Penultimate version. Paper forthcoming as Chapter 23 in Sullivan-Bisset, E. (Ed.) *Routledge Handbook of the Philosophy of Delusion*, Routledge. Please do not cite without permission.

Delusion and Memory

Sarah Robins and Si-Won Song

1. Introduction

Most research on the intersection between memory and delusion has focused on *confabulation* — a term initially used to characterize false memories in dementia patients, which has since expanded to include a broader range of errors in clinical and everyday settings. Confabulation has received a good deal of sustained treatment from philosophers of memory in recent years (Michaelian 2016; 2020; Bernecker, 2017; Robins, 2019; 2020). In the current chapter our aim is to explore the intersection of memory and delusion beyond confabulation: does memory play a role in other delusions?

Capgras delusion has received a good deal of attention from philosophers and other cognitive scientists. Amongst clinicians, and some philosophers, Capgras is viewed as an affective disorder, according to which the connection between familiarity and the response has been somehow severed (Ellis & Young 1990, Ramachandran 1999). Philosopheres have, however, become increasingly interested in pointing out the inadequacies of the affective account, and have offered alternative proposals for understanding Capgras (e.g., Hirstein 2010, Bongiorno 2020, Coltheart & Davies 2021).

While these views are commendable for the increased attention and nuance they bring to the explanation of Capgras, they all continue to face challenges accounting for the full range and nature of the delusional phenomena in Capgras. In this chapter, we propose an alternative characterization according to which the Capgras delusion is a result of a malfunction in the way memories are updated. A version of this *memory updating view* was first proposed by Dennis Staton and colleagues in 1982, but received very little attention and uptake. We update and revive the view here, arguing that it has myriad of advantages that warrant its serious consideration.

2. Capgras Delusion

The Capgras delusion occurs when a person claims that someone familiar to them has been replaced by an imposter. Documented cases are rare. A recent metaanalysis (Pandis et al. 2019) includes 255 documented cases in the century since Joseph Capgras' initial identification. Capgras (1923) identified the delusion in a 52-year-old Parisian woman who insisted that her husband had been replaced by an imposter who bore a great resemblance to him. Subsequent cases have tended to follow this pattern: a person is brought in by a loved one for their insistence that they or another loved one has been replaced by an imposter. Often, these reports are made after repeated reassurances and attempts to provide evidence to the contrary by the alleged imposter and other friends and family. Despite all of this, persons experiencing Capgras delusion continue to insist that the loved one has been replaced with a duplicate.

The delusion has captured the attention of philosophers, as well as other scientists and clinicians, not because of its pervasiveness but because of its peculiarity. The delusion is alarming. It involves a person seeming to simultaneously recognize and fail to recognize someone they know well. After all, persons experiencing this delusion do, in some sense, recognize the familiar person. They are able to accurately identify their physical features, claiming them to be an imposter rather than a stranger. Ramachandran (1999), for example, reports a person experiencing the Capgras delusion to have said, "That guy isn't my father. He just looks like him" (p. 159.) Occasionally, persons experiencing the delusion will offer justifications for their claims, which are based on alleged discrepancies between the imposter and the original. Persons experiencing Capgras delusion have, for example, suggested that the imposter's nose is a different size (O'Reilley, Malholtra 1987) or that their hand is softer (Rojo et al., 1991). However, many of these discrepancies are minor and would not be cause for concern, much less the basis for a claim that the person is an imposter, in standard circumstances.

The Capgras delusion is often directed toward a person who is emotionally significant; a spouse, parent, or child is often the one labeled an imposter. The Pandis and colleagues (2019) metaanalysis suggests that the determination of which of these emotionally significant relationships is affected by the delusion is influenced by the person's life stage. When the delusion onset is at a younger age, the imposter tends to be a parent, while those who first experience the delusion when they are older tend to claim either their spouse or child is the imposter. Most persons experiencing Capgras delusion focus their claim of duplicitous duplication on one or two persons. In a few cases, persons experiencing the delusion identify a handful of imposters. Regardless of who is selected, subjects experiencing Capgras delusion limit their claims about an imposter to only a select few. This is not a disorder that involves indiscriminate misidentification.

There is no standard demographic profile for a person experiencing Capgras delusion, as the delusion can result from different underlying causes. Often, Capgras delusion is a symptom of schizophrenia, but can also be caused by neurodegenerative diseases such as Alzheimer's (Edelstyn NMJ, Oyebode F 1999, for more on delusions in the disorders of old age see Hughes, Ch. 11, this volume). Thus, generally, those affected tend to either be in their early twenties, the standard age of onset for schizophrenia, or in late adulthood, when neurodegenerative diseases are most common. Capgras delusion can also occur as a result of brain trauma - for example, a car accident - in which case the age of onset is more variable. Regardless of its underlying cause, Capgras delusion is persistent. A person experiencing it will continue to claim that the previously familiar person is an imposter for extended periods of time. The duration typically corresponds to the duration of the underlying condition. Those who experience Capgras delusion as a result of the onset of schizophrenia, for example, will generally continue to experience the delusion until their schizophrenia is well-managed. When the delusion occurs as a result of brain trauma or neurodegeneration, the delusion can persist indefinitely.

