Delusions, Dreams and the Nature of Identification

Abstract

Delusional misidentification is commonly understood as the product of an inference on the basis of evidence present in the subject's experience. For example, in the Capgras delusion, the patient sees someone who looks like a loved one, but who feels unfamiliar, so they infer that they must not be the loved one. I question this by presenting a distinction between "recognition" and "identification". Identification does not always require recognition for its epistemic justification, nor does it need recognition for its psychological functioning. Judgments of identification are often the product of a non-inferential mechanism. Delusional misidentification arises as the product of this mechanism malfunctioning.

1. Introduction

The Capgras delusion is the delusion that one or more known persons, often loved ones, have been replaced by identical-looking impostors. Along with a number of other delusions, it is classified as a delusion of misidentification, which is to say that the subject misidentifies someone or something that one would strongly expect them not to misidentify. What is unexpected in the Capgras case is that the subject's perceptual system seems to be working fine: she *recognizes* how things appear, in the sense that she seems to process visual information in the normal way. She is neither partially sighted, nor agnosic, and makes accurate judgments concerning what things and people look like. And yet she mis*identifies* people, often only a strikingly small class of people (namely loved ones). In the presence of, say, her father, the Capgras patient will say, "That man is not my father; he looks exactly like my father, but isn't him." Utterances with similar content can occur in the context of more global pathologies, in particular schizophrenia, but in cases where they are caused by brain damage, the subjects have been known to lead surprisingly ordinary lives, both cognitively and emotionally, when kept away from the individual or individuals that are judged to be impostors (see e.g. Hirstein and Ramachandran 1997).

Over twenty years ago, Ellis and Young (1990) proposed a model for understanding the Capgras delusion that was so well received that it generated an approach to thinking about delusions that still constitutes orthodoxy. I will label this orthodoxy the *inferential-evidential approach*, (IEA for short) since it takes the delusional misidentification to arise as an *inference* on the basis of some kind of *evidence* that is present in the subject's experience. Prominent proponents of IEA include: Maher (1974, 1988, 1999 etc.), Stone and Young (1997), Davies et al. (2001), and, to a much lesser extent, Bayne and Pacherie (2004), among several others.

I will proceed as follows. In section 2, I will explain what I mean by "inference" and "evidence". Then I present Ellis and Young's position as well as some options that have recently been taken by others in elaborating this into different forms of IEA. I look at three problematic issues for IEA. The first two (the "logical" and "epistemological" issues) are philosophical in nature. They consist in explaining how the tracking of individuals (*identification*) differs more profoundly from the recognition of qualitative similarity (*recognition*), both logically and epistemically, than IEA seems to suggest. Furthermore, it is shown to be beneficial that a cognitive system should be capable of considering something's identity entirely independently of its appearance, and that this should involve some kind of tracking mechanism. The third issue (the "psychological" issue) is empirical in nature, and is elaborated throughout section 4. It amounts to claiming not only that the two tasks of identification and recognition are different, and that it would be beneficial to have a dedicated mechanism for the former, but that humans can and do perform them using two dissociable mechanisms. I present evidence for this claim (and in particular the claim that identification can be non-inferential in the relevant sense) from empirical work on dreams and the Frégoli delusion. In section 5, I tentatively present a theory, which introduces the notion of identity files, for the underpinnings of identification, and by extension of *mis*identification. I then examine how the theory compares with its competitors.

2. What is the Inferential-Evidential Approach? (IEA)

This section is dedicated to presenting and clarifying IEA. I will begin by clarifying what I mean by "inference" and "evidence". I will then present the Ellis and Young model and clarify its epistemological and psychological consequences.

