Penultimate draft. Forthcoming in *Synthese*. Please do not cite without permission.
skrobins@ku.edu

Confabulation and Constructive Memory

Sarah Robins
University of Kansas

Abstract

Confabulation is a symptom central to many psychiatric diagnoses and can be severely debilitating to those who exhibit the symptom. Theorists, scientists, and clinicians have an understandable interest in the nature of confabulation—pursuing ways to define, identify, treat, and perhaps even prevent this memory disorder. Appeals to confabulation as a clinical symptom rely on an account of memory's function from which cases like the above can be contrasted. Accounting for confabulation is thus an important desideratum for any candidate theory of memory. Many contemporary memory theorists now endorse *Constructivism*, where memory is understood as a capacity for constructing plausible representations of past events (e.g., De Brigard, 2014; Michaelian, 2012, 2016). Constructivism's aim is to account for and normalize the prevalence of memory errors in everyday life. Errors are plausible constructions that, on a particular occasion have led to error. They are not, however, evidence of malfunction in the memory system. While Constructivism offers an uplifting repackaging of the memory errors to which we are all susceptible, it has troubling implications for appeals to confabulation in psychiatric diagnosis. By accommodating memory errors within our understanding of memory's function, Constructivism runs the risk of being unable to explain how confabulation errors are evidence of malfunction. After reviewing the literature on confabulation and Constructivism, respectively, I identify the tension between them and explore how different versions of Constructivism may respond. The paper concludes with a proposal for distinguishing between kinds of false memory—specifically, between misremembering and confabulation— that may provide a route to their reconciliation.

§1 Introduction

Psychiatric patients often confabulate. Take, for instance, Sergei Korsakoff's description of a person with his eponymous disorder:

> When asked to tell how he has been spending his time, the patient would very frequently relate a story altogether different from that which actually occurred, for example, he would tell that yesterday he took a bike ride into town, whereas in fact he has been in bed for two months, or he would tell of conversations which never occurred (Korsakofff, 1889/1955: 399).

Confabulations are false memories, malfunctions emblematic of Korsakoff's syndrome, Schizophrenia, Alzheimer's, and other forms of dementia and amnesia.[1] Confabulation

[1] This characterization of confabulation is meant only as a first pass at the term's definition. As I discuss in §3, there are many competing characterizations of confabulation, including

plays a key role in psychiatric nosology—aiding in the classification of distinct mental disorders to classify distinct mental disorders—and can be severely debilitating to those who exhibit the symptom. Theorists, scientists, and clinicians have an understandable interest in the nature of confabulation, pursuing ways to define, identify, treat, and perhaps even prevent this memory disorder.

Appeals to confabulation as a clinical symptom rely on an account of memory's function from which cases like the above can be contrasted. Characterizing confabulation as a symptom, reflective of disruption in the mental lives of those who exhibit it, implies a distinction from what remembering looks like in standard cases. If cases of confabulation are those where the ability to remember has gone wrong, then there must be some way in which they differ from cases where the ability to remember has not. The ability to account for clinical confabulation is thus an important constraint on the descriptive adequacy of any candidate theory of memory.

Contemporary theorizing about memory is dominated by *Constructivism*, where memory is understood as a capacity for constructing plausible representations of past events (e.g., De Brigard, 2014; Michaelian, 2012, 2016). Versions of the view are driven by the desire to account for memory errors, with a focus is on normalizing their prevalence in everyday life. Characterizing remembering as constructive makes the general tendency toward false memory understandable. While Constructivism offers an uplifting repackaging of the everyday memory errors to which we are all susceptible, doing so makes it difficult to say how such everyday and understandable errors differ from the confabulation errors appealed to in clinical diagnosis. By accommodating memory errors within our understanding of memory's function, Constructivism runs the risk of being unable to explain how confabulation errors are evidence of malfunction. After reviewing the literature on confabulation and Constructivism, respectively, I identify the tension between them and how different versions of Constructivism may respond. The paper concludes with a proposal for distinguishing between kinds of false memory—specifically, between misremembering and confabulation— that may provide a route to their reconciliation.

---

some broader definitions that capture non-memorial delusions of perception, intention, and emotion as well.

§2 Confabulation as a Clinical Symptom

Confabulation plays a central role in psychiatric theory and practice. In this section, I illustrate the importance of this symptom to psychiatric nosology, stressing along the way the debilitation that confabulations present to those who produce them.

Use of "confabulation" to describe symptomatic behavior of psychiatric patients began in the early 20th century.[2] These early reports were of patients suffering from a disorder that came to be known as Korsakoff's syndrome, a form of dementia produced by thiamine deficiency, often as the result of heavy alcohol consumption.[3] Amongst other symptoms—such as difficulty walking, irritability, and anxiety—patients with Korsakoff's are prone to false reports of their past experiences, for both recent and more distant events. Despite bearing little or no resemblance to their lived experience, these confabulations are characterized as "more or less coherent and internally consistent" (Talland, 1965: 49).

Patients with schizophrenia also produce confabulations. These false memories tend toward the incredible, as illustrated by Kraepelin (1919):

> A patient reported that he had dug up an amputated human arm, but then was compelled by his neighbor with a revolver in front of him to eternal silence; nevertheless he gave information and an inquiry was really made. Another went into a brothel in order to convince himself whether cannibals lived there. People were aiming at his life, but he escaped, though later human flesh was put before him in a restaurant (p. 310).[4]

Most confabulators continue to repeat these reports, in some cases verbatim, over extended periods of time—for weeks, months, and even years. Some embellish these "memories," adding further details with each retelling, whereas other accounts winnow over time. Regardless, patients continue to believe in these experiences (McKenna, Lorente-Rovira, & Berrios, 2009). Since Kraepelin's demarcation of schizophrenia as a distinct psychiatric condition, our understanding of the nature of this disorder has changed

[2] Vernacular use of *confabulation* dates back further—see Berrios (1998).
[3] Korsakoff himself described these as "pseudo-memories." Wernicke (1906) was the first to label them confabulations.
[4] Kraepelin originally characterized these patients as suffering from a distinct disorder, *paraphrenia*, but are now considered a form of schizophrenia (Schnider, 2008).

and is, in fact, still in flux.[5] Nonetheless, theorists and clinicians continue to view confabulation as a central symptom, one that is possibly even pathognomonic (or, as philosophers might put it, necessary and sufficient for the presence the disorder, see Buchanan, 1991).