3. Affective Responsiveness Hypothesis

Clinicians have long favored an affective explanation of the Capgras delusion. Hirstein (2010) dubs it the "affective responsiveness hypothesis," according to which Capgras delusion is the result of a deficit of an autonomic response that is one of the two kinds of recognition involved in facial recognition.

The affective responsiveness hypothesis is standardly credited to Ellis & Young (1990), who developed the view. The general idea that the affective features of recognition are thwarted in cases of this delusion has, however, long been discussed. As Ellis & Young refer in their paper, Capgras (1923) discussed it in his original paper, and similar remarks can be found in Derombies (1935), Lewis (1987), Bauer (1986), Anderson (1988), Brochado (1936) and Weinstein & Burnham (1991). Ellis & Young were, however, the first to develop the claim into a full-fledged view, with neuroanatomical support for the explanation. Ellis and Young were inspired by work on *prosopagnosia*, a disorder where a person is unable to recognize faces. Prosopagnosia patients, peculiarly, show an autonomic response to familiar faces, but not to unfamiliar ones, even when they are unable to recognize the familiar faces (Bauer 1984; Tranel & Damasio 1985, 1988). Facial recognition, in standard cases, involves two distinct kinds of recognition: 1) reidentifying the facial features as belonging to a particular person and 2) an autonomic/physiological response (Bauer 1984). These two forms of recognition normally occur together when we recognize a person, but they are neurologically dissociable. So, for example, when I see a friend of mine, two types of recognition occur: one where I am able to recognize my friend by their physical features and can identify who the features belong to and another that involves a kind of "glow of arousal"[1] where I have a physical response to the perception of familiar facial features. In many cases, the response may be so faint as to remain below the level of awareness, but can nevertheless be detected in experimental contexts.

The first kind of recognition is referred to as *overt recognition* in the literature, while the latter is referred to as *covert recognition*. Despite prosopagnosia patients' inability to overtly recognize familiar faces, they seemed to nonetheless have some covert recognition when looking at familiar faces, as detected by increases in their skin conductance response (SCR) when shown familiar faces. SCR is a measure of the electrical conductivity of a person's skin, which increases when exposed to physically arousing stimuli. It is considered to be an empirical measure of a person's affective response to stimuli. The tendency for prosopagnosia patients to show heightened SCR to familiar faces suggest the persistene of covert recognition of familiar persons they are unable to overtly recognize.

The affective responsiveness hypothesis proposes that Capgras is the inverse phenomenon: overt recognition without a covert recognition response. In Capgras, familiar people are still registered as familiar, but the standard autonomic response to such familiar faces does not come along. Ellis and Young propose that while the patient can correctly recognize that the physical features they see belong to a familiar person, they do not experience the glow of arousal that normally also occurs during recognition. Because these two types of recognition usually occur together, lacking one of the recognition creates an anomalous perceptual experience that is thought to be the cause of the delusion.

4. Etiological Problem

Philosophers have explored the affective response proposal in more detail, asking questions about how precisely the recognition should be understood and how it accounts for the Capgras delusion. A distinction is then made between *explanationist* and *endorsement* accounts, which focuses on how the anomalous experience and delusion are related. From this perspective, Ellis and Young's (1990) account is characterized as an *explanationist* account. The affective response itself has relatively

[1] Term borrowed from Pacherie (2008).

minimal content – i.e., it registers only that the person is unfamiliar. The delusion then arises out of the attempt to explain this anomalous experience.

Alternatively, others have proposed an *endorsement* account (Bayne & Pacherie 2004; Pacherie et al. 2008). On this view, the content of the affective response is understood to be much richer such that the delusion needs only to provide an endorsement of this state rather than an explanation. Such accounts have been viewed as particularly useful amongst philosophers trying to parse the relationship between abnormal perceptual experience and delusional testimony. Bayne and Pacherie (2004), for example, assert that the statements delusional patients make should be understood as a reflection of what their abnormal experience is like. If a patient asserts that they think that their father has been replaced by an imposter, an endorsement account would say that this assertion is more or less what the imposter is experiencing. To the patient, the supposed "imposter" does appear to genuinely be a different person.

Both views face what Bongiorno (2020) calls the *etiology problem*, which highlights the distance between the affective disruption and the delusional content. Consider first the explanationist approach. A thinly characterized affective response could serve as the root of the delusion, but it is unclear how such a response could explain *why* the person goes on to claim that the person who fails to generate an affective response in them is an imposter. That is, the disconnect explains the anomalous experience, but not the subsequent interpretation of the anomalous experience. The challenge is especially striking considering how varied persons experiencing Capgras delusion are, in terms of personality, age, background, and underlying medical condition.

The etiology problem presses on the endorsement account from the other side. While there are benefits to understanding anomalous experiences in the enriched way they favor, it is unclear that the mechanism involved is capable of doing so. The affective mechanisms supporting familiarity responses are not generally thought to be such that they could support elaborate, nuanced contents of the sort that would directly envoke the idea of an imposter.

Ultimately, the etiological problem demonstrates that affective response cannot account for the Capgras delusion. There are competing accounts of the affective content involved in this endorsement (Bongiorno, 2020; see also the expressivist proposal from Bradley & Gibson, forthcoming). It remains unclear which way of thinking about the content is most apt - and more generally, there are no clear guidelines in place regarding how competing content attributions should be adjudicated.