2.1. Clarifying "Inference" and "Evidence"

Both the terms "inference" and "evidence" get used in different ways. As I am using the term, an "inference" is a process that has personal, though not necessarily conscious, steps as part of a process of reasoning (in this case abductive reasoning). I mean "personal" in the sense that it is attributable to the person and not to a subsystem: it is the person, and not some part of them, that performs the inference. It is also the person who, through performing an inference, has grounds for belief, and citing these grounds provides evidential justification. This sense of "inference" is certainly more universally used in philosophy of mind and epistemology than in psychology. In psychology there is a use of "inference" on the basis of which parts of your brain can be described as performing "inferences". This is the sense in which people talk of your visual system performing inferences (e.g. on the basis of ambiguous input). Some recent paradigms (viz. the "Predictive Coding" paradigm, see Clark 2013 for a review) even think of the brain as constantly performing statistical "inferences". When I say that I am interested in whether delusional misidentification is inferential or not, I am referring to personal inference. It would be trivial, or very nearly so, to say that the delusion arises as the result of an inference in the low-level sense, since perhaps almost everything in human cognition can be thought of in this way. These low-level inferences can give rise to certain experiences, which serve as input for personal-level inference. For example, these "inferences" can give rise to certain visual illusions (e.g. the Hollow-mask illusion), and although these low-level "inferences" cannot be interfered with evidentially (namely, knowing that the mask doesn't flip doesn't prevent you from *experiencing* the illusion), knowing that the illusion is an illusion enables you to over-ride what your senses are

telling you, namely, you can perform the personal-level inference that, in spite of appearances, the mask isn't *actually* flipping inside out.

An inference, in the sense that interests us, can be personal but achieved automatically (it can be, to use Pryor's (2000) term, *psychologically immediate*) but the epistemic steps, if it is indeed an inference, should always be ascertainable (it is not *epistemically immediate*). How can we, therefore, tell between an automatic, personal inference, and something that is non-inferential in the personal sense? To use Pryor's own example, I might look at the fuel gauge in my car and, in a psychologically immediate way, come to believe that my car needs filling up. However, this is not epistemically immediate in the sense that if somebody were to tell me that the gauge is not functioning properly, I would thereby cease to take the gauge to be an indication that my car needs filling up.

How we think of "evidence" further clarifies what is meant by inference. "Evidence" is simply that which justifies a belief, drives an inference, and makes it epistemically rational (or irrational). So evidence is also a personal notion. Just as the low-level "inferences" that psychology sometimes appeals to are not inferences in this sense, so low-level information within a subsystem is not evidence.

Evidential steps mediate a judgment that is based on inference, and an inference can be rationally defeated by competing evidence (whereas a low-level "inference" cannot). Crucially, for the orthodox view of delusions, evidence thus construed, firstly, is the sort of thing that you can ignore or misuse, namely, it is that on which reasoning bias operates, and secondly (and in a way that arguably strays from philosophical orthodoxy) allows for private experiential evidence.[1] It is, for example, not "evidence" in the everyday legal sense of publicly accessible entities such as fingerprints.

---

[1] In philosophy, there are two very different readings of 'evidence'. On one, more externalist, reading, x is evidence for p if it actually increases the likelihood of p being true. On another, more internalist reading, x is evidence for p if it makes the subject more likely to believe p. To illustrate, on the first reading, a brain in a vat has no evidence for its "perceptual" beliefs, even though it thinks it has. On the latter reading it could have good evidence. This subjective kind of evidence is what Maher has in mind when he says that "delusional belief is not being held "in the face of evidence strong enough to destroy it," but is being held because evidence is strong enough to support it." (1974, p.99)

With these clarifications out of the way, we turn to Ellis and Young's model, and look at how it is a form of IEA.

2.2 Capgras Delusion as "Inverse Prosopagnosia"

Borrowing Bauer's (1984) model, whereby there are two streams for processing facial information – one covert, affective and anatomically dorsal, the other overt, semantic and anatomically ventral – Ellis and Young (1990) put forward the influential proposal that the Capgras delusion can be understood as a sort of "inverse prosopagnosia." Prosopagnosics have difficulty in the overt recognition of faces. Show them a picture of a familiar face and they will not be able to tell you whose face it is. And yet, surprisingly, some of them appear to have differential autonomic responses to these faces, as measured by heightened skin conductance response (SCR) (see Bauer 1984). Ellis and Young hypothesized that Bauer's two streams can be selectively impaired, leading to double dissociation. According to them, whereas with prosopagnosia the affective stream for "covert recognition" is intact and the semantic stream for "overt recognition" is impaired, with the Capgras delusion it is the other way around. At a personal level this means that the Capgras patient is presented with someone who (thanks to intact subpersonal semantic processing) looks to them exactly like a loved one, but the experience is somehow strange (which is the result of deficient subpersonal affective processing). The perceived person feels unfamiliar and the patient therefore concludes that this person cannot be the loved one in question. This model was given experimental support (Hirstein and Ramachandran 1997, Ellis et al. 1997) when it was discovered that, in contra-distinction to prosopagnosia, Capgras patients show diminished SCR when presented with familiar faces.[2]