Not all clinical cases of confabulation involve false reports that are fantastical. Following an *aneurysm of the anterior communicating artery* (ACoA), many patients produce confabulations that are, in comparison to the above, mundane. Johnson and colleagues (1997), for example, describe a patient who claimed that his ACoA-producing injury occurred because he fell and hit his head while standing outside talking to a friend, when in actuality the rupture occurred indoors during a fight with his daughter (p. 192). Similarly, when asked what he had eaten for lunch, another ACoA patient replied, "a couple of sandwiches, potato chips, and an apple" (Demery, Hanlon, & Bauer, 2008: 296). The response was fitting, but incorrect.

Confabulation is also a common symptom of various forms of dementia, which are classified as *major neurocognitive disorders* in the most recent version of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM–V). Patients with Alzheimer's Disease, the most common form of dementia, often produce confabulations that confuse the temporal order of events (Talberg & Almkvist, 2001). A host of other brain diseases and cognitive disturbances also lead to confabulation, including subarachnoid hemorrhage or encephalitis, Binswanger's encephalopathy, and amnesia resulting from other forms of traumatic brain injury (Dalla Barba, 1993).

Across disorders, these confabulations share many features. First, patients offer them as genuine memory reports, without intent to deceive.[6] Moscovitch (1989) describes confabulation as a form of "honest lying." Second, not all patients produce confabulations and even those who do will not produce them in all situations. Korsakoff's patients, for

[5] Many now believe that schizophrenia lumps together several genetically distinct disorders (see Arnedo et al., 2014).

[6] It is, of course, possible that patients intend some form of deception—to convince the clinician of their improvement, or to conceal the depths of their illness (see Coltheart & Turner (2009) for a discussion of this issue). I do not mean to discount this possibility, only to emphasize what is readily apparent in most cases (i.e., that most patients believe that these reports are experiences from their own past).

example, will often freely admit to not knowing answers to other, impersonal questions (Schnider, 2008). Confabulation also cannot be characterized as an automatic response to a memory deficit. For many of these disorders, it is a common, but not universal symptom. And, importantly, there are other disorders that involve memory loss but not confabulation—for instance, dissociative fugue states (DSM–V, 300.13, F44. 1).

One of confabulation's most startling features is its perseverance. Some patients persist in their confabulations, often even in the face of interrogation or evidence to the contrary.[7] To illustrate the depths of this commitment, it is worth quoting some physician-patient interactions directly. Here, for instance, is DeLuca's (2001) report of a conversation with an ACoA patient:

> *Doctor*: You indicated last night you were working on a number of projects at home…what would you say if I told you you were actually in the hospital last night
> *Patient*: I'd be surprised, because my experience, what I learn from my eyes and ears tells me differently…I'd want some evidence. I'd want some indication that you knew about my private world before I gave any cognizance.
> *Doctor*: Would you believe me?
> *Patient*: Not out of the blue, especially since we haven't met [an illustration of the patient's amnesia].
> *Doctor*: What if your wife was here and she agreed with me, what would you think at that point?
> *Patient*: I'd continue to resist but it would become more difficult.

Moscovitch offers an even more striking case, featuring a patient with dementia resulting from damage to the frontal cortex:

> *Psychologist*: How long have you been married?
> *Patient*: About 4 months
> …
> *Psychologist*: How many children do you have?
> *Patient*: Four. [Laughs.] Not bad for four months.
> …
> *Psychologist*: How old are your children?

[7] For this reason, Langdon and Turner (2010) recommend a distinction between the initial adoption of confabulated content and the maintenance of belief in this confabulated content over time. Coltheart, Menzies, & Sutton (2010) argue that confabulations need not be considered delusional until the latter justificatory confabulations are produced. The cases quoted in the text exemplify such persistent confabulations.

> *Patient*: The eldest is 32; his name is Bob. And the youngest is 22. His name is Joe.
> *Psychologist*: How did you get these children in four months?
> *Patient*: They're adopted.
> …
> *Psychologist*: Immediately after you got married you adopted these four older children?
> *Patient*: Before we were married we adopted one of them, two of them. The eldest girl Brenda and Bob, and Joe and Dina since we were married.
> *Psychologist*: Does it all sound a little strange to you, what you are saying?
> *Patient*: I think it is a little strange.
> *Psychologist*: I think when I looked at your record it said you've been married for over 30 years. Does that sound more reasonable to you if I told you that?
> *Patient*: No.
> *Psychologist*: Do you really believe that you have been married for 4 months?
> *Patient*: Yes.
>
> (1989: 135–136).

The aim of this section has been to demonstrate the nature of confabulation, as it is displayed across persons with a range of psychiatric diagnoses. Given confabulation's puzzling features, it is not surprising that this symptom has captured the interest of psychiatric clinicians and theorists. Confabulations are wide-ranging—these false reports span the recent and distant past, describing alleged events that stretch from the everyday to the incredible. Across variations in content, these confabulations share many core features: they are offered in earnest, generated without prompting or encouragement, and maintained in light of contravening evidence.

§3 Defining Confabulation

Exploring the connection between confabulation and theories of memory may seem premature. There is at present no agreed upon definition of confabulation. Shouldn't reconciliation be postponed until the phenomenon is better understood? In what follows, I argue that this concern can be set aside. All proposed definitions of confabulation include false memories and characterize confabulations as malfunctions. Regardless of which definition wins the day, the cases described in §2 will fall within its extension.

