chance of 'highlighting' the relevant memory trace than poorer retrieval cues (De Brigard 2012, pg. 7).

De Brigard notes that this line is supported by Tulving's (1982) *synergistic ecphory* model. This model states that the degree of subjective recollection is the degree to which the memory traces interface with the retrieval cue. Unfortunately, there isn't much evidence in the way of neurology for this model.

Moving forward, there are two main reasons for thinking that attention is not sufficient for conscious recollection. Because this part is not as important for my purposes, I will only list them briefly:

- 1) Successful recollection requires that the memorial contents one attends to be reinstated during retrieval. This is a matter of priority. If certain contents are missing prior to attentional processing, then there will be a recollective failure not attributable to attention.
- 2) Attention is not an “all or nothing” process for broadcast. So, even if attention acts on the memorial content to increase gamma-band responses, it might not reach threshold levels.

If the above evidence is convincing, then we have good reason for thinking that attention is necessary for conscious recollection of episodic memories. As for the case of RB, we now have a tool on the workbench that goes some way towards an alternative explanation. Episodic memories contain self-referential content, e.g., “*I am eating pizza at XXX in Isla Vista.*” I want to suggest that attention is not appropriately operating over those contents (the “*I*”) to give rise to the feeling of ownership. Here, RB experiences the rich contextual details and can even see himself in the memory (there is some reference to oneself as a participant in the episode), but he doesn’t feel ownership due to a lack of self-biasing. In the next section, I’ll review evidence for a separate self-attentional network (SAN) that will explain why attention does not appropriately operate over the self-referential content of the memory. Indeed, the evidence for SAN will indicate a particular mechanistic explanation for RB’s condition.

3. The Self-Attentional Network and Self-Bias

In the previous section, I laid one tool on the workbench to explain RB’s phenomenology. In this section, I will lay the second tool on the workbench-- the self-attentional network. Humphreys and Sui (2016) think that self-bias modulates attention. In their discussion paper, the authors review three experimental paradigms for assessing self-biases and argue that self-related information has a differential impact on allocation of attention and furthermore, that it can alter the saliency of a stimulus. To account for this function, they propose a self-attention network that responds specifically to self-related stimuli and interacts with general attentional resources.

3.1 Behavioral Evidence of Self-Bias

Imagine being at a cocktail party. You are discussing the recent election with your friends and the conversation is getting heated. All of a sudden, you whirl around because you hear your name mentioned across the room. This phenomenon is known as the “cocktail party effect”. It turns out that humans have a self-bias toward their own name over other names. But this case is particularly puzzling because you were talking amongst friends on a complex topic that demanded many attentional resources. Yet, seemingly miraculously you attended to your name being mentioned from across the room. Presumably, there is lots of chatter at cocktail parties and your attentional processing filters out most of it so that you can talk with your friends without distraction. But when your name is called, it seems that the filter fails to stop the name from making demands on attention. What accounts for this? The evidence suggests that the processing of our own name is pre-attentive, automatic, and cues attention. For example, Moray (1959) found that subjects could report more words heard in an unattended ear after the presentation of their own name in that ear. Or consider the work of Bargh & Pratto (1986) and Wood & Cowan (1995) which shows that the central presentation of one’s name can disrupt detection in peripheral vision. There is a general worry here though-- perhaps our own name is just extremely familiar to us versus other names or words. This would mean that there is nothing special about self-related information-- that it does not activate special self-representations or modulate attention any differently *qua* self.

Similar self-bias effects have been reported for one’s own face-- that is, there is a self-bias for own-face versus other-face. In general, subjects are faster and more accurate at making perceptual judgments about the orientation of faces if they see their own face versus the face of another. One’s own face is more salient than other faces. Here, the worries about familiarity are still present but perhaps less so than own-name effects. For example, Sui and Humphreys (2013) showed that

[R]esponses to a friend’s face varied across categorization tasks as did responses to the face of a stranger, suggesting a qualitative distinction between the processing of self-faces and other faces. Given that a friend’s face will be highly familiar relative to the face of a stranger, then the self-face advantage cannot easily be attributed to (at least linear) effects of familiarity (Humphreys and Sui 2016, pg.8).

There is also evidence that own-face bias can cue attention. Liu, He, Rotshtein, and Sui (2016) showed that when subjects are cued with their own face versus the face of others, they could better identify a target letter.

Finally, this last experimental paradigm is supposed to set aside the familiarity objection. Subjects are trained on shape-name pairs, e.g. own name-Square, friend's name-Circle, stranger's name-Triangle. And then presented with pairs that may or may not be the same as the training set, e.g., own name-Triangle. Researchers found that there was a substantial advantage for matching self-related pairs over other-related pairs, i.e., self-related pairs are more salient than other-related pairs. Interestingly, when researchers increased the frequency of familiar and self-related pairs equally, there was still only a bias for the self pairs. This indicates that the self-related stimuli were differentially preferred to familiar and stranger stimuli.

3.2 The Neural Basis of Self-Bias: The Self-Attentional Network (SAN)

For my part, the behavioral evidence does not avoid the familiarity objection because the self bias might persist in virtue of being more familiar than the friend pair stimulus, e.g., own name-Square is more familiar than friend's name-Circle, and that is why subjects have the advantage for self pairs over friend pairs. But setting this aside, the neural evidence is perhaps more convincing. fMRI imaging revealed preferential activation for self-related stimuli in the ventral-medial prefrontal cortex (vmPFC) and the left-posterior superior temporal sulcus (lpSTS). Humphreys and Sui note that the vmPFC is suggested to be the seat of self-representation and the lpSTS is the seat of social attention to the environment. Temporally speaking, they note that the vmPFC is activated prior to the lpSTS suggesting that the self-representation primes attentional control. Contrastingly, for other-related pairs, fronto-parietal network was preferentially activated. So it turns out that self-related stimuli activates independent neural correlates. More interesting still, the authors note that damage to the vmPFC results in hypo-self responses (self-bias is eliminated) and damage to the lpSTS results in hyper-self responses.

To account for this data, Humphreys and Sui propose a self-attention network with three components (see figure 1):

- 1) A general purpose top-down attentional control mechanism (includes dorsolateral PFC and intraparietal sulcus),
- 2) Self-representation by the vmPFC,
- 3) And bottom-up orienting by the lpSTS.

They claim that vmPFC is triggered by the presence of self-related information which modulates subsequent attentional processing to the lpSTS and the general attentional control network. In general, there are bottom-up and top-down influences between the SAN and the general attentional control network.

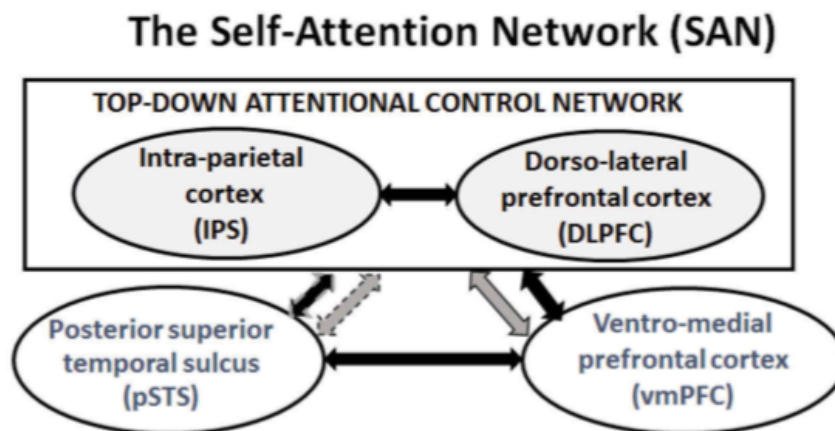


FIGURE 1. The Self-Attention Network (SAN). Here we distinguish between a top-down attentional network (including the intraparietal sulcus (IPS) and the dorsolateral prefrontal cortex (DLPFC)) and a network that responds in a bottom-up fashion to self-related information (the posterior superior temporal sulcus (pSTS) and the ventromedial prefrontal cortex (vmPFC)). Black arrows indicate excitatory connections. Grey arrows indicate inhibitory connections. Dotted arrows (to and from the pSTS) highlight that there is currently little direct evidence for these functional connections (from Humphreys and Sui 2016, pg.12).