---

[2] It is worth noting, however, that Bauer's dual stream model is far from being universally accepted. Breen, Caine and Coltheart (2000), for example, deny that the dorsal stream is capable of recognition. Rather, with prosopagnosics the ventral route is damaged enough to prevent conscious recognition, but, in some cases (in particular in associative rather than apperceptive prosopagnosia), not so damaged as to prevent referral to the limbic system. This critique undermines the claim that prosopagnosia and Capgras delusion exhibit a double dissociation, but it can leave the epistemic gist of the Ellis and

With this in mind, let us examine the notions of overt and covert recognition.

2.3.“Overt Recognition”, “Covert Recognition” and the Delusional Inference

As we saw, Ellis and Young think that the Capgras patient, when presented with her father, judges, “This man has the appearance of my father”. This constitutes overt recognition. “Overt” information (on the basis of which one makes an overt judgment) can be understood as that which is accessible to consciousness and provides citable justification for the subject. If asked, “Why does this man have the appearance of your father?” she might reply, “Because he has the same nose, same eyes, same hair etc.” And these are demonstrable features that (assuming her perceptual apparatus is working normally and she's not hallucinating) are available for others to see, and in principle could convince them of the correctness or plausibility of her claim. Granted, our facial recognition is modularized and holistic (see Palermo and Rhodes 2007 for a review), so we have a great facility in dealing with facial features: we don't *form* the judgment by consciously matching them up in the way just described, but the crucial point is that *post hoc* justification (and support) is available to the subject.

What is meant by “covert”? One might think that “covert” means subpersonal, but we need to be careful here. The damage to affective processing is subpersonal, but insofar as the covert recognition is a judgment, it is personal, and just as personal as the judgment of overt recognition. It is the Capgras patient, the person, who judges “This man does not feel familiar/like my father.” This is covert, not in the sense that it is subpersonal, but rather in the sense that it is not overt in the sense just described. Namely, if someone were to ask her, “Why doesn't this man feel like your father?” there is no justification that can be adequately

Young model intact, namely that a central *reason* (i.e. evidential subjective grounds) for the delusional misidentification is a lack of affective response. Breen, Caine and Coltheart (2000) themselves endorse this much, and attribute it (as Hirstein and Ramachandran 1997 do) to a disconnection between the FRU (“face recognition unit”) in the temporal cortex and the amygdala. This explains the emotional response in associative prosopagnosia, which can still be thought of as “covert recognition” in the absence of overt recognition, without claiming that the dorsal stream is capable of recognition.

It is also worth noting that the claim that some prosopagnosics exhibit covert recognition remains untouched. Rather the *anatomical* claim that this is due to intact *dorsal* processes is to be rejected.

put into words and, relatedly, there is no inter-subjectively available property that can be appealed to. The subject is thus robbed of the ability to justify her judgment to others (although it may be supported by private, experiential evidence) and would have to reply something like, "He just doesn't." An inability to (successfully) justify judgments to others is an integral part of any case that is likely to be called delusional.[3] Delusions (whatever else they may be) are beliefs that lack justification or reasonableness by other people's lights.[4] Although this covert judgment isn't yet delusional (she *could* claim: "Although this person doesn't feel like my father, I can see that it is my father"), it is an important precursor on Ellis and Young's model.

The delusional subject isn't merely claiming that this person doesn't feel familiar, as her father should feel. She is claiming that he actually *isn't* him. So where does the full-blown delusional judgment "This man is not my father" come in? Well, according to Ellis and Young, it is the product of a personal-level inference, more specifically, an abductive inference to the best explanation (Ellis and Young themselves call this a "rationalization strategy"). It takes something like this form:

1. This man looks like my father.

2. This man feels unfamiliar (hence doesn't feel like my father).[5]

[How do I explain that, although this man looks like my father (and hence should feel like my father), he fails to feel like my father?]

3. He must not *be* my father.

2.4. Accounting for Tenacity: Two-factor Theories

[3] Some delusional patients try to offer (probably confabulated) justification for their delusions, whereas others do not. Sometimes they confabulate slight differences in appearance, like "his moustache looks different" (cf. e.g. Hirstein and Ramachandran 1997).