Proposed definitions of confabulation differ not only in content, but in scope. Some focus on cases like those illustrated above, using three features to characterize this symptom: confabulations are 1) false 2) memory 3) reports. Berrios (1998), for example,

defines confabulation as "inaccurate or false narratives…issued by subjects intent on 'covering up' for a putative memory deficit" (p. 225). Such definitions are often labeled as *narrow* because they focus on memory exclusively (Bortolotti & Cox, 2009). Amongst narrow definitions, a further distinction can be made between those that restrict confabulations to pathological distortions that occur in clinical contexts (e.g., Fotopoulou et al., 2008) and those that count all false memory reports, even in non-clinical cases, as confabulations. The latter is most associated with the work of Elizabeth Loftus. Loftus and colleagues have shown that fabricated reports of past experiences can arise in the absence of any psychiatric disorder. Specifically, they have shown that, in response to mildly suggestive questioning, normal adult participants can be led to create rich, elaborate "memories" for events that never occurred, such as being lost in a shopping mall as a child or spilling punch at a wedding (e.g., Loftus & Pickrell, 1995). Such errors are often described as "a kind of confabulation in non-clinical subjects" (French, Garry, & Loftus, 2009).

Others reject the narrow definition of confabulation and propose instead expansion, in recognition of the similarity between confabulation and delusional states. Patients with anosognosia—a lack of awareness of their illness—are often described as confabulating. Anton's Syndrome patients suffer from cortical blindness, but retain the belief that they can see and so will often explain away behavioral missteps that result from their lack of vision as the result of poor lighting or some other environmental feature (Swartz & Brust, 1984). Similarly, many patients experience paralysis following a stroke, but are unaware of the paralysis and will confabulate reasons for their lack of movement. As Ramachandran (1996) describes, one patient claimed: "these medical students have been probing me all day and I'm sick of it. I don't want to use my left arm" (p. 125). Other theorists consider delusional beliefs to be a non-memorial form of confabulation (Coltheart & Turner, 2009), as when patients with Capgras delusion claim that a close loved one has been replaced with an imposter (Ellis et al., 1994) or patients with asomatognosia claim that their paralyzed limbs belong to someone else (Feinberg et al., 2010). And still others press the term into even broader service, describing the post hoc rationalizations offered to explain decision-making (Nisbett & Wilson, 1977) and moral judgment (Haidt, 2001) as confabulatory.

In an effort to reflect its expanded use, many now favor a wide, *epistemic* approach to defining "confabulation," aiming to capture the shared features of ill-grounded beliefs that result from memory, perception, intention, emotion, and the like. Hirstein's (2005) offers the most extensive philosophical account:

> Jan confabulates that p if and only if
>
> (1) Jan claims that p.
> (2) Jan believes that p.
> (3) Jan's thought that p is ill-grounded.
> (4) Jan does not know that her thought is ill-grounded.
> (5) Jan should know that her thought is ill-grounded.
> (6) Jan is confident that p. (2005: 187).[8]

The hunt for a definition of confabulation is intriguing, raising as it does questions about the nature of rationality, self-knowledge, and other treasured philosophical concepts. These ongoing debates need not be viewed as reason to delay exploration of the interface between confabulation and theories of memory, however. Regardless of whether one's favored definition is narrow or wide, false memory reports will be included. Terminological disputes are waged at the periphery; cases illustrated in §2 are central and so covered by definitions both narrow and wide. To my knowledge, no candidate definition has proposed their exclusion. There may be debates as to whether to include delusions *in addition to* false memories, but not debates over the inclusion of false memories.

The boundary between clinical and non-clinical cases of confabulation is also unimportant for my purposes here. All proposed definitions share a common, evaluative claim: confabulation is a malfunction. It is an error, disorder, or disruption—the result of an agent, cognitive system, or neurological mechanism performing other than it should. This is illustrated in condition (5) of Hirstein's (2005) analysis, quoted above, where the agent is held accountable for lacking knowledge of her thought's poor grounding. In a later defense of his view, Hirstein makes the point even more forcefully:

---

[8] Hirstein's account does not lack for critics. Bortolotti and Cox (2009) challenge Hirstein for his failure to acknowledge that confabulation can confer some benefits upon the confabulator expressing them. Bortolotti continues, however, to view confabulations as malfunctions, arguing elsewhere that any account of them must be capable of distinguishing between pathological and non-pathological cases (see Bortolotti, 2011).

> If the confabulator's brain were functioning properly, she would know that the claim is ill-grounded and not make it (2009: 652).

The portrayal of confabulation as a malfunction can also be seen in accounts of confabulation pitched at a lower level. Coltheart and colleagues, for example, prefer a neurocognitive account, characterizing confabulation and other delusions at the level of brain mechanisms and cognitive systems. Specifically, they endorse a 2-factor model of confabulation: first, the mechanism for retrieving information breaks down; second, the process for evaluating possible contents is impaired (Coltheart, Langdon, & McKay, 2011). Even though these accounts differ in their preferred level of explanation, they share commitment to the idea that confabulation is best illustrated in contrast to an account of the mechanism's normal functioning.

The insistence on characterizing confabulation as malfunction reflects the severity of the disruption that confabulations cause in the lives of those who exhibit this symptom. We owe it to those who struggle with these conditions to continue the pursuit of ways to identify, treat, and prevent confabulation and the disorders from which it arises. Doing so requires, first, an account of how memory functions under normal conditions. A sense of how the system should work can then guide our understanding of its failure, limitation, or malfunction in cases of psychiatric disorder. Accounting for confabulation is thus an important constraint on any candidate theory of memory, one that should receive attention in current debates over whether and how to rethink the nature of memory. With this understanding of confabulation and its significance now to hand, I turn to Memory Constructivism.