Many philosophers have thus been motivated to look for ways that the affective response could be supplemented by patterns or forms of reasoning. This then leads to debates between one- and two-factor accoutns over what these additional reasoning features are and whether they should be understood as abnormal or not.[2] The ensuing discussion often focuses on the additional reasoning factors, moving attention further away from the initial abnormality. We are setting this debate aside, proposing a return of focus to how the basic disruption to recognition and familiarity

[2] Philosophical accounts of delusions can generally be divided into one- and two-factor accounts. One-factor accounts (Maher 1974, Gerrans 2002 and Sullivan-Bissett and Noordhof 2021, Ch. 29, this volume) locate the abnormality entirely within sensory or perceptual systems, arguing that all additional reasoning elements fall within the normal range. Two-factor accounts (Langdon, Coltheart 2000, Garety and Freeman 1999; Garety et al. 2001; McKay 2012; Davies and Coltheart, Ch. 30, this volume) couple the phenomenal abnormality with an additional cognitive, interpretive, or reasoning abnormality.

in the Capgras delusion should be understood. For this reason, we focus our discussion below on accounts that offer an alternative to the affective response hypothesis.

Hirstein (2010) and Wilkinson (2016) both suggest abandoning the affective view of the Capgras delusion and replacing it with something else entirely. They differ over what they propose as a suitable replacement.

Hirstein (2010) suggests that Capgras is the result of damage to the representations featured in the person's theory of mind.[3] Mindreading is the ability to predict and explain the behavior of oneself and others in terms of mental states, most particularly beliefs and desires. Philosophers debate about how to best characterize this capacity as well as its scope in human development and in other organisms, but there is nonetheless consensus about the existence of such a capacity (Andrews, Spaulding, & Westra 2020). To explain how the Capgras delusion is related to this overall capacity, Hirstein emphasizes the role of perspective-taking in mindreading. Specifically, Hirstein characterizes perspective-taking in mindreading in terms of the ability to create egocentric representations of the minds of others. Not all mindreading involves such intense perspective-taking. But, Hirstein argues, in cases where we know a person well, it becomes possible to employ not just generic viewpoints, but person-specific perspectives on the world.

Suppose you are set to meet someone at the movie theater, and you arrive a few minutes late to find that they are not at the agreed-upon location. Given the context, where use of cell phones is prohibited, you cannot contact them. Instead, you have to reason about what they might have done. If the person is a casual acquaintance, you may engage in mindreading by taking a generic perspective: what is a person likely to do in such a situation? If, however, it is someone you know well, then you can deploy your own egocentric representation of their perspective, asking more specifically what *that person* is likely to do in such a situation. The ability to incorporate specific features presumably provides richer information that can yield more accurate predictions.

Damage to this ability for egocentric mindreading is, Hirstein (2010) claims, at the basis of the Capgras delusion. As he argues, it can be understood as "caused by damage to the mind-representing part of this large egocentric representation system" (2010: 243). How does the loss of an egocentric representation of another person lead to the belief that the person is an imposter? Hirstein claims that the deficit occurs only to the egocentric representations of the person, leaving other information about the person intact. That is, a Capgras patient can still represent what their loved one looks like, but can no longer generate any sense of what the world is like from their perspective. It is this dissonance that leads to the conviction that the person is an imposter.

Wilkinson (2016) offers a different proposal for the primary deficit in the Capgras delusion, which he characterizes in terms of mental files. In developing this account, Wilkinson appeals to Recanati's (2012) approach to informational semantics, which explains the ability to have thoughts about individuals in terms of mental files that refer to those individuals. Mental files explain how thought about individuals is possible. As a disordered way of thinking about familiar individuals, Wilkinson reasons, Capgras delusion is best understood in terms of damage to this system - as the

[3] Hirstein's (2010) mindreading account is meant to explain not only Capgras delusion, but also other misidentification syndromes like asomatognosia.

"mismanagement of files" (2016: 396). The *mental file* is the referring concept that identifies a particular individual as the one we are thinking about. When we encounter an individual for the first time, we open a file on them and fill it with pertinent information about them. Subsequent thoughts about and interactions with that same person reopen the file, allowing for further information to be added.[4] Appeal to mental files has been used in the literature on reference to explain errors in identifying individuals - e.g., cases when a new file is opened for a person that has been encountered previously. Capgras delusion is, then, a particular way of mismanaging files, where a completely new file is created for a well-known person rather than merging the new open demonstrative file with the file of the familiar person. What happens in the case of Capgras is that the subject notices the resemblance between the newly opened file and a previously built one, and concludes that the latter individual is an impersonator of the former rather than the same person. As Wilkinson sees it, the mental files approach offers a helpful way of distinguishing between identifying individuals and predicating over them, which can be used to explain the distinct elements of what goes wrong and right in Capgras, respectively.