Neuroimaging evidence for these connections comes from Sui et al. (2013).¹⁴ These researchers showed preferential activations and coupling between the lpSTS and vmPFC for self-related contents in a matching task (as in the own name-shape/stranger name-shape task), whereas there was preferential activation only in the DLPFC and IPS for other-related matching. The idea here is that when self-related content is present, vmPFC and lpSTS activate and exert bottom-up modulation on the general attentional

¹⁴ See De Brigard et al. (2015) for evidence that the vmPFC is preferentially involved in tasks (episodic recollection, future mental time travel, and episodic counterfactual thinking) with self-related contents. See Denny et al. (2012) for a meta-analysis of over 100 neuroimaging studies on self- versus other-related stimuli. Finally, see Wagner, Waxby, and Heatherton (2012) for an extensive review of the mPFC's role in self-related processing.

control network. When there are no self-related contents present, there is no such bottom-up modulation from these nodes-- the attentional control network operates business-as-usual.

If this is convincing, then we have good reason for thinking that there is a separate attentional network dedicated to processing self-referential information. This network, constituted by the vmPFC and lpSTS, explains self-bias processing, e.g., the cocktail party effect, by having a dedicated, bottom-up mechanism that activates and exerts influence on the general attentional control network to increase the salience of self-referential content.

4. SAN and Retrieval Processes

The discussion of the last section focused on self-bias and attention in *perception*. What about the role of self-bias in episodic recollection? If the SAN is supposed to be part of the explanation for RB's lack of feeling of ownership, then there must be evidence that the SAN is implicated in memorial processing. This section provides evidence for cortical midline regions' involvement in retrieval processes.¹⁵ Although there is no direct evidence that the SAN is implicated in retrieval processes (although see Cunningham 2016), there is ample evidence of bias for self-referential content in recollection (Symons and Johnston 1997). Moreover, ventral medial areas of prefrontal cortex are consistently found to be implicated in this processing (for a review see Cabeza and Nyberg 1997; Nyberg et al. 1996). Since we have independent reasons for thinking that internal attention is necessary for conscious episodic recollection and the SAN's involvement in self-bias processing generally, the evidence provided here gives us reason to think that the SAN is involved in retrieval processes of episodic recollection specifically. I will briefly discuss three studies and summarize the key takeaway message.¹⁶

The self-reference effect in memory (SRE) is well documented (Brédart 2016; Symons and Johnston 1997; Turk et al. 2008; Turk et al. 2011; van der Bos et al. 2010). When stimuli are processed with reference to the self, they are generally better remembered than stimuli processed otherwise. For example, Turk et al. (2012) asked subjects to encode items under imagined ownership by themselves or

¹⁵ By "cortical midline regions", I mean areas such as medial PFC, posterior cingulate cortices, and anterior cingulate cortices, among others. Of course, I am most interested in medial PFC.

¹⁶ Additional evidence can be found at Arzy and Schacter (2019); Dafni-Merom and Arzy (2020).

another person. The researchers found a significant advantage for recall of personally owned items compared to other-owned items. Using a “remember-know” paradigm, they found the SRE correlated with higher “remember” responses. More interestingly, when a divided attention condition was implemented at encoding, the SRE was eliminated. Turk and colleagues hypothesized that the SRE relies on attentional resources at encoding.

4.1 SRE at Retrieval Implicates MPFC¹⁷

Our present concern, however, is with attentional resources at *retrieval*. Fossati et al. (2004) investigated the neural correlates of retrieval success using an SRE paradigm. The researchers had subjects encode positive and negative personality traits in self-referential, semantic, and phonemic conditions. They then gave the subjects a recognition task where they were presented with new and old words and asked to select the previously presented words. The behavioral data showed a significant SRE for recall of both positive and negative traits. fMRI imaging during the recognition task revealed, “the successful retrieval of self-encoded emotional personality traits reactivated the right MPFC and activated a specific neural network including posterior cortical regions and sub-cortical areas” (Fossati et al. 2004, pg.1603). Contrastingly, while the MPFC was active for other encoded stimuli, greater activation was reported for the self-encoded stimuli. Since attention is necessary for conscious recollection and the SAN implicates the MPFC, a plausible explanation for this greater activation is that retrieval for self-related contents triggers the MPFC to bias general attentional resources in broadcasting. So not only is MPFC crucial for encoding of self-related stimuli, it is also crucial to retrieval processes of self-related content.

4.2 Incorporation of Self in Episodic Memories Implicates MPFC

Additional evidence comes from Kurczek et al. (2015) showing a double dissociation between hippocampal and MPFC contributions to autobiographical memory recall. The researchers had subjects recall a past event given a neutral word cue. Some subjects had hippocampal damage, some had MPFC damage, and the others were healthy controls. Quantifying contextual details and self reference, Kurczek

¹⁷ The reader may have noticed that I am discussing the MPFC now and not the vmPFC. But the vmPFC is a part of the MPFC. The studies discussed in this section used a coarser grain of analysis.

and colleagues found that patients with hippocampal damage, “showed disruptions in their ability to construct highly detailed episodic events across all time periods, but this deficit did not extend to their ability to incorporate themselves in the narratives of those (re)constructions” (2015, pg.123). In contrast, patients with MPFC damage, “showed the opposite pattern. They were able to construct highly detailed episodic events but incorporated themselves in the narratives of those (re)constructions less often than healthy participants” (pg.123). Of present interest is the latter finding-- that MPFC seems to be implicated in self-reference for autobiographical recall. Kurczek et al.’s study was not testing encoding because these events were memories previously encoded. Thus, MPFC is implicated in episodic retrieval processes especially for self-referential content.

4.3 Retrieval of Self-Related Content in Episodic Recollection Implicates MPFC

Finally, a series of studies by Moroz (1999) showed aged-related differences in “remember” versus “know” responses. Subjects encoded stimuli in terms of self, other, social desirability, and syllable.¹⁸ After the encoding task, subjects were given a recognition test similar to the one given in Fossati et al. (2004), but with a “remember-know” paradigm. Young adults reported more “remember” responses than older adults especially for self-encoded words. Moroz hypothesizes that this is due to age-related reductions in cognitive resources, e.g., attentional resources.

To test this hypothesis, Moroz implemented a divided attention paradigm at retrieval for young adults. The results show that young adults in the divided attention condition had reduced “remember” responses on par with the older adults in a full attention condition. The “remember” responses were especially reduced for self-encoded words. This suggests attentional resources are implicated in retrieval of previously encoded stimuli, but especially for self-encoded stimuli.

Finally, Moroz conducted a PET study at retrieval. The imaging revealed preferential activation of anterior-frontal areas of the right PFC for retrieval of self-encoded stimuli. Other-encoded stimuli showed preferential activation of more posterior-frontal areas of right PFC. Moreover, differences were

¹⁸ The stimuli were trait adjectives. For the social desirability condition, subjects judged how desirable each trait is. For the syllable condition, subjects judged how many syllables the word has.

revealed along the left-right axis. Moroz reports, “whereas the self [condition] was associated with a pattern of neural activity that included bilateral frontal areas, the social desirability [condition] was associated with a pattern of neural activity that included a left frontal area and medial anterior cingulate, and the other [condition] was associated with a pattern of neural activity that included only medial anterior cingulate” (1999, pg.108). Taken together, Moroz (1999) provides compelling evidence that attentional resources are recruited at retrieval for episodic recollection. More specifically, right prefrontal areas (including MPFC) are recruited for retrieval of self-related stimuli.