§4 Memory Constructivism

Views about the nature of memory and ideas about which kinds memory errors are possible are deeply interconnected. Traditionally, memory has been characterized as a preservative capacity, which has in turn directed attention toward errors of *omission* rather than errors of *commission*—that is, toward forgetting rather than false memory. In contemporary theorizing about memory, however, the influence runs in the opposite direction. Accumulating evidence that false memories are commonplace in everyday

remembering has led many to urge a rethinking the nature and function of memory. I label the views that result from this rethinking *Constructivist* (e.g., De Brigard, 2014; Michaelian, 2012, 2016). Constructivists argue that memory is a capacity for building (i.e., constructing) representations of past events from a generalized network of information. Precisely which features of this general characterization are emphasized, and how, differs across the view's proponents.[9] In this section, I first review the problems Constructivists see with traditional, preservative accounts of memory and then outline the evidence that motivates Constructivism and its refashioned approach to memory and memory errors.

Constructivists identify two problems with the traditional, preservative focus on forgetting errors. First, traditional accounts portray forgetting negatively; forgetting errors are seen as instances of cognitive vice. Failing to retain information is in conflict with memory's aim at preservation. As Constructivists like Michaelian (2011) have noted, this is at best an oversimplification. Forgetting is often a productive way of streamlining memory's contents, making the most critical information available more efficiently and effectively. Constructivists' second concern is with neglect of false memories. Memory science has shown, repeatedly, that errors of commission are common. Representations produced in remembering often contain information from many sources, even if they feel to the rememberer as if they depict a particular past event (Brainerd & Reyna, 2005). False memories are readily observed in both laboratory and naturalistic settings. Loftus' misinformation paradigm is one of the prominent experimental techniques for eliciting false memories. In this paradigm, participants witness an event and are then quizzed on its details. Use of this paradigm shows that—despite participants' confidence in their memories—their reports contain many errors. After viewing a video of a car accident, for example, a quarter of participants "remembered" the roadway scene as involving a stop sign rather than a yield sign (Loftus, Miller, & Burns, 1978). When the prompt involved misleading information, false reports increased—to 60%. In such cases, the substitutions are relatively minor. But in others they errors are more substantial; participants have been shown to misidentify central actors and confuse the order of critical details, even when asked within minutes of their witnessing the event (e.g., Christianson & Loftus, 1987).

[9] For a discussion of the differences between accounts of philosophical Constructivism, see Robins (2016).

The effects of misinformation are also evident in people's memories of their own experiences, even those thought to be significant and indelible. Studies of these "flashbulb memories" reveal that the details presented in these recollections change over time. For example, a person might initially report having heard of the event while at work, only later to recall first seeing the event reported on television.[10] People are often confident in their reports of details that were not part of their previous experience, and in some cases retain this confidence in light of evidence that tells against the accuracy of their retelling (Paradis, Solomon, Florer, & Thompson, 2004).

Constructivists argue that the accumulation of such errors presents a serious challenge to the traditional view of memory, placing pressure on its preservative portrayal. It is strange to characterize memory as a capacity designed for preservation if it rarely manages to keep memories of past events intact. As De Brigard explains, "saying that false and distorted memories are a failure of memory may force us to accept that we have a memory system that regularly and systematically malfunctions" (2014: 159). A better option, the Constructivist claims, is to think that we have misunderstood memory's function. If we instead view memory errors as the primary explanandum—asking what purpose these false and distorted representations of the past might serve—then what emerges is an altogether different account.

So what is this alternative account of memory's function? The characterization is usually contrastive, selecting a feature of the traditional preservationist picture against which the Constructivist alternative can be understood. The general Constructivist approach is to situate remembering within a broader cognitive system, as one use of an episodic (i.e. self-directed) capacity amongst many. Below I discuss the two most prominent versions: De Brigard's (2014) *Episodic Hypothetical Thinking* account, and Michaelian's (2012; 2016) *Simulation Theory*.

De Brigard (2014) advocates for a change in the goal of memory processing, to *plausibility* rather than *accuracy*. This change in function fits memory's place in a larger cognitive system devoted to episodic hypothetical thinking—a system for "self-referential

[10] Even Presidents are not immune to such errors (Greenberg, 2004).

mental simulations about what happened, may happen, and could have happened to one self" (2014: 174–175). He argues that the memory system is best viewed as producing optimal reconstructions of what was likely to have occurred during a past event, rather than faithful reproductions of what actually happened during a past event.

Characterizing remembering in this way helps to make memory errors understandable. First, forgetting becomes an important part of effective remembering. It is by forgetting that the system develops the expertise required to produce optimal reconstructions. Second, false memories are an expected outcome of this constructive process. Information is stored in ways that highlight similarities across events; construction privileges patterns and themes. This is done in service to one's anticipated uses of memory—information encountered frequently is likely to be of continued relevance and rarely does such re-use require elaborate detail. And so remembering will tend to be successful when one is trying to recall information that frequently recurs. Similarly, false memories will occur when one is trying to recall less frequent events or details. The system will tend to replace these details with others that are more common or likely, as when Loftus' participants swapped the less expected yield sign for the more common stop sign.[11] This is, of course, an error, but an understandable one. It is not a deep mistake, indicative of problems with the remembering process. Instead, it is the outcome of a generally effective process in less than optimal circumstances. Remembering was unsuccessful not because something within the system went wrong, but because the information requested was not what the system was designed to favor. Despite being an error, it reflects the system working well. Much as visual illusions offer a window into the underlying structures that support the act of seeing, so too memory errors serve as mnemonic illusions, revealing the underlying, constructive nature of remembering (Roediger, 1996). De Brigard characterizes his revised view of memory's function in this way:

> Most of the time what you recall accurately depicts the witnessed event. Sometimes it does not. In both cases, however, the system is doing what it is supposed to do (2014: 172).

[11] See De Brigard for a more detailed Constructivist explanation of Loftus' misinformation effect (2014: 172). For a critique of De Brigard's frequentist interpretation of memory errors, see Robins (2016).