6. Selectivity Problem

Each of the accounts of the Capgras delusion reviewed in the previous section offers a plausible way to address the etiological challenge. For the purposes of this chapter, we are not interested in adjudicating between them further on this point. Instead, we claim that even if the etiology challenge can be met, existing accounts of the Capgras delusion all face an additional and largely-overlooked challenge: *the selectivity problem*. As discussed above in Section 2, persons experiencing Capgras delusion focus that delusion on very few people. In most cases, the Capgras delusion is focused on a single person. In some cases, two or three familiar persons are claimed to be imposters. Why is the delusion selective? Any successful account of the Capgras delusion will have to explain how the anomalous experience can arise in one or only a handful cases. This is the selectivity problem.[5]

The original Affective Responsiveness Hypothesis of Ellis and Young (1990) attempted to address selectivity via familiarity. If it is the reduced affective response that triggers the delusion, then the reduced response would be most prominent and most likely to do so, in the case of the person who is most familiar to the patient. This helps to explain why it is standardly a person who is emotionally significant to the patient who is claimed to be an imposter. But familiarity is not nearly selective enough. Nearly everyone is familiar with many people. Coltheart, Langdon, and McKay (2011) thus propose a slightly more nuanced approach to Capgras' selectivity, suggesting that it is only in cases where the familiarity reaction should have been so strong that the discrepancy is large enough to produce the delusion. Familiarity admits of degree so there may be some such threshold. Yet it seems strange that the delusion would be set such that it picks whatever level corresponds to one or possible two to three persons across all those who experience the delusion. There are also cases that directly challenge these proposals - for example, Nuara and colleagues (2020) discuss a patient who developed the delusion toward his son, but not his daughter, when all reports indicate equal familiarty with and fondness toward both children.

---

[4] As Wilkinson explains it, the information in the file includes things like "what they have done, when the subject has encountered them in the past, character traits etc., as well as what they look like" p. 397.

[5] Note that the selectivity problem as posed is not a question about who was selected. We do not take the issue to be *Why did the patient claim their mother was an imposter rather than their father?* We suspect answers about which individual is selected are likely to be determined by the particular details of a patients' life. The selectivity question is thus not *Why X and not Y* but instead *why 1 not 100 (or all)?*

The selectivity problem is not only a challenge for explanationists. Endorsement accounts of the Capgras delusion are also not well-equipped to address the selectivity problem. Such accounts are focused on the delusion's content, not its origin. That is, endorsement accounts take hold at the point of the abnormal perceptual experience. They do not offer any account of how the abnormal perceptual experience was generated. This could mean that an account that answers the selectivity problem could be supplemented with an endorsement view of the delusion's content, but on its own endorsement cannot explain why Capgras is selective.

Replacement accounts like Hirstein (2010) and Wilkinson's (2016) show tacit sensitivity to the selectivity problem. In rejecting the affective approach, they both offer attempts to replace it with a mental process that includes discrete representations of individual people. Hirstein locates these representations within the mindreading system; Wilkinson locates them within mental files. Each architectural proposal makes it possible to lose or disrupt some representations but not others.

No account of the Capgras delusion is complete without an explanation of its selectivity. It must offer a way to make sense of the delusion's focus on only a select few individuals. Accounts based in the affective responsive hypothesis appear to lack the resources to account for Capgras' selectivity. Replacement accounts, in contrast, show promise. The question thus turns to evaluating the various replacement proposals: do either of the proposed capacities offer the right kind of individualized representations of persons? If not, what other candidates are available. We turn to these questions in the next section.

7. Selectivity and Cognitive Architecture

Replacement views of the Capgras delusion show potential for addressing the selectivity problem because they locate the delusion in a system that, in standard conditions, stores representations of individual persons. The delusion then represents a disruption to that system such that one or more representations are damaged. This makes it possible to have a delusion focused on a single person, leaving the majority of person representations intact and preventing the claim of imposter from spreading more widely. Both Hirstein (2010) and Wilkinson's (2016) replacement accounts are well-structured to address selectivity, but it remains to be determined whether the systems in which they locate this selectivity can provide such representations of individual persons and do so in a way that makes sense of the delusion.

6.1 Mindreading. Hirstein (2010) situates individual person representations within the mindreading system. According to Hirstein, Capgras delusion is the result of an egocentric representation of a familiar person becoming damaged and inaccessible. There are, however, at least three issues that arise for his account.

First, while there are many accounts of mindreading that include representations of individual persons, it is unclear that the kind of representations they use could accommodate the distinctively egocentric perspective that Hirstein proposes. According to "model" views of mindreading (e.g., Spaulding 2018; Maibom 2009, Godfrey Smith 2005), mindreading happens in much the same way as much of scientific reasoning: it relies on the creation and systematic manipulation of a model of the world. Model views of mindreading incorporate many features, including representations of individual persons for individuals that the particular mindreader knows well. This, at least, is well-aligned with Hirstein. It is unclear, however, how these representations could account for a deficit in egocentric ways of representing those individuals. Model-based

accounts include systematized information about individuals - what they believe, what they're prone to do, etc. They do not include information about what the world is like from that person's perspective. Such information would be not only burdensome to store, and of limited use for predicting and explaining behavior, which may require reasoning about situations where the mindreader was not present and situations that haven't happened yet.

Second, it is not clear that patients with Capgras delusion have a mindreading deficit. Clinical reports of Capgras patients provide evidence of retained mindreading ability toward the person they now claim is an imposter. Staton and colleagues (1982) describe a patient who justified his claim that his father had been replaced by an imposter by detailing how his father's behavior contradicted his expectations. The patient expressed certainty that it was an imposter because, "my father would never have expanded the milk business" (Staton et al 1982). Similarly, Fraser and Roberts (1994) describe a patient who claimed, "my son would never kiss me," as the reason as to why she believed her son had been replaced with a stranger. There has not been, to our knowledge, any systematic testing of mindreading abilities in Capgras patients, but offhand remarks like these are suggestive of retained capacity for mindreading about the same persons toward whom the delusion is targeted.