At this point, one might object that none of this evidence demonstrates that an impaired SAN gives rise to the kind of disownership experience that RB has. That is, none of the evidence canvassed here explicitly shows that if the modulatory connections between the SAN and general attention are severed, then subjects will report feeling disownership over their memories. In way of response, I don’t think that means this account is mistaken. Rather, from what I gather, the question of whether an impaired SAN can give rise to feelings of disownership has not even been pursued. A lack of evidence is not evidence of a lack. Indeed, I think this is what makes my view novel, interesting, and worth pursuing further.

We started this section with the question of whether there is any evidence that midline regions (specifically MPFC), are implicated in episodic retrieval processes. The question is important because we have good reason for thinking that attention is necessary for conscious episodic recollection (De Brigard 2012) and that the SAN biases general attentional resources for self-related *perceptual* stimuli (Humphreys and Sui 2016), but if these tools are going to explain RB’s phenomenology, we need evidence that the SAN biases attentional resources for broadcast of self-related *memories*. The three studies discussed in this section all implicated the MPFC (Brodmann’s area 8/9) in retrieval processes of self-related memories. Fossati and colleagues showed greater activation in this area for recall of self-encoded personality traits. Kurczek and colleagues demonstrated a double dissociation between hippocampal contributions to contextual details and MPFC contributions to self-related contents of autobiographical memories. Moroz reported reduced “remember” responses, especially for self encoded

stimuli, if attention was divided at retrieval. Furthermore, Moroz provided neuroimaging that showed preferential activation of BA 8/9 for retrieval of self-encoded stimuli. In the next section, I will be tying these threads together for an explanation of RB's lost feeling of ownership.

5. The Self-Attentional Model of RB's Lost Feeling of Ownership

5.1 Outline of Self-Attentional Model

Having reviewed neuroscientific evidence for the necessity of attention in conscious recollection and the self-attentional network, I turn now to tying these threads together for an explanation of RB's lack of ownership over his memories. In broad strokes, there are three pieces to the self-attentional model. Firstly, as the global workspace model states, internal attention operates over memorial content in local processing networks to increase gamma-band responses above threshold so that the memory is broadcast to the global workspace. This, as de Brigard has argued, is what accounts for conscious episodic recollection. Secondly, the typical high salience of self-content is partially constitutive of the supervenience base for the feeling of memory ownership. That is, the typical salience of self-related contents is necessary for the feeling of ownership. I will elaborate on this more below. And lastly, self-bias processing (on memorial content) is sufficient for the production of the typical threshold level of salience. As I showed in previous sections, the self-attentional network is biased toward self-related contents and increases the salience of such contents. Altogether, in a neurotypical subject, self-related memorial content activates the self-attentional network and increases its salience. This typical high level of salience is partially constitutive of the supervenience base for the feeling of ownership. As per the global workspace model, internal attention operates over this supervenience base. And assuming phenomenology supervenes on content, the feeling of ownership comes about. I have provided evidence for the first and last parts of this model already, so I now turn to evidence for the necessity of the typical high level of salience of self-related content for the feeling of ownership. After that, I will use this model to explain RB's lost feeling of ownership.

5.2 The Necessity of Salience

That self-content is more salient than other-content is not controversial. The self-bias literature briefly reviewed here demonstrates that self-content is distinctive in its salience. But that it is *necessary* for the feeling of ownership is controversial. So here I defend premise 1. Assume for reductio that self-content does not differ in salience from other-content. Then, all else being equal, the experience of each should not differ either. But the experience of each does differ. Indeed, we have a feeling of ownership for self-content, and not for other-related content. So, salience must differ between self- and other-content.

For example, when I remember taking a road trip with my partner to Duluth, I have a feeling of ownership over this memory-- I feel as if I experienced that road trip. Contrastingly, suppose a friend of mine tells me about a road trip she took recently. It is not the case that I have a feeling of ownership associated with *that* road trip-- I do not feel as if that road trip happened to me. This difference in experience can be explained by the difference in salience. In the former case, my friend is the participant in the road trip (other-content). In the latter case, I am a participant in the road trip (self-content).

But we need not rely on intuitions. William James (1890) famously characterized the feeling of recollecting personal memories as warm and intimate. I take it that the salience of self-content is (at least part of) what generates this feeling.¹⁹ But how does this happen? Amygdaloid contributions to episodic recollection are well-established (Cahill 1995; Phelps and Anderson 1997). In fact both Markowitsch et al. (1997) and Fink et al. (1996) demonstrated amygdaloid activation for recollection of personal, but *not* impersonal memories. It is thought that the amygdala contributes a unique affective quality to the recollection of personal memories, hence the warmth and intimacy that James described.²⁰

But why does the amygdala activate in recollection of personal memories? Schmitz and Johnson (2006) identified a ventral MPFC network that includes the amygdala, insula, and nucleus accumbens. These researchers performed fMRI imaging during an appraisal task. Subjects were asked to decide if presented personality traits applied to themselves (self-appraisal condition) or whether the personality

¹⁹ Because I think saliency is only a necessary condition, not sufficient.

²⁰ It is also worth noting that Hirstein and Ramachandran (1997) argued that missing affective glow for perception of loved ones in Capgras' syndrome is caused by missing amygdala connections.

trait was positive (non-self appraisal condition). In the self-appraisal condition, the ventral MPFC network was activated. In the non-self appraisal condition, a more dorsal MPFC network was activated. This suggests that self-related information activates the ventral MPFC and the limbic structures in this network. In contrast, non-self related information does not activate the ventral MPFC. Thus, the amygdaloid contributions in the self-appraisal condition are uniquely modulated by the ventral MPFC.

The most compelling evidence for the necessity of salience for feelings of ownership comes from studies on depersonalization. Depersonalization is a disorder characterized by feelings of “unreality”. The disorder can take many forms. Patients report feeling like their body is not real, they are in a dream, and/or their thoughts are not real. Depersonalization can be local in the sense that only some parts of patients’ experience feel unreal, or global in the sense that every aspect of their experience feels unreal. Additionally, depersonalization can be primary or secondary. When depersonalization is secondary, it is considered a symptom of a larger disorder, e.g., PTSD (for review see Sierra and David 2011. For book length treatments see Sierra 2009; Simeon and Abugel 2006).

Sierra (2009) distinguish between four main symptoms in primary depersonalization:

- (1) Anomalous body experience,
- (2) Emotional numbing,
- (3) Anomalous subjective recall, and
- (4) Alienation from surroundings (pp.26-27).

Anomalous body experience can take many forms, but feeling like your body is not real is one example. Anomalous subjective recall can also take many forms, but a common reporting is feeling like one’s memories are not one’s own. In emotional numbing, patients report neither feeling happy nor sad, just neutral or nothing at all. Finally, patients report feeling alienated from their environment, like the world is constructed or dream-like. Importantly, patients intellectually know that things are not the way they feel them to be, but nonetheless cannot shake the feelings. Consider the report of the following patient:

I feel some degree of ‘out of it’ all of the time, but it has almost become to be what I am used to now. I get times when I feel very out of my body. I am looking at people, know who they are, but

can't place myself there. *I remember events from the past, but don't always see 'me' there.* Even photos of me look different. I don't like the person I remember being. Looking in the mirror proves difficult as I don't always recognize the person looking back at me (Sierra 2009, pg.27; emphasis added).

In the neuroscientific literature on depersonalization, the consensus view seems to be that the disorder arises from emotional inhibition resulting from hyperactivity in prefrontal regions. For example, Sierra and Berrios (1998) propose a cortico-limbic disconnection model (see also Phillips and Sierra 2003). In this model, hyperactive prefrontal regions *inhibit* anterior cingulate cortex and amygdala. However, Simeon et al. (2000) conducted a PET study on a verbal memory task between patients and healthy controls. The result of this study showed higher metabolism in occipital and parietal regions with no changes in prefrontal regions for depersonalized patients. Poulin and Black (2006) commented on Simeon and colleagues' study saying the following:

The choice of a verbal memory task by Simeon and associates might account for the discrepancy with the Sierra and Berrios model. An autobiographic rather than a verbal memory task might have recruited limbic and medial prefrontal regions involved in self-referential mental activity (pg.3).