Michaelian's (2012; 2016) *Simulation Theory* presents a similar but distinct version of Constructivism. There are two key differences between De Brigard and Michaelian's versions of Constructivism. First, they offer distinct portrayals of the constructive process. Michaelian characterizes the broader cognitive system as one dedicated to producing episodic imagination—i.e., the simulation of possible episodes. The only difference between remembering and other forms of episodic imagination is the aim of the simulation process. In remembering, the aim is to produce an episode from the rememberer's past (2016: 105). Second, they differ over whether the constructive system is characterized as aiming at and maintaining accuracy in its constructions. While De Brigard says no, Michaelian says yes. These episodic simulations, whether intended for imagining a possible future or remembering the personal past, will function at their best when the cognitive system is streamlined. And so, Michaelian's view helps to make sense of why forgetting occurs and why it may be best construed as an epistemic virtue:

> Given limited computational resources, forgetting is necessary to enable the system to achieve the balance of reliability, power, and speed appropriate for it given its function; given the necessity of sensitivity to interests for virtue, virtuous memory involves preferentially forgetting uninteresting records (Michaelian, 2011: 423).

The Simulation Theory also makes sense of why memory errors occur. Michaelian argues that most of the additional content utilized in the construction of episodic remembering serves as information rather than *mis*information. Nonetheless, a system that aims at simulating episodes from one's personal past may occasional fail to do so; incorporating misinformation into one's simulation can lead to error.

§5 The Tension between Confabulation and Constructivism

Constructivists propose a change to the traditional view of memory. The shift is in many respects welcome, as it helps to make sense of otherwise troubling evidence of widespread memory errors in everyday life. But this change creates problems for our understanding of other psychological phenomena that relied on or presupposed the traditional account of memory. In this section, I explore the tension between the characterization of

confabulation as a clinical symptom and the approach to memory errors favored by Constructivists.

For Constructivists, the primary aim is the normalization of everyday memory errors, including forgetting and most especially false memory. Despite differences in *how* the constructive process is understood, there is general agreement that this process is at work in cases of successful and unsuccessful remembering alike. The goal of Constructivism is to explain how memory errors can be compatible with our understanding of memory's function. The tension comes when we consider the place of clinical confabulations in this project. Are they evidence of memory's function or malfunction? Viewed from the perspective of psychiatric theory and practice—with an eye toward the disruptive role of confabulations for those receiving diagnosis—false memories are malfunctions. But viewed from the perspective of theorizing about memory, memory errors like confabulations are a sign of standard, or perhaps even well functioning memory. Constructivists have altered our conception of memory's function, removing along with it the account of memory against which clinical confabulations could be contrasted and understood. In the attempt to make sense of certain memory errors, Constructivists have crafted an account of memory that downplays the significance of others. Confabulation errors, those that were the focus of the opening sections of this paper, are memory errors that resist normalization. The question is: does Constructivism offer a way to distinguish between memory errors that are compatible with memory's function and memory errors that are evidence of memory's malfunction?

Before turning to investigate how specific versions of Constructivism might approach this question, it helps to clarify what this problem is *not*. It is not a worry that the boundary between everyday and clinical memory distortions has been blurred. The possibility of such a continuum has long been acknowledged, and some happily endorse broad, epistemic accounts of confabulation, as discussed in §3. The worry here is distinct. The concern is that Constructivism may lack the ability to distinguish between memory errors that are functional and those that are not. Which memory errors are to be normalized, which are evidence of malfunction, and how do we tell the difference? The function/malfunction division may correspond to the boundary between everyday and clinical errors, but it need not.

Versions of Constructivism, as outlined in §4, will offer distinct responses to this question. De Brigard's (2014) Episodic Hypothetical Thinking view collapses the processing distinction between memory errors and successful remembering. All are instances of a general constructive process. It may still be possible to distinguish them, at least in cases where external confirmation of what happened during the alleged event is available. But there is no way to tell the difference simply by observing the process by which the constructions are produced. In fact, not only are false memories not instances of malfunction, they are evidence that memory is functioning *well*. Susceptibility to false memories may be indicative of other cognitive virtues of skills. In support of this claim, De Brigard cites correlations between the production of false memory errors and measures of creativity (Howe et al., 2011) and convergent thinking (Dewhurst, Thorley, Hammond, & Ornerod, 2011). Viewed in this light, memory errors are not only understandable, but beneficial:

> Assuming that an increase in problem solving abilities, such as those tapped by these insight-based and convergent thinking tasks, confers cognitive organisms like ourselves an advantage in our current environment, then some tendency toward misremembering may prove advantageous rather than detrimental (De Brigard, 2014: 165).

On this version of Constructivism, confabulation is not a malfunction. The report offered in the confabulation may be inaccurate, but nothing within the memory system has gone wrong. In some sense, all attempts at remembering are confabulations. They are constructions of a system geared toward tracking general patterns.

There may still be a way for De Brigard's Constructivism to enforce a distinction. The healthy confabulations of everyday remembering and the pathological confabulations that occur in clinical cases may not be distinguished by appeal to the underlying cognitive process, but they may still be distinguishable in terms of their content. Everyday constructions are reasonable; they exploit patterns in past events and use this to guide future inferences. This is less clear for clinical confabulations. As shown in §2, these confabulations describe unlikely, implausible events like body farms and cannibalism, persecution by doctors, and the belief that newlyweds would adopt several adult children. Even if the boundary between clinical and non-clinical constructions is not sharp, it may

still be clear: constructions illustrative of confabulation are strange, abnormal, bizarre. We know them when see them.