Finally, there may be cases of Capgras delusion that target non-human or even non-minded individuals. A small but still notable proportion of Capgras cases involve the misidentification of animals and inanimate objects. According to the meta-analysis by Pandis and colleagues (2019), inanimate objects are the 3rd most common target of Capgras delusion, behind spouse and parent. Islam and colleagues (2015) describe a patient who claimed that both her pet dog and the paintings in her home had been replaced with duplicates. Others claimed that their cutlery, book, or even favorite songs have been replaced with duplicates (Ghatak, Agrawal et al 2023). In most accounts of mindreading, the capacity is focused on fellow humans. Even if it is taken to extend to include pets and other familiar animals, this would not explain cases where the delusion targets inanimate objects. Locating the individual representations inside of mindreading thus seems poorly fit to the full scope of individuals at which the delusion can aim.

6.2 Mental Files. Wilkinson's appeal to mental files is intuitive. Mental files were proposed to account for our ability to have thoughts that refer to individuals (Recanati 2012). Wilkinson straightforwardly applies this framework to cases of Capgras, arguing that the delusion occurs when the files are 'mismanaged' such that multiple files are mistakenly created for the same individual.

First, Wilkinson's appeal to mental files inherits challenges raised to the overall framework. Much of the work on mental files operates at a metaphorical level - discussing cognitive processes in terms of office workflows. Many find the metaphor quite useful, and this has likely helped promote attention to and support for this view. It remains unclear, however, whether the notion of a mental file has any psychological or neural plausibility. Is there more to the view than metaphor?[6] This matters for the application to Capgras delusion because the implementation of this framework is critical to solving the selectivity problem. If mental files are just a loose way of speaking about the mind's organization that does not actually impose any such structure in the neurocognitive architecture, then there would be no way for this framework to account for damage to a particular component of that architecture.

[6] See Goodman and Gray (2022) for an extended and compelling argument that there is not.

Second, while the mental files approach is focused on thoughts about individuals, the concerns around which it is built are quite different from those at the heart of the Capgras delusion. Mental files were developed from broader concerns in the philosophy of language about the nature of singular thought and reference. The focus is largely on explaining how thoughts about and inferences over individuals are possible. There is a lot of work on how mental files can be used to make determinations of identity or distinctness. In cases of Capgras delusion, however, what needs to be answered is why retrieval of the correct file failed. And why the file cannot be mended and updated, as happens in all cases other than the imposter.

Here the metaphorical nature of mental files compounds the confusion. What does it mean to 'mismanage' files and how might we make progress on understanding that in terms of plausible cognitive processes? In thinking through this question, we find a discussion in Ramachandran's (1999) helpful. Ramachandran describes his patient as a young man who developed Capgras delusion toward his parents as the result of a car accident. The patient had no prior history of psychiatric conditions, or other signs of psychiatric illness. Much like other Capgras patients, his delusion was selective. He took his parents to be imposters, but his ability to recognize other people remained intact. Ramachandran and colleagues ran many tests to investigate the nature and scope of the patient's facial recognition deficit. They engaged the patient in a task to test his ability to track gaze direction, showing the patient a series of pictures featuring strangers and asking him to judge which direction the person was looking and whether the direction of gaze changed across photos. The patient performed well across photos where there were slight changes in the subject's direction of gaze. When there were significant changes in the person's direction of gaze, however, the patient reported that it was a new model in the picture. The patient would offer some explanation for their judgment, claiming the person in the photo looked older, etc.

This intriguing result suggests that the Capgras deficit may involve some sort of failure of updating information about individuals over time. A failure to update is, in a sense, a form of mismanagement - but it is a more specific form, and one more closely tied to cognitive processing. Rather than attempt to expand the mental files approach to make sense of this, we think it is potentially more promising to develop an alternative replacement account centered on memory updating.

8. Memory Updating Account

In this final section, we explore the idea that the Capgras delusion should be explained in terms of a deficit in memory updating. This idea has been around for a while. Staton and colleagues (1982) proposed it in a case report four decades ago. With its focus on a single study in the era just before Ellis and Young's (1990) affective responsiveness hypothesis took off, it was largely dismissed. Given how the literature on the nature of delusions and memory have both changed in the decades since, we think the view warrants a second look. We return to it here, elaborating and expanding on the initial proposal. Ultimately, we think the memory deficit theory is a plausible and promising account of the Capgras delusion, which is well-positioned to address both the etiological and selectivity problems discussed above.

In a short case report of a patient experiencing Capgras delusion, Staton and colleagues (1982) propose that the delusion derives from a memory deficit. More specifically, they argue that it is the patient's failure to update their memory of a familiar person that leads to the belief that the familiar person is an imposter. They describe a patient, RK with the Capgras delusion. RK was a 31

year old male with no prior psychiatric or neurological impairment, who developed the Capgras delusion following brain damage from a car accident when he was 23. At the time of referral, he held the Capgras delusion about his parents, relatives and friends. In defending his delusional states, RK repeatedly appealed to his memory as evidence that the current instantiations being imposters. For example, when RK claimed that familiar people had been replaced, he cited differences between his recollections and their current appearance and behavior.