Back in 2005, while describing the phenomenon of anomalous subjective recall, Sierra and colleagues commented that no imaging studies of autobiographical recall in depersonalized patients had been conducted. But that changed in 2007 when Hooper et al. conducted a study on dissociative PTSD patients.²¹ Dissociative PTSD is a subtype of PTSD wherein patients dissociate from their traumatic memories, i.e., they have anomalous subjective recall. Patients often report that recollection of these events is like watching a film where the events did not really happen to them (Ataria 2014). In contrast, patients with hyperarousal PTSD re-experience the traumatic events in high detail and vividness. Hooper et al. (2007) showed that trauma reminders induced a dissociative response in dissociative PTSD subjects. Imaging revealed hyperactivity of MPFC and hypoactivity of insula and amygdala. These results are consistent with Sierra and Berrios' cortico-limbic disconnection model (for a review see Lanius et al. 2011).

²¹ There are studies by Medford et al. (2006) and Medford et al. (2016), but these tested emotional memory, rather than autobiographical memory.

Evidently, what explains the feelings of alienation, unreality, or lack of ownership in depersonalized patients with anomalous subjective recall is hyperactivity of MPFC and hypoactivity of limbic structures. Thus, the distinctive contribution that limbic structures make in retrieval is necessary for the feeling of ownership. These limbic contributions are part of what it is for self-related content to be more salient than other-related content. Without such saliency, the feeling of ownership is lost. This is because the salience of such content is partially constitutive of the supervenience base for the feeling of ownership. Furthermore, if the limbic contributions are part of what it is for self-related contents to be salient, then limbic contributions are part of self-bias processing.²² Indeed, depersonalized patients do not benefit from self-bias (Farmer et al. 2020).

5.3 Application of the Model to RB

If this has been convincing, then we have evidence for each of the three parts of the self-attentional model. We have evidence that (1) internal attention operates over memorial content in local processing networks to increase gamma-band responses above threshold so that the memory is broadcast to the global workspace (section 2), (2) the typical salience of self-related contents is necessary for the feeling of ownership (section 5.2), and (3) self-bias processing (on memorial content) is sufficient for the production of the typical threshold level of salience (sections 3 and 4). If any one of these pieces is disrupted in some way, then the feeling of ownership is disrupted.

For RB, he does have recollective experiences, so (1) is not disrupted. It is my contention that RB has lost modulatory connections between the SAN and general attentional resources. Effectively, this eliminates bias toward self-referential content thus diminishing the salience of that content. Typically, self-related contents enjoy a higher salience than other-related contents in virtue of attentional biases toward self-related content. So if the bias is eliminated, the saliency drops to the level of other-related

²² It is worth noting that Humphreys and Sui (2016) state that “future work should not only attempt to better characterize the effective and functional connectivity between vmPFC and pSTS, but also try to elucidate *which intermediate cortical or subcortical nodes are missing in the SAN model* and what modulatory role they may play” (pg.25; emphasis added). My suggestion, informed by the evidence canvassed here, is that the amygdala and probably the insula are modulated by MPFC and contribute to the high salience of self-related content. I agree further work should attempt to test this.

content. This was demonstrated in Moroz's (1999) divided attention at retrieval experiments. Young adults, in a divided attention paradigm, showed reduced "remember" responses for self-encoded stimuli. Because self-related contents demand more attentional resources, retrieval of those contents under divided attention conditions diminishes performance. We also saw this in Kurczek et al.'s (2015) MPFC damaged patients who showed far less self-reference in episodic recall than healthy controls. In the case of RB, I am not claiming that he has a damaged vmPFC. RB does not have an identity issue-- he can self-represent. Rather, I am suggesting that the vmPFC cannot exert its normal influence on general attention to increase the salience of self-content. So the self-content is broadcast at the *same* salience level as other-related content. This effectively causes RB to not feel ownership over those memories. Finally, Even if RB has typical connections between MPFC and limbic structures, the influence that MPFC has on general attention is eliminated. It is possible, however, that RB has cortico-limbic disconnections *in addition* to severed modulatory connections. Without a neuropsychological profile, it is hard to know for sure. But given the evidence canvassed here, I think we can be confident that RB lacks typical salience of self-related content, and this explains his lost feeling of ownership. Moreover, given the similarity in phenomenology between RB and depersonalized patients with anomalous subjective recall, the self-attentional model likely has wide applicability and is worth pursuing further (see section 6).

5.4 An Objection to the Necessity of Salience

Before moving forward, I want to consider an objection to the necessity of salience of self-content for the feeling of ownership.²³ Consider the following two memories:

Warm Football Game: I remember being at the football game, seeing the bright shining sun, the players running up and down the field, and the beautiful mountains that formed the horizon. My team won 36-30. It was such a close game! The final play was a hail mary from halfway down the field.

Cold Football Game: I remember being at the football game, how cold the air was on my face, my nose was running uncontrollably, my hands were shaking, dry and cracked from the whipping wind. I couldn't get warm. I couldn't wait for the game to end. I kept thinking to myself: *how much longer is this gonna take? I can't wait to get home and warm up by the fire.*

²³ Thank you to an anonymous reviewer for this fascinating and challenging objection.

It seems plausible that the subject of these memories, call her Hayley, feels ownership over both of these memories-- that they are experienced as things that happened to them. But it also seems plausible that in *warm football game*, the saliency of self-content is diminished compared to *cold football game*. Indeed, the details of *warm football game* are more focused on the external world, whereas the details of *cold football game* are more focused on Hayley. But this is problematic because my view would predict that Hayley lost the feeling of ownership with respect to *warm football game* if the saliency is diminished. So saliency is not necessary for the feeling of ownership.

In way of response, I want to deny that the self-attentional model predicts that Hayley would lose the feeling of ownership with respect to *warm football game*. Firstly, I agree that the self is more salient in *cold football game*, than in *warm football game*, but this does not entail a lost feeling of ownership. On the self-attentional model, there is a threshold level of saliency required for the feeling of ownership. So there can be differences in saliency, granted the differences are above threshold, that do not make a difference to the feeling of ownership. However, once the saliency of the self-content drops below threshold, the feeling of ownership is lost.²⁴

5.5 Explanation of RB's Isolated Lack of Ownership

I now want to address a dimension of RB's pathology that I noted in section 1. RB reported that he lacked the feeling of ownership *only* for memories prior to the accident. That is, he felt that he owned memories formed post-morbid. It would seem that the self-attentional model cannot explain this asymmetry. This is because if there are severed modulatory connections between the vmPFC and general attention, the lack of self-bias at retrieval would affect all memories regardless of when they were formed. And if RB feels ownership despite the lack of self-bias, then salience of self-related content is not necessary for the feeling of ownership.

²⁴ At this point, the reader might reasonably wonder what the threshold is for the feeling of ownership. The reader might also wonder what the proportion of saliency is compared to other content. I submit that while these are interesting questions, they are empirical and warrant future investigation. As such, answering these questions is beyond the scope of this paper.

Intuitively, this seems right, but empirically this is likely not the case. Neuropsychological studies on so-called “isolated amnesia” have revealed independence of retrograde and anterograde memory processes (Goldberg et al. 1981; Hunkin et al. 1995; Kapur et al. 1992, 1996; Mattioli et al. 1996). Isolated amnesia is described as having either retrograde or anterograde amnesia, but not both. Typically, retrograde and anterograde amnesia co-occur. But there are some cases where patients have severe retrograde amnesia, but relatively mild or non-existent anterograde amnesia.

Kapur et al. (1992) provided a neuropsychological and behavioral profile for a patient LT with isolated retrograde amnesia. LT suffered a traumatic closed head injury. fMRI revealed multiple lesions in the temporal lobes. Behavioral tests for anterograde and retrograde memory performance revealed typical performance for anterograde, but marked deficiency for retrograde. In a comparison study for cued recall of events pre- and post-morbid, LT could not recall any events premorbid, but recalled 20 post-morbid events (the maximum for the test). To explain this asymmetry, Kapur and colleagues conclude that different mechanisms must subservise retrograde and anterograde memory processes. Indeed, this seems to be the consensus view among researchers studying isolated amnesia (see above cited works, and Kopelman 1993 for a review). Moreover, retrograde memory processes seem to be further mechanistically distinguishable. Kopelman (1993) argues on the basis of clinical evidence that temporal damage is correlated with storage deficits, but *prefrontal* damage is correlated with retrieval deficits.