Such an amendment would offer little help to the philosophers and scientists engaged in questions of psychiatric nosology. First, the appeal to content would not catch all clinical cases, some of which are mundane. Consider the patient, described in §2, who reported (falsely) eating sandwiches, chips, and an apple for lunch. Second, and more importantly, relying on judgments about content would conflict with guidelines of psychiatric nosology. Modern psychiatry is, for good reason, reluctant to let intuitions about abnormality guide taxonomy. A brief review of psychiatry's history serves as a sufficient, and uncomfortable, reminder as to why. In 19th century America, slaves attempting to flee captivity were diagnosed with *Drapetomania*—a disorder characterized by insufficient concern for the property rights of one's owner. Homosexuality was listed as a disorder by the American Psychiatric Association until 1973.[12] Deciding which constructed memories are confabulations by evaluating their content would invite a return to the unsavory practice of pathologizing deviance. What is disapproved of is not thereby disordered.

Indeed, most agree that the only way for psychiatric nosology to avoid such value-laden pitfalls is to embrace a two-stage model of psychiatric classification:

> The first project is what determines that someone has a frontal lobe lesion, a depressive cognition, a genetic susceptibility to anxiety or a serotonin imbalance. The second project asks if human beings can flourish if they have such physical or psychological abnormalities (Murphy, 2006: p. 19).

On such an account, a psychiatric diagnosis requires, first, the identification of some way in which the cognitive system or neural mechanism has deviated from standard functioning. It is only *after* such identification that the evaluative project can begin—i.e., asking whether the departure presents a severe enough challenge to our ideal of *well* functioning to merit labeling it a disorder. Murphy goes on to argue, quite forcibly, that the two-stage approach will work best if the first project is closely aligned with our best cognitive science. The two-stage approach to psychiatric nosology appears to make retaining confabulation as a

[12] Murphy (2006) offers a comprehensive discussion of these and other cases psychiatry's missteps.

clinical symptom impossible on De Brigard's account of Constructivism. The first stage of the process cannot be completed. When patients offer false reports of their past experiences, there is no cognitive process or neural mechanism that is behaving abnormally. As De Brigard states (quoted above in §4) the system is "doing what it is supposed to do" in both cases. And so there is no justification for asking the further question of whether this malfunction represents a disorder worthy of treatment.

If De Brigard's form of Constructivism reflects our best cognitive science of memory, then we may be forced to accept this conclusion. But the disorders in which confabulations feature disrupt patients' severely enough to caution us against conceding the point prematurely. At the very least, we owe it to those who experience this symptom, and those devoted to its treatment and prevention, to search for an alternative.

Michaelian's (2016) *Simulation Theory* offers a different approach, one that allows for a distinction between memory errors consistent with the episodic system's functioning and memory errors indicative of malfunction. It may thus be better suited to accommodating clinical confabulations. Michaelian draws the distinction in terms of the system's tendency toward accuracy. As he explains:

> In a healthy subject, the system recombines information, whether or not it originates in the target episode, following procedures designed to enable it to produce a representation of the episode which is (within certain limits) accurate. In the confabulating subject, in contrast, the system malfunctions, following procedures which tend to produce inaccurate representations (2016: 109).

Michaelian goes on to explain that memory's tendency toward accuracy is sustained in two ways. First, the simulation process is governed by heuristics that favor information over misinformation. Second, the simulation process is overseen by a source monitoring system whose job is to reject simulations that have gone too far afield. These mechanisms ensure that the system is by and large reliable, but the mechanisms are not infallible. Their operation is consistent with the occasional memory error. It is when memory errors become more frequent—when they become the rule rather than the exception—that the system changes from functioning to malfunctioning. Michaelian's account thus allows us to say that the memory errors that occur in everyday cases are consistent with memory's function because they are outnumbered by cases where remembering is reliable. Clinical

confabulations, on the other hand, are malfunctions because these errors are a more common result of attempts at remembering.

Michaelian's approach distinguishes everyday memory errors from confabulations by the frequency with which they occur. We can now pause to ask whether this account of the difference accords with the felt difference between many everyday memory errors and cases of clinical confabulation. Errors may be more common for clinical patients, or it may be only that the errors produced are more noticeable or that reports from such patients are met with more skepticism than everyday attempts at remembering. Determining how many attempted rememberings are errors, in either everyday or clinical cases, is difficult outside of controlled experimental conditions.

I want to close by suggesting an alternative approach to distinguishing between everyday memory errors and clinical confabulations—one that focuses on the *type* of error made, rather than the *frequency* of error. It is standard practice for the terms *memory error*, *misremembering*, *false memory*, and *confabulation* to be used interchangeably. Little attention has been paid to ways of distinguishing amongst them. Given the range and extent of errors observed, it seems possible, even likely, that they could come in multiple, distinct forms. If a difference can be found between cases of clinical confabulation and everyday false memories, then the symptom and the theory would no longer be in conflict.

A good place to look for such a difference may be perceptual illusions, those to which false memories are often compared already (Roediger, 1996). There are two general types of perceptual error: illusion and hallucination.[13] Illusions involve perceiving an object as having properties that it does not—seeing, for instance, a barren tree as possessing leaves and fruit. Hallucination is a more extensive error; it occurs when the entire perceptual experience (both the tree *and* its features) is illusive. It may be possible to draw a similar distinction between memory errors. Elsewhere I have drawn a distinction between *misremembering* and *confabulation* (Robins, 2016). Misremembering errors are those that result from distortions of retained information. They are, in this way, comparable to perceptual illusions. Confabulation errors, on the other hand, are wholly

[13] These remarks are intended as an appeal to a widely accepted way of characterizing the distinction between these two forms of perceptual error, not as an endorsement of any particular theory of perception.

inaccurate, reflecting no influence of information retained from a particular past event. Confabulations are thus akin to perceptual hallucinations.