In describing his experience, RK is quoted as saying, "The issue is that my memory is just too good"(Staton et al 1982). We agree with the patient, and the assessment of Staton and colleagues, that the deficit appears to be memory-related. We also think it is worth probing further how a cognitive capacity performing too well could be the source of a deficit or delusion. In describing his memory as working too well, RK appears to be appealing to a common assumption about memory - namely, that the purpose of memory is to retain information. On such an "archival view" of memory (Robins 2016), RK's would be exceptionally good, keeping track of so many details about familiar persons that even the smallest changes trigger a mismatch. This way of thinking about memory was more prominent several decades ago, at the time Staton and colleagues (1982) were evaluating RK's case. The archival view of memory helps to make sense of RK's remark, but it does not help to make sense of his deficit. How could memory performing its function too well be a malfunction?

In the decades since, an alternative way of thinking about the function of memory has become increasingly influential. According to the "constructive view" of memory, memory is better understood as a fluid system, where information is retained, but in a way that permits dynamic changes and alterations over time (e.g., Schacter & Addis, 2007). On such a constructive view, RK's deficit is more easily understood. A well-functioning memory retains information in a more loose and labile way so that representations of individuals can absorb a range of changes. On such a view, a memory that retains too many details in too fixed of a format would be malfunctioning. Thinking of Capgras in these terms thus fits well with current theorizing about the nature of memory.

Additionally, there are advantages to thinking of the Capgras delusion in terms of a memory deficit. The delusion clearly involves a deficit of recognition. Attempts to capture this deficit as part of the perceptual process and the accompanying affective response have not fared well, as was discussed above. Thinking of the recognition process in terms of memory may fare better. Failure to update a memory could lead to the mismatches that produce the kind of partial recognition experienced in Capgras, where the person is registered as familiar but suspicious. This would provide a straightforward response to the etiological problem. Further, it could situate Capgras within a broader set of strange phenomenal experiences that occur in cases of recognitional mishaps in memory. Much like deja vu is the false recognition of a new event as familiar; Capgras delusion could be the failed recognition of a familiar person.

The memory updating view also appears better equipped to account for phenomena that posed challenges to Hirstein (2010) and Wilkinson's (2016) accounts above. Consider again Ramachandran's (1999) patient who had difficulty tracking a single, novel individual across photos where there were significant changes in the person's direction of gaze, expression, etc. The case does not fit well with Hirstein's explanation in terms of psychological representations of well-known persons. Similarly, Wilkinson (2016) could characterize what's happening as the mismanagement of mental files, but this does not in any way explain the case. It merely describes the phenomenon. If we think about the case in terms of memory updating, a richer explanation is possible. A memory of

the individual depicted is being formed, but the features included are represented in a rigid or inflexible way, such that a significant change in how the same person is depicted registers as the kind of break indicative of a new individual. Expanding this kind of explanation further, we can see how the memory updating view fares better than other available accounts in explaining how the Capgras delusion can, in some cases, be directed toward mere acquaintances rather than close friends or family.

Memory updating may also better explain cases where the Capgras delusion is directed toward pets, inanimate objects, and environmental contexts like one's house or city. Views like Hirstein's (2010) which situate the deficit within the mindreading capacity, struggle to explain cases of this kind. Wilkinson's (2016) mental files approach may have slightly wider scope, but is still limited to individuals. Memory for particular people, places, and things need not be similarly restricted. There can be memory related to any item that is tracked over time and across events - and the same general issue of updating can explain the selective loss or disruption of particular memories over time. In this way, the memory updating view has the advantage of making clearer the connection between Capgras delusion and other delusional misidentification syndromes, such as reduplicative paramnesia where the failure to recognize previously familiar items is extended to locations. In fact, the patient RK (from Staton et al 1982) had reduplicative paramnesia in addition to Capgras delusion. He not only reported his family members as imposters, but also thought that his city and pet had been replaced. Staton and colleagues describe him as claiming that Fargo (the city he was hospitalized in) could not be the real Fargo,[7] because "they didn't have this kind of hospital in Fargo" (1982: 24), and his cat was a fake because of a new scar on its ear.

Framing the Capgras delusion in terms of memory updating also allows for a straightforward response to the selectivity problem. It makes sense for memory to be organized in a way that allows for discrete retention of familiar people, places, and things that are repeatedly encountered across time in different contexts. This kind of organization in memory makes it possible to explain how memory for some items could be degraded or lost while other memories are left intact.

To fully explore and develop this view requires a richer understanding of how memory for individuals is organized and updated. To do this would require a significant shift in the focus of current theorizing about memory. Theorizing about memory is currently guided by what James Openshaw (2022) refers to as *eventism* - our understanding of how memories are encoded, stored, and retrieved is guided by events. The standard way of thinking about declarative memory involves separating it into two distinct forms: episodic memory, which is memory for particular past experiences, and semantic memory, which is memory for general facts and information. This way of thinking about memory's organization only allows for memory of individuals as it is situated within events or tied to general pieces of information. Openshaw's focus is on the need to account for memory of objects. Here our aim is to echo this need, and encourage the expansion of this category to include persons.