RB seems to have lost the feeling of ownership for premorbid memories (retrograde), but not for post-morbid memories (anterograde). Thus, RB’s pathology can be seen as a very specific kind of isolated retrograde disruption at retrieval. In particular, a disruption of *retrograde* self-bias processing at retrieval. This is consistent with Kopelman’s position that prefrontal damage is correlated with retrieval deficits, because I am claiming that RB has disrupted modulatory connections between the vmPFC and general attention that impairs self-bias at retrieval.

If RB has an isolated retrograde self-bias deficit, this implies that there is *anterograde* self-bias processing subserved by different mechanisms. RB’s anterograde self-bias processing seems to be typically functioning because he reports that he feels ownership for memories formed post-morbid. If

that's right, then salience of self-content as a necessary condition for the feeling of ownership is not threatened. As far as I know, there is no literature on this distinction, but it could explain RB's curious pathology. Therefore, it is worth further investigation.

In summary then, the self-attentional model of the feeling of ownership explains RB's lack of ownership by suggesting that RB has severed modulatory connections between the vmPFC and general attentional resources. Severed connections eliminate self-bias processing at retrieval effectively diminishing the saliency of self-related memorial content on par with other-related content. As I argued, the typical high salience of self-related content (provided by self-bias processing) is necessary for the feeling of ownership. Because self-bias processing at retrieval is impaired, salience is diminished below typicality (to the level of other-related content). The self-content is then broadcast at the same salience level as other-related content. Thus, RB lacks the feeling of ownership.

However, as noted, RB's lack of feeling of ownership is not global, but isolated to pre-morbid memories. This asymmetry can be accommodated by the self-attentional model by distinguishing between retrograde and anterograde processes. Thus, RB has disrupted retrograde self-bias processing at retrieval, but intact anterograde self-bias processing. Studies of isolated amnesia suggest mechanistic independence of retrograde and anterograde memorial processes. So, it is possible that anterograde self-bias processes are subserved by different mechanisms than the one identified in this paper. Indeed, it seems that *must* be the case because if retrograde and anterograde self-bias are subserved by the same mechanism, then RB should lack ownership for memories formed post-morbid as well. As I mentioned above, this specific distinction has not been discussed in the literature, and as such, warrants further investigation. In the next section, I will be discussing predictions of the self-attentional model, as well as potential future work.

6. Predictions and Further Work

In this section, I will discuss two predictions that the self-attentional model makes with respect to RB. Then I will discuss two potential applications of the self-attentional model.

6.1 Predictions about RB

Firstly, I argued that what explains RB's loss of feeling of ownership over his (premorbid) memories is diminished salience of self-related content as a result of impaired self-bias at retrieval. One might think that impaired self-bias should disrupt other self-related processes and states besides memories, e.g., occurrent perceptions, dreaming, imaginings. Perhaps RB should no longer experience the cocktail party effect, for example. This is an interesting, and admittedly open question. Klein reports that his interview with RB only focused on memories.²⁵ So it is entirely possible that RB experiences lack of ownership over other kinds of self-related states (or other kinds of disruptions). At least, there is no evidence that RB's disruptions are *only* related to memorial processes.

Second, and relatedly, there is much evidence that prospection, or imagining one's personal future, deploys similar neural resources as episodic recollection (De Brigard 2014; Gilbert and Wilson 2007; Schacter and Addis 2007). Indeed, Schacter, Addis, and Buckner (2007) argue, based on neuroimaging results, that remembering the past and imagining the future is subserved by a core network including medial prefrontal regions, posterior regions in the medial and lateral parietal cortex, the lateral temporal cortex, and the medial temporal lobe. If I am right that medial prefrontal regions (vmPFC) are responsible for self-bias processing at retrieval of episodic memories, then it's likely that those regions are playing the same role in imagining the personal future. Thus, RB might be similarly impaired for planning for and imagining his personal future.

There is perhaps some minimal evidence that this prediction is true. Consider the following quote from RB's interview with Klein:

[I] was able to plan for the future. Although my working memory loss and lack of skill at compensation made it *challenging*. When I slowly returned to work, *it was hard to plan a complex strategy*. I had to think of useful things to do and then do them. The best compensation I found was to separate the planning of the strategy from the execution. It worked best if I made a list of Things To Do. Then I could handle doing them one at a time (Klein and Nichols 2012, pp.688-689; emphasis added.).

²⁵ Personal correspondence.

It is true that RB could plan for the future, but as he says, it was difficult. Admittedly, it is not clear why it was difficult. That is, it is not clear that lack of felt ownership was making it hard. But it is clear that we cannot rule it out.²⁶

6.2 Applications and Future Work

Having discussed two predictions of the self-attentional model, I turn now to potential applications of this model to two other psychological conditions: depression and schizophrenia. Patients with these conditions have self-related processing, attentional, and autobiographical memory deficits. The self-attentional model might have resources to explain these deficits holistically. For sake of space, discussion will be brief.

6.2.1 Depression

Depression is often characterized by increased self-focus with respect to negative stimuli (Smith and Greenberg 1981). That is, increased attention to negative stimuli in relation to the self. Lemogne et al. (2012) argue, on the basis of several fMRI results, that there are, “two modes of elevated MPFC activation in major depression, either tonic in ventral MPFC or phasic in dorsal MPFC” (Lemogne et al. 2012; pg.e7). The authors suggest that ventral MPFC activation is associated with *automatic* depressive self-focus, whereas dorsal activation is associated with *strategic* depressive self-focus. Automatic self-focus is implicit and unconsciously driven. In contrast, strategic self-focus is explicit.²⁷ Both cases of elevated activation arise from lack of inhibition of the default network. Indeed, the researchers report that “compared to controls, and consistent with activation findings, depressed patients displayed an increased functional connectivity between [ventral] MPFC region, the DLPFC and the dorsal MPFC, including the supragenual ACC” (Lemogne et al. 2012; pg.e8). The DLPFC, as we saw in the discussion of Humphreys and Sui (2016), is implicated in general attentional control. Overall, it seems that depressive self-focus is explained by heightened activation of both (or either) ventral and dorsal MPFC. This heightened

²⁶ It is also worth noting that this quote is the first (and as far as I know, only) time we are told that RB had working memory loss.

²⁷ Lemogne and colleagues claim that automatic tagging of incoming information is a paradigmatic case of automatic self-focus and self appraisal judgements are cases of strategic self-focus (see pg.e8).

activation evidently makes demands of general attentional resources toward negative stimuli. This results in depressive phenomenology.

As it turns out, this (self-) attentional impairment is not unrelated to another symptom of depression: impaired episodic autobiographical recall. Lemogne et al. (2005) gave depressed individuals an autobiographical memory task to assess their recall of positive and negative personal memories along three dimensions: specificity, auto-noetic consciousness (remember/know), and perspective (field versus observer). The results indicate that, compared to controls, depressed subjects score higher for negative memories and lower for positive memories along all three dimensions.

The self-attentional model that I have put forward to explain RB's pathology, can be used to explain both the attentional and episodic memory impairments. First, ventral MPFC was implicated in heightened depressive self-focus. If Humphreys and Sui are right about the interactions between the SAN and general attention, then the ventral MPFC, in being biased toward negatively valenced self-related memorial contents, influences general attention to increase the salience of those contents. When broadcast to the global workspace, those negatively valenced self-related contents are experienced more vividly than unbiased positive memories. Moreover, due to the comparative salience of negative versus positive memories (negative being more salient), all else being equal, the negative memories will have a heightened feeling of ownership than the positive memories. At least, this is a prediction made by my explanation of RB's lack of ownership. It is worth further investigation. In summary, the attentional and autobiographical memory impairments in depressed populations can potentially be explained holistically by the self-attentional model.