[4] Indeed, nothing prevents delusions from being accidentally true. The schizophrenic who claims that "the end of the world is nigh", may happen to be correct thanks to an unforeseeable disaster.

[5] To avoid possible confusion, Ellis and Young do not claim that there is a specific feel for the individual and that it is this that is lacking. Rather there is a general lack of familiarity (or general presence of unfamiliarity) and that, for the subject, excludes the relevant familiar people. In this case the relevant familiar person is the father, since the person the subject sees looks just like his father.

If the inference is to be an epistemically rational one, where "epistemic rationality" roughly means that a judgment goes only as far as evidence warrants, then 2 (and what it suggests) would have to be strong enough evidence to defeat 1. If this is plausible, (and it's far from obvious that it is) it explains the initial formation of the delusion. But the delusion is also highly tenacious: it is not just formed briefly but maintained over time and in the face of much contrary evidence. For example, no matter how much one tries to convince the subject, she still thinks that the man we know is her father, and whom she agrees looks exactly like her father, is not her father. Therefore 2, the affectively unusual experience, has to defeat not just 1 but also a great deal else: the testimony of others, including people of authority such as doctors, and, of course, the obvious implausibility of the delusional scenario; aside from very rare cases (viz. identical twins) people statistically *just don't* look exactly like a certain individual and yet somehow fail to be them.[6]

Maher insists that the affectively anomalous experience can play such an epistemic role: "delusional belief is not being held "in the face of evidence strong enough to destroy it," but is being held because evidence is strong enough to support it." (1974, p.99; Reimer 2009 has recently supported this view, claiming that we simply can't underestimate the strength or strangeness of these patients' experiences). However, most have found this implausible. Instead, they claim, delusional subjects aren't rational and must have, in addition to their unusual experience, a reasoning bias or deficit (Stone and Young 1997; Davies et al. 2001). In epistemic terms, delusional subjects attach more credence (i.e. more weight) to 2 than they ought to. These are called "two-factor theories". Since these naturally build on IEA, we will address them later. They amount to claiming that the patient's judgment is based on an irrational inference.

---

[6] The patients are sometimes sensitive to this implausibility. Patient DS (Hirstein and Ramachandran 1997, p.438) when asked why this man was pretending to be his father, he replied "That is what is so surprising, doctor; why should anyone want to pretend to be my father? Maybe my father employed him to take care of me … paid him some money so that he could pay my bills." Note, however, that this assumes that the judgment of misidentification is correct: the implausibility is located at the level of the impostor's motives!

2.5. Explanationist and Endorsement Accounts of the Role of the Affective Anomaly

Whether one is a two-factor theorist or not, if one adopts IEA, the affectively anomalous experience still has to constitute strong subjective evidence in favor of the delusional hypothesis. There are, broadly speaking, two different views on the role of the anomalous experience in the formation of the delusion. So-called explanationist models are closer to Ellis and Young's initial proposal. They (e.g. Maher 1999) say that the content of the affectively anomalous experience is something sparse like, "This experience is strange" or perhaps "This man feels strange", and that the delusional judgment *explains* the bizarre experience (roughly, the subject reasons: "This man, in spite of looking like my father, doesn't *feel* like my father would feel, therefore he cannot *be* my father"). One *prima facie* problem with such a view is that, if the experience is sparse and non-specific, why is there not a wider array of potential hypotheses used to explain it? ("Maybe I don't like dad anymore", "Maybe I'm tired" etc.).[7]

So-called endorsement models (e.g. Bayne and Pacherie 2004) say that the delusional content is encoded directly in the unusual experience, and all that suffices is endorsement of that content. According to them, there are not two separate judgments (one about how the person looks, and another about how the person feels) that through a process of inference lead to the delusional judgment. Rather there is one experience that "tells" the subject something like: "This man may look like your father, but it's not him". As will become clear, I am far more sympathetic to endorsement theories that to explanationist ones. In a sense what I am doing in this paper is giving us reason to "pack a lot into the experience". Although this is conducive to the endorsement models, I am not in complete agreement with them, and still consider them, in some sense, to be remaining within IEA. Here is why.

According to them, one in principle could have exactly the same subjective experience as the Capgras patient and not (even *should* not) go on to form the delusion. In other words, the Capgras patient in principle could (and *ought* to) disregard what his experience is "telling him", namely the *evidence* provided by experience, as one might, for example, with a well-known visual illusion (both Bayne and Pacherie 2004, and Pacherie

---

[7] Thanks to an anonymous referee for making this point.