Invoking a distinction between misremembering and confabulation could help in two ways. First, it would reduce the overlap between the memory errors that occur in clinical cases and the memory errors that drive Constructivist theorizing. Most of the memory errors discussed in §2 are confabulations; most of the memory errors discussed in §4 are misrememberings. The distinctions may not line up perfectly. Some cases of everyday memory error may be confabulations and some clinical memory errors may only be misrememberings. For example, the case of Alzheimer's patients confusing the temporal order of events (as reported in §2) looks to be a case of misremembering.[14] The difference can be seen by considering different experimental manipulations performed by Loftus, which offer evidence of each type of error. Loftus' suggestibility studies, introduced in §2, show evidence of participants producing entirely false memories—representations of events that never occurred. They, and the other confabulation reports discussed in this section, are mnemonic hallucinations. The rememberer was never lost in the mall as a child or served human flesh at a restaurant. In contrast, the results of Loftus' misinformation paradigm, from §4, involve misremembering. The memories produced are not wholly inaccurate. Instead, participants remember the past event as having details that it did not—as involving a stop sign rather than a yield sign.

As further support for the distinction between misremembering and confabulation, it is worth noting that confabulations are somewhat difficult to produce in non-clinical subjects. Loftus' suggestibility studies work best for events that are easy for participants to imagine—being lost in a mall or spilling punch at a wedding. Most participants cannot be led to produce confabulations of unlikely events, like being abducted by aliens (Clancy et al., 2002). Attempts to produce confabulations in non-clinical subjects rely on an experimental paradigm known as the *forced fabrication technique* (Chrobak & Zaragoza, 2009). This technique involves coaching participants to invent detailed accounts of imagined events and then, in sessions held weeks later, asking them to discuss these imagined events. Most participants resist speculating about events they do not remember,

[14] Thanks to an anonymous reviewer for pointing out this case as an example of misremembering.

but their hesitation can be overridden in certain conditions. As Chrobak and Zaragoza note, studying how self and uncertainty monitoring mechanisms can be swamped in a laboratory context may provide some insight into the mechanisms that are malfunctioning in clinical cases of confabulation.

The second advantage of the misremembering/confabulation distinction is that it would make it possible to retain a key Constructivist insight: namely, that memory's constructive nature can be advantageous in some circumstances. Misremembering may be advantageous in a number of circumstances. Keeping track of only an event's general features may make memory more efficient and useful than it would be otherwise. And modifying one's account of past events over time may yield benefits for mental health and social cohesion. It is difficult to construe confabulations as similarly useful.[15] Confabulations are not simply false memory reports, but reports that lack any substantive contact with information retained from a particular past event. To confabulate is to claim memory of a past experience that is false in its entirety, not only in detail. Memory's function may be much as Constructivists' suppose: it may be more sensitive to future constraints than to those of the past. Still, memory does not perform well when it abandons the past entirely. As with Michaelian's proposed distinction in terms of the frequency of errors, it may be difficult to discern whether a particular memory report is accurate. But at least in experimental contexts, where the encoding and retrieval conditions can be controlled, the amount of error can be assessed.

The remarks here are intended only as initial speculation. More research into these memory errors is required in order to determine whether the difference between misremembering and confabulation can be maintained. Such investigation may reveal that the distinction between misremembering and confabulation is blurry; these errors may be better understood as ends of a continuum. Or it may be that the distinction only helps to sort between misrememberings and the more implausible confabulations. But the identification of any distance between misremembering and confabulation is likely to be

[15] Confabulations do not appear to be useful to the act of remembering, but they may of course be of broader use to the person—e.g., by promoting self-esteem (e.g., Fotopoulou 2010). Thanks to an anonymous reviewer for raising this point.

useful for limiting, if not entirely removing, the tension between clinical confabulation and Memory Constructivism.

Acknowledgments:
I am grateful to attendees of the *Early Career Scholars Conference in Philosophy of Psychiatry* at the University of Pittsburgh and the audience at a University of Kansas colloquium for helpful comments on previous drafts of this paper. Special thanks to Serife Tekin and Kourken Michaelian for their feedback and conversations about confabulation.

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* Washington, DC: American Psychiatric Association.

Arnedo, J., Svrakic, D.M., del Val, C., Romero-Zaliz, R. Hernandez-Cuervo, H., Fanous, A.H., et al. (2014) Uncovering the hidden risk architecture of the schizophrenias: Confirmation in three independent genome-wide association studies. *American Journal of Psychiatry*, *172*. Published online.

Berrios, G.E. (1998). Confabulations: A conceptual history. *Journal of the History of the Neurosciences, 7,* 225-241.

Bortolotti, L. (2011). Psychiatric classification and diagnosis: Delusions and confabulations. *Paradigmi, 1*, 99-112.

Bortolotti, L., & Cox, R. (2009). Faultless ignorance: strengths and limitations of epistemic definitions of confabulation. *Consciousness and Cognition, 18*, 952-965.

Brainerd, C. J., & Reyna, V. F. (2005). *The Science of False Memory.* Oxford, UK: Oxford University Press.

Buchanan, A. (1991). Delusional memories: First-rank symptoms? *British Journal of Psychiatry, 159,* 472—474.

Christianson, S., & Loftus, E.F. (1987). Memory for traumatic events. *Applied Cognitive Psychology*, 1*, 225-239.*

Chrobak, Q.M., & Zaragoza, M.S. (2008). The cognitive consequences of forced fabrication: Evidence from studies of eyewitness suggestibility. In W. Hirstein (Ed.) *Confabulation: Views from neuroscience, psychiatry, psychology, and philosophy*. Oxford: Oxford University Press (pp. 67—90).

Clancy, S.A., McNally, R.J., Schacter, D.L. ,Lenzenweger, M.F., & Pittman, R.K. (2002). Memory distortions in people reporting abduction by aliens. *Journal of Abnormal Psychology, 111*, 455-461.

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology, 62*, 271-298.

Coltheart, M., Menzies, P., & Sutton, J. Abductive inference and delusional belief. *Cognitive Neuropsychiatry, 15*, 261–287.

Coltheart, M. & Turner, M. (2009). Confabulation and delusion. In W. Hirstein (Ed.) *Confabulation: views from neuroscience, psychiatry, psychology, and philosophy*. Oxford: Oxford University Press (pp. 173—188).