6.2.2 Schizophrenia

Self-disturbances are a hallmark of schizophrenia. Self-disturbances range from source-monitoring deficits, lack of sense of bodily and agential ownership, and impaired autobiographical memory encoding and retrieval (for a review see Nelson et al. 2014a,b). For autobiographical memory, patients typically have overgeneral retrieval and a lack of specificity in details. Overgeneral retrieval is when a subject is asked to recall an event from a specific time and place, but instead retrieves multiple,

extended events and/or repeated events (McLeod et al. 2006). Patients also experience auditory hallucinations which are thought to result from inner speech not recognized as such (Frith and Done 1989).

Recent neuroimaging studies have revealed abnormal activity in self-related processing areas. For example, Chai et al. (2017) measured resting state activity using fMRI and found decoupling between ventral MPFC and DLPFC. In a meta-analysis of resting state imaging studies, Kühn and Gallinat (2013) found hypoactivity in ventral MPFC. Finally, Mier et al. (2017) had patients perform a social-cognitive task during fMRI and found hyperactivity of right and left posterior superior temporal sulcus (pSTS).²⁸

Recall that the SAN has two components: the ventral MPFC and the left pSTS. The SAN picks up self-related contents and modulates general attention (including DLPFC). If schizophrenic patients have decoupled connectivity between ventral MPFC and DLPFC, then self-bias should be minimized or non-existent. Moreover, hypo-ventral MPFC activity suggests deficient self-representation. Lastly, increased activity of pSTS, because it modulates ventral MPFC in the SAN, suggests inhibited self-representation.²⁹

All of this is convincingly explanatory for the schizophrenia patients' degraded sense of self. But as far as I can tell, no studies have been done on the sense of ownership over personal episodic memories in schizophrenia patients. The prediction of the self-attentional model is that these patients would have a degraded or eliminated sense of ownership. One barrier to testing this, however, is that many patients have difficulty even retrieving personal memories (Gold et al. 1992), and when they do retrieve them, they often are overgeneral and lacking details. A degraded sense of ownership might be a result of the lack of detail, and not necessarily lack of self-bias.

Relatedly, Jimenez et al. (2018) performed an fMRI study on controls and schizophrenic patients to assess the self-reference effect (SRE) at encoding and retrieval. Interestingly, there was no significant difference in performance between the groups for the self-condition. However, fMRI revealed differences

²⁸ See also Shin et al. (2015).

²⁹ It is worth mentioning that Pankow et al. (2015) argue that schizophrenia patients attribute salience to non-self related stimuli and this is (at least partly) explained by the hypoactivity of ventral MPFC.

in activity between the controls and patients. Controls showed higher activation in ventral MPFC for self versus other, but patients showed no difference in activation. The imaging data is consistent with the above evidence for hypoactivity in ventral MPFC, but inconsistent with the self-attentional model's prediction that self-bias would be eliminated. Nonetheless, Jimenez et al. conclude that patients did *not* benefit from SRE at retrieval. Here's what they say:

These findings suggest that either the conceptual or abstract representation of self may not be as well established in schizophrenia, or that perhaps patients have an impaired ability to distinguish self from other, or both. Furthermore, at retrieval, patients failed to demonstrate differentiation at the neural level between words previously encoding with respect to the self versus other. This suggests that at retrieval patients did not benefit from enhanced self-processing at encoding to the same degree as controls (2018, pg.26).

This is hardly satisfying given the behavioral results. Mere neural differences do not seem to be sufficient to show that patients did not benefit from SRE, given that they performed on par with controls. At any rate, this is the only study I know of that tests recall in schizophrenia using an SRE paradigm. More work needs to be done.

7. Conclusion

Episodic recollection typically involves a feeling that the remembered events happened to the rememberer-- the feeling of ownership. Klein thinks that this feeling is a constituent of auto-noetic consciousness. But sometimes the feeling of ownership can be disrupted or eliminated. Patient RB lost the feeling of ownership for memories formed pre-morbidly. This paper put forward a causal explanation of RB's phenomenology-- the self-attentional model. On this model, RB lost modulatory connections between the self-attentional network and general attentional resources. More specifically, the modulatory connections between the ventral MPFC and general attention has been disrupted. This disruption effectively eliminates self-bias processing at retrieval which diminishes the saliency of self-related content. Typically, self-related content enjoy a higher level of salience than other-related content. This higher level of salience, I argued, is provided (at least in part) by limbic contributions modulated by the MPFC. When salience of self-related content is diminished to the level of other-related content, the feeling of ownership is lost because that high salience is necessary for the feeling of ownership. Thus,

when RB's memorial content is broadcast to the global workspace, it is experienced in the same way as other-related content.

Additionally, RB claims that he feels ownership over memories formed post-morbidly. This asymmetry, I argued, can be accommodated by the self-attentional model by distinguishing between retrograde and anterograde memory processes, a well-established distinction in the literature on isolated retrograde amnesia. The novel contribution I made here was to suggest that RB's post-morbid feelings of ownership can be explained by independent anterograde self-bias processes that are not impaired. Whereas, his lost feelings of ownership for premorbid memories is explained by retrograde self-bias processes that are impaired. If it is the case that self-bias takes place at retrieval to endow self-related contents with their distinctive salience, then self-bias processes are memorial processes that can be either retrograde or anterograde. Further studies need to be done to test this hypothesis, however.

The self-attentional model does not merely explain RB's condition, but has much wider applicability. I discussed anomalous subjective recall in depersonalization disorder where patients feel like the events in their memories did not happen to them. The similarity in phenomenology to RB suggests that the self-attentional model can potentially explain some aspects of depersonalization. I also discussed how the attentional, memorial, and self-related impairments in depression and schizophrenia might be holistically explained by the self-attentional model as well. But there are other psychopathological cases involving disturbances of the self that might benefit from the self-attentional model, e.g., dissociative identity disorder and bipolar disorder.

Psychopathological cases aside, if the self-attentional model can (at least partially) explain the feeling of ownership, and the feeling of ownership is a constituent of auto-noetic consciousness, then we have made some progress toward identifying the neural correlates of auto-noetic consciousness.³⁰ In discussing the case of RB, Klein suggests that episodic memory is a kind of mental experience that is distinguished from semantic memory, not by content, but by whether auto-noetic consciousness attaches to

³⁰ I say that the self-attentional model can at least partially explain the feeling of ownership because I think that the model puts forward only a necessary condition for ownership, not a sufficient condition.

content at retrieval. I take it that I have gone some way toward identifying the “binding mechanism” that Klein had in mind.

One limitation of this project is that RB does not have a neuropsychological profile. Indeed, the only data on RB that we have are selected quotes from an interview with Klein. As such, the proposed model was created based solely on these phenomenological reports. However, given the similarity in phenomenology between RB and depersonalized patients, and given that we have neuropsychological studies on depersonalized patients, the self-attentional model is not baseless. Indeed, the model is empirically testable and explanatorily and predictively robust. The hope is that further empirical work will be done to confirm or disconfirm the model’s predictions.