2009 explicitly cite the Muller-Lyer illusion in support of this). My proposal seeks to go beyond the endorsement theory by denying this possibility. First, the evidence that supports the claim that there are patients who have "the same experience", but who don't go on to form the delusion (Tranel, Damasio and Damasio 1995) is, as we shall see, not very convincing. For starters, "sameness of experience" is asserted on the grounds of diminished skin conductance response. But this physiological response can be interfered with in any number of ways and does not demonstrate "sameness of experience". The second point is more philosophical in nature. On an attractive, but overlooked, view of judgment, experience and the relation between the two, the difference between judging and not judging *is itself an experiential difference* (see, e.g. Pitt 2004). The sense of trusting or not trusting your senses is an experiential (or phenomenological, if you prefer) difference (see Ratcliffe 2004 and 2008 for an approach running along similar lines). .

There is, therefore, even for the endorsement theorist, inference going on, namely, the weighing up of evidence from what the current perceptual experience is telling the subject, against the subject's background beliefs (including the general implausibility that people can look exactly like others and yet somehow fail to be them). This means that endorsement theorists need to rely on a reasoning deficit to explain the tenacity of the delusion just as much as the explanationists. Indeed endorsement theorists (Bayne and Pacherie 2004) express this explicitly.

3. Problems with IEA: Queries from the Armchair

We have seen that several theories fall under IEA (one-factor, two-factor, explanationist and endorsement theories). What they all have in common is that the delusion is formed on the basis of an inference based on evidence in the subject's experience. The subject recognizes that this person *looks like* A, and this would usually be evidence to support the judgment that this person *is* A, but there is an experiential anomaly that epistemically defeats this.

There are three problematic issues that I'd like to raise for IEA. The first two are non-empirical: they constitute a logical issue and a related epistemological issue. The third, empirical issue, the "psychological" issue, is the most important and will constitute section 4.

3.1 The Logical Issue

Ordinary language allows that you can "recognize" an individual as well as a property or kind: you can "recognize" a rook, but you can also "recognize" your mother. Here superficial linguistic similarity is obscuring a deep logical difference. The former ("This [bird] is a rook") is of the form Fa whereas the latter ("This [woman] is my mother") is of the form $a = b$. Namely, one predicates a property of an individual (the referent of a singular term), whereas the other makes a connection of identity between two individuals (or rather implicitly claims that two singular terms have the same referent). This is why I prefer to use "identification" rather than "recognition" for judgments of the form $a = b$.

This terminological distinction between identification and recognition is not arbitrary, but serves to mark what I take to be a deep difference in the epistemology and psychology of the tracking of qualitative similarity and the tracking of numerical identity, namely, the judgment that something looks a certain way and the judgment that something is a certain individual. The former attributes a property to an individual; the latter picks out an individual as a previously encountered individual, which is, in principle, independent of attributed properties.

3.2 The Epistemological Issue

Before making my central epistemological claim it is important to see precisely what IEA epistemologically amounts to.

An analogy might be as follows. An ornithologist, who owns many acres of woodland, has a favorite rook, call it Bob. However, he often gets confused between Bob and the vast number of very similar looking rooks (although he can tell after a while from the rook's behavior when he has got his initial identification wrong). To put an end to this

confusion, he puts a ring on Bob's leg. Now when he sees a certain rook he might think, "This has the appearance of Bob (and therefore, so far, could *be* Bob)". Upon closer inspection, the ring is absent; therefore the possibility of it being Bob is ruled out. For IEA, the affectively underpinned covert recognition (or lack thereof) is epistemically analogous to the ornithologist seeing this ring (or seeing its absence). Both are about a supplementary and identifying piece of evidence.

However, this supplementary evidence can be defeated, and, more specifically, undercut.[8] If the ornithologist had reason to think that the ring had fallen off (e.g. he had found it on the ground), he would think twice about judging on the basis of the ring's absence that the rook in question isn't Bob.[9]

This is analogous to the Capgras case according to IEA. And yet, if we look at actual cases, the patients do have reason to believe that that evidence would be missing: they are made aware by their doctor of their brain damage and its effects. However, unlike our ornithologist, Capgras patients do not refrain from judging on the basis of this lack of identifying information that the individual they encounter is not the individual in question. One way that IEA has of explaining this, as we saw, is by saying that the delusional subject doesn't give the undercutting evidence as much weight as she ought to (and gives her experiential evidence too much weight). However, another way, which I am proposing in this paper, is to say that the delusional judgment is not formed on the basis of evidence, it is non-inferential.