In sketching a proposal for object memory, Openshaw (2022) proposes that object memory be understood as a collection of information about an individual. This initial proposal is promising, but does not yet have the fluidity or flexibility to explore how the memory of an individual, whether it's an object or a person, could be updated to allow for changes over time without loss of unity.

[7] This delusion is called Reduplicative Paramnesia, which is a monothematic misidentification delusion involving location.

Even for inanimate objects, there will be changes over time - in the perspective from which they are viewed, the available lighting, as well as surrounding items and noise, and so on. For pets and persons there will be all of these features and more as the individuals grow, age, and move around.

In thinking about the complexity of such cases, we are struck by the poignancy of an example from Mark Rowlands (2016), where he describes a childhood memory including his father. In reflecting on this memory, Rowlands comes to realize that, in his memory, his father's face is depicted as it looks now, rather than how it looked in his childhood. This kind of case is suggestive and compelling for thinking about how memory of persons in particular may exert an influence on episodic memory and its content rather than the other way around. There is much more to be explored here. We contend that the appeal of using such a framework to account for Capgras delusion provides key motivation for its development.

9. Conclusion

In this chapter we have explored standard accounts of the Capgras delusion, highlighting an explanatory limitation that they all share: the selectivity problem. The Capgras delusion often targets a single person, or at most a handful of persons and other animals or objects. It is a constraint on any adequate account of this delusion to explain how such a selective disruption of recognition processes is possible. We argue that, at a minimum, this requires an account to include representations of individuals. Views based in an affective response cannot do this, and so accounts that seek to replace the affective responsiveness hypothesis have an advantage over views that seek instead to supplement the affective view. We argue that a view of the Capgras delusion in terms of memory updating offers the best such replacement account. Building off of a long-neglected proposal from Staton and colleagues (1982), we explain how such a view aligns with contemporary theorizing about memory and well accounts for the selectivity of Capgras.

References

Anderson, D.N. (1988) "The Delusion of Inanimate Doubles," *British Journal of Psychiatry*, 153(5), pp. 694–699.

Andrews, K, Spaulding, S & Westra, E. (2020). Introduction to Folk Psychology: Pluralistic Approaches. *Synthese* 199 (1-2):1685-1700

Bauer, R.M. (1984) "Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test," *Neuropsychologia*, 22(4), pp. 457–469.

Bauer, R.M. (1986) "The Cognitive Psychophysiology of Prosopagnosia," *Springer eBooks*, pp. 253–267.

Bayne, T. and Pacherie, E. (2004) "Bottom-Up or Top-Down: Campbell's Rationalist Account of Monothematic Delusions," *Philosophy, Psychiatry, & Psychology*, 11(1), pp. 1–11.

Bernecker, S. (2017) A Causal Theory of Mnemonic Confabulation. *Frontiers in Psychology* 8.

Bongiorno, F. (2020) "Is the Capgras delusion an endorsement of experience?," *Mind & Language*, 35(3), pp. 293–312.

Brochado, A. (1936) "Le syndrome de Capgras." *Ann. M´edico-Psychologiques* 15, 706–717

Capgras, J.-B. and Reboul-Lachaux, J. (1994) "L'Illusion des 'sosies' dans un délire systématisé chronique," *History of Psychiatry*, 5(17), pp. 119–133. (Original work published 1923)

Coltheart, M. and Davies, M. (2021) "What is Capgras delusion?," *Cognitive Neuropsychiatry*, 27(1), pp. 69–82.

Coltheart, M., Langdon, R. and McKay, R. (2007) "Schizophrenia and Monothematic Delusions," *Schizophrenia Bulletin*, 33(3), pp. 642–647.

Coltheart, M, Langdon, R and McKay, R (2011) "Delusional Belief" *Annual Review of Psychology* 62:1, 271-298

Davies, M. *et al.* (2001) "Monothematic Delusions: Towards a Two-Factor Account," *Philosophy, Psychiatry, & Psychology*, 8(2), pp. 133–158.

Derombies, M. (1935) L'illusion de sosie, forme particuliers de la reconnaissance systematique. *Annales Medico-Psychologiques*, 94, 706–717

Edelstyn, N.M.J. and Oyebode, F. (1999) "A review of the phenomenology and cognitive neuropsychological origins of the Capgras syndrome," *International Journal of Geriatric Psychiatry*, 14(1), pp. 48–59.

Ellis, H.D. *et al.* (1997) "Reduced autonomic responses to faces in Capgras delusion," *Proceedings of the Royal Society B: Biological Sciences*, 264(1384), pp. 1085–1092.

Ellis, H.D. and Young, A.J. (1990) "Accounting for Delusional Misidentifications," *British Journal of Psychiatry*, 157(2), pp. 239–248.

Freeman, D., & Garety, P. A. (1999) Worry, worry processes and dimensions of delusions: An exploratory investigation of a role for anxiety processes in the maintenance of delusional distress. *Behavioural and Cognitive Psychotherapy, 27*(1), 47–62

Garety, P. A., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. E. (2001) A cognitive model of the positive symptoms of psychosis. *Psychological medicine*, *31*(2), 189–195. https://doi.org/10.1017/s0033291701003312

Gerrans, P. (2002) Multiple Paths to Delusion. *Philosophy, Psychiatry, and Psychology* 9 (1):65-72

Ghatak, M. *et al.* (2023) "The Capgras Delusion of inanimate objects: A duplicate world," *Psychiatry Research*, 2(1), p. 100106.