References

- Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences*, *110* (Supplement 2), 103.
- Arzy, S., & Schacter, D. L. (2019). Self-agency and self-ownership in cognitive mapping. *Trends in cognitive sciences*, *23*(6), 476-487.
- Ataria, Y. (2014). Acute peritraumatic dissociation: In favor of a phenomenological inquiry. *Journal of Trauma & Dissociation*, *15*(3), 332-347.
- Ataria, Y. (2015). Dissociation during trauma: The ownership-agency tradeoff model. *Phenomenology and the Cognitive Sciences*, *14*(4), 1037-1053.
- Baddeley, A. D., Lewis, V., Eldridge, M., and Thomson, N. (1984). Attention and retrieval from Long-term memory. *J. Exp. Psychol. Gen.* *13*, 518–540.
- Bargh, J. A., & Pratto, F. (1986). Individual construct accessibility and perceptual selection. *Journal of Experimental Social Psychology*, *22*, 293–311.
- Bayne, T., & Pacherie, E. (2007). Narrators and comparators: the architecture of agentive self-awareness. *Synthese*, *159*(3), 475-491.
- Bernecker, S. (2010). *Memory: A philosophical study*. Oxford University Press.
- Berryhill, M. E., Phuong, L., Picasso, L., Cabeza, R., and Olson, I. R. (2007). Parietal lobe and episodic memory: bilateral damage causes impaired free recall of autobiographical memory. *J. Neurosci.* *27*, 14415–14423.
- Billon, A. (2017). Mineness first: three challenges to contemporary theories of bodily self-awareness. In Adrian J. T. Alsmith & Frédérique de Vignemont (eds.), *The Subject's Matter: Self-Consciousness and the Body*. Boston, USA: MIT Press. pp. 189-216.
- Bisiach, E., and Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex* *14*, 129–133.
- Bradley, A. (2019). The Feeling of Bodily Ownership. *Philosophy and Phenomenological Research*.
- Brédart, S. (2016). A self-reference effect on memory for people: We are particularly good at retrieving people named like us. *Frontiers in Psychology*, *7*, 1751.
- Cabeza, R., & Nyberg, L. (1997). Imaging cognition: An empirical review of PET studies with normal subjects.

- Journal of cognitive neuroscience*, 9(1), 1-26.
- Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, 377(6547).
- Chai, X. J., Whitfield-Gabrieli, S., Shinn, A. K., Gabrieli, J. D., Castanón, A. N., McCarthy, J. M., ... & Öngür, D. (2011). Abnormal medial prefrontal cortex resting-state connectivity in bipolar disorder and schizophrenia. *Neuropsychopharmacology*, 36(10), 2009-2017.
- Chun, M., Golomb, J., and Turk-Browne, N. B. (2011). A taxonomy of external and internal Attention. *Annu. Rev. Psychol.* 62, 73–101.
- Craik, F. I., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., & Kapur, S. (1999). In search of the self: A positron emission tomography study. *Psychological science*, 10(1), 26-34.
- Cunningham, S. J. (2016). The function of the self-attention network. *Cognitive neuroscience*, 7(1-4), 21-22.
- De Brigard, F. (2012). The role of attention in conscious recollection. *Frontiers in psychology*, 3, 29.
- De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155-185.
- De Brigard, F., Spreng, R. N., Mitchell, J. P., & Schacter, D. L. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. *Neuroimage*, 109, 12-26.
- De Vignemont, F. (2007). Habeas corpus: The sense of ownership of one's own body. *Mind & Language*, 22(4), 427-449.
- De Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness and cognition*, 20(1), 82-93.
- Dehaene, S., and Changeux, J. (2000). Reward-dependent learning in neuronal networks for planning and decision making. *Prog. Brain Res.* 126, 217–229.
- Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of cognitive Neuroscience*, 24(8), 1742-1752.
- Dugas, L., Sierra, M., & Berrios, G. E. (1996). A case of depersonalization. *History of Psychiatry*, 7(27), 455-461.
- Farmer, H., Cataldo, A., Adel, N., Wignall, E., Gallese, V., Deroy, O., ... & Ciaunica, A. (2020). The detached self: Investigating the effect of depersonalisation on self-bias in the visual remapping of touch. *Multisensory Research*, 1(aop), 1-22.
- Fernández, J. (2019). *Memory: A Self-Referential Account*. Oxford University Press, USA.
- Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J., & Heiss, W. D. (1996). Cerebral representation of one's own past: neural networks involved in autobiographical memory. *Journal of neuroscience*, 16(13), 4275-4282.
- Fossati, P., Hevenor, S. J., Lepage, M., Graham, S. J., Grady, C., Keightley, M. L., ... & Mayberg, H. (2004). Distributed self in episodic memory: neural correlates of successful retrieval of self-encoded positive and negative personality traits. *Neuroimage*, 22(4), 1596-1604.
- Frith, C. D., & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological medicine*, 19(2), 359-363.
- Gilbert, D. T., & Wilson, T. D. (2007). Propection: Experiencing the future. *Science*, 317(5843), 1351-1354.
- Gold, J. M., Randolph, C., Carpenter, C. J., Goldberg, T. E., & Weinberger, D. R. (1992). Forms of memory failure in schizophrenia. *Journal of abnormal psychology*, 101(3), 487.
- Goldberg, E., Antin, S. P., Bilder, R. M., Gerstman, L. J., Hughes, J. E., & Mattis, S. (1981). Retrograde amnesia: possible role of mesencephalic reticular activation in long-term memory. *Science*, 1392-1394.
- Hirstein, W., & Ramachandran, V. S. (1997). Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1380), 437-444.

- Humphreys, G. W., & Sui, J. (2016). Attentional control and the self: the Self-Attention Network (SAN). *Cognitive neuroscience*, 7(1-4), 5-17.
- Hunkin, N. M., Parkin, A. J., Bradley, V. A., Burrows, E. H., Aldrich, F. K., Jansari, A., & Burdon-Cooper, C. (1995). Focal retrograde amnesia following closed head injury: A case study and theoretical account. *Neuropsychologia*, 33(4), 509-523.
- James, W. (1890). The perception of reality. *Principles of psychology*, 2, 283-324.
- Jimenez, A. M., Lee, J., Wynn, J. K., & Green, M. F. (2018). The neural correlates of self-referential memory encoding and retrieval in schizophrenia. *Neuropsychologia*, 109, 19-27.
- Kapur, N., Ellison, D., Smith, M. P., McLellan, D. L., & Burrows, E. H. (1992). Focal retrograde amnesia following bilateral temporal lobe pathology: a neuropsychological and magnetic resonance study. *Brain*, 115(1), 73-85.
- Kapur, N., Scholey, K., Moore, E., Barker, S., Brice, J., Thompson, S., ... & Fleming, J. (1996). Long-term retention deficits in two cases of disproportionate retrograde amnesia. *Journal of Cognitive Neuroscience*, 8(5), 416-434.
- Klein, S. B. (2013). Making the case that episodic recollection is attributable to operations occurring at retrieval rather than to content stored in a dedicated subsystem of long-term memory. *Frontiers in behavioral neuroscience*, 7, 3.
- Klein, S. B. (2015). What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(1), 1-38.
- Klein, S. B. (2016a). Auto-noetic consciousness: Reconsidering the role of episodic memory in future-oriented self-projection. *The Quarterly Journal of Experimental Psychology*, 69(2), 381-401.
- Klein, S. B. (2016b). Lost feeling of ownership of one's mental states: the importance of situating patient RB's pathology in the context of contemporary theory and empiricism. *Philosophical Psychology*, 29(4), 490-493.
- Klein, S. B. (2019). The phenomenology of REM-sleep dreaming: The contributions of personal and perspectival ownership, subjective temporality, and episodic memory. *Psychology of Consciousness: Theory, Research, and Practice*, 6(1), 55.
- Klein, S. B., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind*, 121(483), 677-702.
- Kopelman, M. D. (1993). The neuropsychology of remote memory. *Handbook of neuropsychology*, 215-238.
- Kühn, S., & Gallinat, J. (2013). Resting-state brain activity in schizophrenia and major depression: a quantitative meta-analysis. *Schizophrenia bulletin*, 39(2), 358-365.
- Kurczek, J., Wechsler, E., Ahuja, S., Jensen, U., Cohen, N. J., Tranel, D., & Duff, M. (2015). Differential contributions of hippocampus and medial prefrontal cortex to self-projection and self-referential processing. *Neuropsychologia*, 73, 116-126.
- Lanius, R. A., Vermetten, E., Loewenstein, R. J., Brand, B., Schmahl, C., Bremner, J. D., & Spiegel, D. (2010). Emotion modulation in PTSD: Clinical and neurobiological evidence for a dissociative subtype. *American Journal of Psychiatry*, 167(6), 640-647.
- Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *science*, 303(5661), 1208-1210.
- Lemogne, C., Piolino, P., Friszer, S., Claret, A., Girault, N., Jouvent, R., ... & Fossati, P. (2006). Episodic autobiographical memory in depression: Specificity, auto-noetic consciousness, and self-perspective. *Consciousness and cognition*, 15(2), 258-268.
- Lemogne, C., Delaveau, P., Fretton, M., Guionnet, S., & Fossati, P. (2012). Medial prefrontal cortex and the self in major depression. *Journal of affective disorders*, 136(1-2), e1-e11.
- Levine, B., Black, S. E., Cabeza, R., Sinden, M., McIntosh, A. R., Toth, J. P., ... & Stuss, D. T. (1998). Episodic memory and the self in a case of isolated retrograde amnesia. *Brain: a journal of neurology*, 121(10),