Dalla Barba G. (1993). Different patterns of confabulation. *Cortex, 29*, 567–81.

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191, 155-185.

DeLuca, J. (2001). A cognitive neuroscience perspective on confabulation. *Neuropsychoanalysis, 2,* 119-132.

Demery, J.A., Hanlon, R.E., & Bauer, R.M. (2008). Profound amnesia and confabulation following brain injury. *Neurocase,* 7, 295—302.

Dewhurst, S. A., Thorley, C., Hammond, E. R., & Ormerod, T. C. (2011). Convergent, but not divergent, thinking predicts susceptibility to associative memory illusions. *Personality & Individual Differences, 51*, 73-76.

Ellis, H.D., Whitley, J., & Luaute, J.P. (1994). Delusional misidentification: the three original papers on the Capgras, Fregoli, and intermetamorphosis delusions. *History of Psychiatry, 5,* 117–146.

Feinberg, T.E., Venneri, A., Simone, A.M., Fan, Y., Northoff, G. (2010). The neuroanatomy of asomatognosia and somatoparaphrenia. *Journal of Neurology, Neurosurgery, and Psychiatry, 81*, 276–281.

French, L., Garry, M., & Loftus, E (2008). False memories: a kind of confabulation in non-clinical subjects. In W. Hirstein (Ed.) *Confabulation: views from neuroscience, psychiatry, psychology, and philosophy*. Oxford: Oxford University Press (pp. 33—66).

Fotopoulou, A. (2010). The affective neuropsychology of confabulation and delusion. *Cognitive Neuropsychology, 15*, 38–63.

Fotopoulou, A. Conwat, M., Griffiths, P., Birchall, D., & Tryer, S. (2007). Self-enhancing confabulation: revisiting the motivational hypothesis. *Neurocase, 13,* 6–15.

Greenberg, D. L. (2004). President Bush's false 'flashbulb' memory of 9/11/01. *Applied Cognitive Psychology, 18*, 363–370.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Hirstein, W. (2005). *Brain Fiction*. Cambridge, MA: MIT Press.

Hirstein, W. (2009). The name and nature of confabulation. In J. Symons and P. Calvo (Eds.) *The Routledge Companion to Philosophy of Psychology*. New York: Taylor & Francis (pp. 647-658).

Howe, M. L., Garner, S. R., Charlesworth, M., & Knott, L. (2011). A brighter side to memory illusions: False memories prime children's and adult's insight-based problem solving. *Journal of Experimental Child Psychology*, *108*, 383–393.

Johnson, M.K., O'Connor, M., & Cantor, J. (1997). Confabulation, memory deficits, and frontal dysfunction. *Brain and Cognition, 34*, 189-206.

Klein, S. (2013). The temporal orientation of memory: It's time for a change of direction. *Journal of Research in Applied Memory and Cognition, 2*, 222-234.

Korsakoff, S. (1889). Psychic disturbance in conjunction with peripheral neuritis. Trans. M. Victor and P.I. Yakovlev. *Neurology* (1955)*, 5,* 394-406.

Kraepelin, E. (1919). *Dementia Praecox and paraphrenia*. Ed. G.M. Roberston. Trans. R.M. Barclay. Edinburgh: E&S Livingstone.

Langdon, R., & Turner, M. (2010). Delusion and confabulation: Overlapping or distinct distortions of reality? *Cognitive Neuropsychiatry, 15*, 1–13.

Loftus, E. F. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews: Neuroscience, 4*, 231—234.

Lofuts, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology, 4*, 19–31.

Loftus, E. F. & Palmer, J. C. (1974). Reconstruction of auto-mobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585 -589.

Loftus, E. F. & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25*, 720-725.

McKenna, P.J., Lorente-Rovira, E., & Berrios, G.E. (2009). In W. Hirstein (Ed.) *Confabulation: views from neuroscience, psychiatry, psychology, and philosophy*. Oxford: Oxford University Press (pp. 159—172).

Michaelian, K. (2011). The epistemology of forgetting. *Erkenntnis, 74*, 399-424.

Michaelian, K. (2012). Generative memory. *Philosophical Psychology 24*, 323—342.

Moscovitch, M. (1989). Confabulation and the frontal systems: strategic vs. associative retrieval in neuropsychological theories of memory. In H. Roediger and F.I.M. Craik (Eds.) *Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving.* Hillside, NJ: Erlbaum Associates (pp. 133—160).

Murphy, D. (2006). *Psychiatry in the Scientific Image*. MIT Press.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Paradis, C. M., Solomon, L. Z., Florer, F. & Thompson, T. (2004). Flashbulb memories of personal events of 9/11 and the day after for a sample of New York City residents. *Psychological Reports, 95*, 304–310.

Ramachandran, V.S. (1996). What neurological syndromes can tell us about human nature: some lessons from phantom limbs, Capgras syndrome, and anosognosia. *Cold Spring Harbor Symposia on Quantitative Biology, 65*, 115-134.

Robins, S.K. (forthcoming). Misremembering. *Philosophical Psychology*.

Roediger, H.L. (1996). Memory illusions. *Journal of Memory and Language, 35*, 76-100.

Schnider, A (2008). *The Confabulating Mind: How the brain creates reality*. Oxford: Oxford University Press.

Swartz, B.E., & Brust, J.C. (1984). Anton's syndrome accompanying withdrawal hallucinations in a blind alcoholic. *Neurology, 34,* 969-973.

Talberg, I.M., & Almkvist, O. (2001). Confabulation and memory in patients with Alzheimer's disease. *Journal of Clinical Experimental Neuropsychology, 23*, 172-184.

Talland, G.A. (1965). *Deranged Memory*. New York: Academic Pres.

Wernicke, C. (1906). *Grundriss der Psychiatrie (2nd ed).* Leipzig: Thieme.