Recognition, one might think, is needed to set up the possibility of identification, and hence misidentification. Someone's not feeling like your father is not evidence that defeats the judgment that this man is your father if this man looks nothing like your father, and you were not tempted to judge that he was your father anyway. But identification should not

---

[8] Following Pollock (1986) we can distinguish, within defeating evidence, between undercutting and rebutting evidence. Undercutting evidence defeats evidence e for p by putting the reliability of e in question. Rebutting evidence, on the other hand, defeats e by directly presenting stronger evidence for not-p.

[9] Conversely, if he had reason to think that a prankster had attached a similar ring on other identical looking rooks (e.g. if he saw two rooks with the same kind of ring), he would think twice about judging, on the basis of a rook's having the ring, that it is Bob.

always be thought of like this. Intuitive as this may seem, we shall see in the next section, especially when we examine the Frégoli delusion, that identification demonstrably isn't always parasitic on recognition in this way, but can operate entirely independently.

The fact that it can operate independently, though perhaps counterintuitive (and at times making it hard, or impossible, to verbally justify judgments of identification), is no bad thing when we consider the central epistemic point I want to make, which is as follows. Judging that two things are the same individual (*identification*) and judging that two things look the same (*recognition*) are very different kinds of judgment. You can judge, and indeed be justified in judging, that x at t1 and y at t2 are the same individual, in spite of them looking nothing like each other. Indeed (and this is the way around that is relevant to the Capgras delusion) you can judge that x at t1 and y at t2 are distinct individuals in spite of looking the same.

Some theorists, (e.g. Evans 1982) are sensitive to the idea that you might make identification judgments not on the basis of recognitional evidence, but on background knowledge, such as spatiotemporal location (i.e. you might be able to locate your old friend James, since you've been told that he's working the morning shift at the sandwich shop, but you might also be warned that he has changed beyond all recognition). I want to take things even further. Sometimes identificational judgments are made on the basis of no evidence at all, not even spatiotemporal location. They are not the product of inference. These "intuitive" judgments can be based on a reliable mechanism. Sometimes this reliability breaks down, and you get misidentification (and if this breakdown is stable and persistent enough, you get *delusional* misidentification). This constitutes the psychological issue I have to raise against IEA.

3.3. The Psychological Issue

The Epistemological Issue was a necessary claim, stemming from the difference between the metaphysics of unique individuals and that of multiply instantiable appearances. In any possible world, for any intelligent subject, a judgment of identification is such that it can in

principle be correct (and reliable/justified/non-accidental) without any support from recognition.

The Psychological Issue is very different. It is a contingent claim about how we humans happen to make such judgments of identification. The whole of the next section is dedicated to this.

4. On the Independence of Tracking Individuals (Identification)

Here I will just state the claim plainly: human judgments about numerical identity can and do bypass judgments about qualitative similarity (and spatiotemporal trajectory) altogether. In such a situation the judgment of identification (which could, of course, be wrong, as in delusional cases) is not something that is inferred. In other words, it is not the product of personal-level inference, biased or otherwise. It is with this that I introduce and elucidate the notion of tracking individuals, namely, judgments of identification.

4.1 Bypassing both Similarity of Appearance and Spatiotemporal Considerations

I don't want to deny that we often make judgments of identification on the basis of judgments about appearance, with the addition of some kind of abductive inference. For example, we say to ourselves, "This looks exactly like my tennis racket; it is unlikely that there should be a tennis racket that looks exactly like mine, and that isn't mine; therefore it is mine". And of course, there are also certain principles about spatiotemporal location. If the racket is where I left it, then that further supports my judgment. Or if I see it in one place, and then an identical-looking racket somewhere where it couldn't possibly be, then I ought not to judge that this racket, on the second encounter, is mine. The error of IEA is to think that this generalizes across the board for all judgments of identification (including delusional misidentification). There are many routes that lead to judgments of identification. There are certainly different *inferential* routes, different kinds of *evidence* that can lead to a judgment of

identification (e.g. perceiving similarity or keeping spatiotemporal track of an individual). However, my empirical claim is that sometimes there are routes to identification that don't make use of any personal level evidence, viz. that aren't inferential at all. Sometimes, the judgment of numerical identity relies entirely on a non-inferential mechanism, which involves neither the matching of qualitative similarity nor the application of spatiotemporal principles.[10]

Consider what happens when you encounter someone and you (implicitly) ask yourself "Have I met this person before?" In the vast majority of cases, there is no personal level inference. You don't *work out* whether you've met them before. The question is just answered.