Godfrey-Smith, P. (2005) Folk psychology as a model. *Philosophers' Imprint* 5:1-16

Goodman, R. & Gray, A. (2022) Mental filing. *Noûs* 56 (1):204-226.

Hirstein, W. (2010) "The misidentification syndromes as mindreading disorders," *Cognitive Neuropsychiatry*, 15(1–3), pp. 233–260.

Islam, L., Piacentini, S., Soliveri, P., Scarone, S., & Gambini, O. (2015) Capgras delusion for animals and inanimate objects in Parkinson's Disease: a case report. *BMC psychiatry*, *15*, 73. https://doi.org/10.1186/s12888-015-0460-7

Lewis, S. (1987) "Brain Imaging in a Case of Capgras' Syndrome," *British Journal of Psychiatry 150,* pp. 117–121

Maibom, H, (2009) In defence of (model) theory theory. *Journal of Consciousness Studies* 16 (6-8):6-8

Coltheart, M. & Davies M. (2022) "What is Capgras delusion?," *Cognitive Neuropsychiatry*, 27:1, 69-82

Maher B. A. (1974) Delusional thinking and perceptual disorder. *Journal of individual psychology*, *30*(1), 98–113.

Maibom, H. L. (2009). Feeling for Others: Empathy, Sympathy, and Morality. *Inquiry: An Interdisciplinary Journal of Philosophy* 52 (5):483-499

McKay, R. (2012), Delusional Inference. *Mind & Language*, 27: 330-355.

Michaelian, K (2016). *Mental Time Travel: Episodic Memory and our knowledge of the personal past.* MIT Press.

Michaelian, K. (2020). Confabulating as Unreliable Imagining: In Defence of the Simulationist Account of Unsuccessful Remembering. *Topoi* 39 (1):133-148.

Ghatak, M., Agrawal, A., *et al* (2023) "The Capgras delusion of inanimate objects: A duplicate world," *Psychiatry Research Case Reports*

Noordhof, P., Sullivan-Bissett, E. (2021) The clinical significance of anomalous experience in the explanation of monothematic delusions. *Synthese* 199*,* 10277–10309

Nuara, A. *et al.* (2020) "Catching the imposter in the brain: The case of Capgras delusion," *Cortex*, 131, pp. 295–304.

Openshaw, J. (2022). Remembering objects. *Philosophers' Imprint* 22:1–20.

O'Reilly, R.J. and Malhotra, L. (1987) "Capgras Syndrome – An Unusual Case and Discussion of Psychodynamic Factors," *British Journal of Psychiatry* 151:263-5

Pacherie, E. (2008). Perception, Emotions and Delusions: The Case of the Capgras Delusion. In Tim Bayne & Jordi Fernàndez (eds.), *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation.* Psychology Press. pp. 107-125.

Pandis, C., Agrawal, N. and Poole, N. (2019) "Capgras' Delusion: A Systematic Review of 255 Published Cases," *Psychopathology*, 52(3), pp. 161–173.

Recanati, F. (2012) *Mental Files*, Oxford: Oxford University Press

Ramachandran, V.S. (1999) *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. Harper Collins.

Robins, S. (2020). Mnemonic Confabulation. *Topoi* 39 (1):121-132.

Robins, S. (2019). Confabulation and constructive memory. *Synthese* 196 (6):2135-2151Rojo, V.I. *et al.* (1991) "Capgras' syndrome in a blind patient," *American Journal of Psychiatry*, 148(9), pp. 1271b–11272.

Rojo, V.I. *et al.* (1991) "Capgras' syndrome in a blind patient," *American Journal of Psychiatry*, 148(9), pp. 1271b–11272

Rowlands, M. (2017) *Memory and the Self: Phenomenology, Science and Autobiography*. Oxford University Press.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1481), 773–786.

Spaulding, S. (2018) Mindreading beyond belief: A more comprehensive conception of how we understand others. *Philosophy Compass*. 13:e12526.

Staton, R.D., Brumback, R.A. and Wilson, H.R. (1982) "Reduplicative Paramnesia: A Disconnection Syndrome of Memory," *Cortex*, 18(1), pp. 23–35.

Noordhof, P., Sullivan-Bissett, E. (2021) The clinical significance of anomalous experience in the explanation of monothematic delusions. *Synthese* 199, 10277–10309

Tranel, D et al. (1985) "Electrodermal discrimination of familiar and unfamiliar faces: a methodology." *Psychophysiology* vol. 22,4

Tranel, D et al. (1988) "Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity." *Neurology* vol. 38,5 690-6.

Turkiewicz, G., Zanetti, M., Zung, S., & Cordeiro, Q. (2009). Coexistence of Capgras and Fregoli syndromes associated to frontotemporal volume reduction and cerebral white matter hyperintensities. *Revista De Psiquiatria Clinica*, 36(6), pp. 240– 247.

Weinstein, E.A. and Burnham, D.L. (1991) "Reduplication and the Syndrome of Capgras," *Psychiatry* 54:1 pp.78-88

Wilkinson, S. (2016) "A Mental Files Approach to Delusional Misidentification," *Review of Philosophy and Psychology*, 7(2), pp. 389–404.