- 1951-1973.
- Markowitsch, H. J., Fink, G. R., Thone, A., Kessler, J., & Heiss, W. D. (1997). A PET study of persistent psychogenic amnesia covering the whole life span. *Cognitive Neuropsychiatry*, 2(2), 135-158.
- Mattioli, F., Grassi, F., Perani, D., Cappa, S. F., Miozzo, A., & Fazio, F. (1996). Persistent post-traumatic retrograde amnesia: a neuropsychological and (18F) FDG PET study. *Cortex*, 32(1), 121-129.
- McLeod, H. J., Wood, N., & Brewin, C. R. (2006). Autobiographical memory deficits in schizophrenia. *Cognition and Emotion*, 20(3-4), 536-547.
- Medford, N., Brierley, B., Brammer, M., Bullmore, E. T., David, A. S., & Phillips, M. L. (2006). Emotional memory in depersonalization disorder: a functional MRI study. *Psychiatry Research: Neuroimaging*, 148(2-3), 93-102.
- Medford, N., Sierra, M., Stringaris, A., Giampietro, V., Brammer, M. J., & David, A. S. (2016). Emotional experience and awareness of self: Functional MRI studies of depersonalization disorder. *Frontiers in psychology*, 7, 432.
- Michaelian, K. (2011). Is memory a natural kind?. *Memory Studies*, 4(2), 170-189.
- Mier, D., Eisenacher, S., Rausch, F., Englisch, S., Gerchen, M. F., Zamoscik, V., ... & Kirsch, P. (2017). Aberrant activity and connectivity of the posterior superior temporal sulcus during social cognition in schizophrenia. *European archives of psychiatry and clinical neuroscience*, 267(7), 597-610.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56-60.
- Moroz, T. (1999). *Episodic memory for personally relevant information, evidence from aging, divided attention at retrieval and positron emission tomography* (Doctoral dissertation, National Library of Canada=Bibliothèque nationale du Canada).
- Nelson, B., Whitford, T. J., Lavoie, S., & Sass, L. A. (2014a). What are the neurocognitive correlates of basic self-disturbance in schizophrenia?: Integrating phenomenology and neurocognition. Part 1 (Source monitoring deficits). *Schizophrenia research*, 152(1), 12-19.
- Nelson, B., Whitford, T. J., Lavoie, S., & Sass, L. A. (2014b). What are the neurocognitive correlates of basic self-disturbance in schizophrenia?: Integrating phenomenology and neurocognition: Part 2 (Aberrant salience). *Schizophrenia research*, 152(1), 20-27.
- Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., and Mesulam, M. M. (2004). Orienting attention to locations in perceptual versus mental representations. *J. Cogn. Neurosci.* 16, 363-373.
- Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieval: The HERA model. *Psychonomic Bulletin & Review*, 3(2), 135-148.
- Pankow, A., Katthagen, T., Diner, S., Deserno, L., Boehme, R., Kathmann, N., ... & Schlagenhauf, F. (2016). Aberrant salience is related to dysfunctional self-referential processing in psychosis. *Schizophrenia bulletin*, 42(1), 67-76.
- Perrin, D., Michaelian, K., & Sant'Anna, A. (2020). The Phenomenology of Remembering Is an Epistemic Feeling. *Frontiers in psychology*, 11, 1531. <https://doi.org/10.3389/fpsyg.2020.01531>
- Phillips, M. L., & Sierra, M. (2003). Depersonalization disorder: a functional neuroanatomical perspective. *Stress*, 6(3), 157-165.
- Poulin, S., & Black, D. N. (2006). Neurobiology and Neuroimaging of Depersonalization, Conversion, and

- Dissociation. *Psychiatric Times*, 23(7), 47-47.
- Robins, S.K. (2017). Confabulation and constructive memory. *Synthese*, 1-17.
- Schmitz, T. W., & Johnson, S. C. (2006). Self-appraisal decisions evoke dissociated dorsal—ventral aMPFC networks. *Neuroimage*, 30(3), 1050-1058.
- Shin, J. E., Choi, S. H., Lee, H., Shin, Y. S., Jang, D. P., & Kim, J. J. (2015). Involvement of the dorsolateral prefrontal cortex and superior temporal sulcus in impaired social perception in schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 58, 81-88.
- Sierra, M. (2009). *Depersonalization: A new look at a neglected syndrome*. Cambridge University Press.
- Sierra, M., & Berrios, G. E. (1998). Depersonalization: neurobiological perspectives. *Biological psychiatry*, 44(9), 898-908.
- Sierra, M., & David, A. S. (2011). Depersonalization: a selective impairment of self-awareness. *Consciousness and cognition*, 20(1), 99-108.
- Simeon, D., Guralnik, O., Hazlett, E. A., Spiegel-Cohen, J., Hollander, E., & Buchsbaum, M. S. (2000). Feeling unreal: a PET study of depersonalization disorder. *American Journal of Psychiatry*, 157(11), 1782-1788.
- Simeon, D., & Abugel, J. (2006). *Feeling unreal: Depersonalization disorder and the loss of the self*. Oxford University Press, USA.
- Sui, J., & Humphreys, G. W. (2013). The boundaries of self-face perception: response time distributions, perceptual categories and decision weighting. *Visual Cognition*, 21, 415–445.
- Sui, J., Rotshtein, P., & Humphreys, G. W. (2013). Coupling social attention to the self forms a network for personal significance. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 7607–7612.
- Tulving, E. (1972). Episodic and semantic memory. *Organization of memory*, 1, 381-403.
- Tulving, E. (1982). Synergistic ephory in recall and recognition. *Can. J. Psychol.* 36, 130–147.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, 26(1), 1.
- Tulving, E. (1993). What is episodic memory?. *Current directions in psychological science*, 2(3), 67-70.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annual review of psychology*, 53(1), 1-25.
- Tulving, E. (2005). Episodic Memory and Autonoesis: Uniquely Human?. *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*, 1.
- Turk, D. J., Cunningham, S. J., & Macrae, C. N. (2008). Self-memory biases in explicit and incidental encoding of trait adjectives. *Consciousness and cognition*, 17(3), 1040-1045.
- Turk, D. J., Van Bussel, K., Brebner, J. L., Toma, A. S., Krigolson, O., & Handy, T. C. (2011). When “it” becomes “mine”: Attentional biases triggered by object ownership. *Journal of Cognitive Neuroscience*, 23(12), 3725-3733.
- Turk, D. J., Brady-Van den Bos, M., Collard, P., Gillespie-Smith, K., Conway, M. A., & Cunningham, S. J. (2013). Divided attention selectively impairs memory for self-relevant information. *Memory & cognition*, 41(4), 503-510.
- van den Bos, M., Cunningham, S. J., Conway, M. A., & Turk, D. J. (2010). Mine to remember: The impact of ownership on recollective experience. *Quarterly Journal of Experimental Psychology*, 63(6), 1065-1071.
- Wagner, D. D., Haxby, J. V., & Heatherton, T. F. (2012). The representation of self and person knowledge in the medial prefrontal cortex. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4), 451-470.
- Wood, N. L., & Cowan, N. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 255–260.