This appeal to intuitions about everyday cases may not convince, so I turn to more robust support for the possibility of non-inferential judgments of identification.

4.2 Support for Bypassing: Identification and Recognition come apart in Dreams and Delusions

4.2.1 Dreams

Schwartz and Maquet (2002) compared the content of dream reports with functional imaging during REM sleep (i.e. the sleep stage during which vivid dreams are reported) (see also J. Hobson 1999). The central idea is that, in principle, dream features can be mapped onto specific distributions of brain activity.[11] This has generally been taken to show a distinction between "feature binding" (the construction of a perceptual or quasi-perceptual representation) and "contextual binding" (the organization of the representations into a

---

[10] A rare example of where (purported) identification not only doesn't use, but also actively flouts, spatiotemporal principles is in reduplicative paramnesia: the delusion that a place or individual is existing in two places at once.

[11] What is assumed by this approach is that the anatomical segregation of brain functions remains similar when one is dreaming to when one is awake. It has been shown that this assumption holds for audition: the presentation of auditory stimuli during REM sleep elicits responses that are similar to those elicited during wakefulness (so with no contrary evidence presenting itself the assumption is, not without justification, extrapolated across the board).

narrative or theoretical explanation) (see Gerrans 2012).[12] Instead, I would here like to focus

on what dreams tell us about the distinction between identification (tracking individuals,

perceiving identity) and recognition (the perceiving of appearance or similarity). This

distinction is located *within* feature binding, although it is clearly contentious whether and in

what sense identity can enter into a perceptual or "quasi-perceptual" representation.

Dreaming uses similar resources to waking cognition, and depends on a large-scale

neural network.[13] As Schwartz and Maquet point out, the content of certain bizarre elements

of dreams, as well as the underlying brain activity, may resemble the content and brain

activity associated with the delusions of misidentification that we are concerned with here. I

want to suggest that recognition and identification regularly come apart, not only in

delusional cases, but also in the dreams of healthy subjects. Consider this dream report: "I had

a talk with your colleague, but she looked differently, much younger, like someone I went to

school with, perhaps a 13-year-old girl." Or this one: "I recognize A's sister […] I am

surprised by her beard, she looks much more like a man than a woman, with a big nose"

(both quoted in Schwartz and Maquet 2002, p. 26 and p. 29). This dissociation between how

people look and who they are taken to be, coupled with imaging results, shows, as Schwartz

and Maquet neatly put it,

> that neuronal processes during sleep can *simultaneously and independently* engage
>
> (1) unimodal visual areas underlying the internal generation of a perceptual
>
> representation of an individual's face [i.e. "recognition"][…] and (2) distinct
>
> multimodal associative areas in the *temporal lobe* responsible for triggering the

---

[12] Like me, Gerrans (2012) uses dreams to shed light on delusional misidentification. He however, uses them to show that delusions should still be thought of as responses to experiences (viz. to support "the standard account" against "the revisionist account"). Although not incompatible with what I say here, the emphasis is very different. For me the main point is to distinguish recognition and identification, to see that they are epistemically very different, and (and here is where dreams are illustrative) psychologically independent.

[13] However, during sleep there is highly diminished activity in the frontal lobe (which is normally taken to perform supervisory control functions). This is perhaps why dream experiences can be extremely strange, and yet go completely unquestioned by the dreamer. We will also see that frontal lobe lesions or atrophy may contribute to delusions.

retrieval of a familiar individual's identity [i.e. "identification"]. (2002, p.26,

emphasis added)

Firstly, although simplistic, it is worth mentioning that the Capgras delusion is associated with damage to the right temporal lobe (see, e.g. Staton et al. 1982 and Johnson and Raye 1998), to which Schwartz and Maquet ascribe the role of "identity retrieval" (we will, in particular, see the relevance of this in section 5). Secondly, such a postulation of two distinct mechanisms for the tracking of numerical identity and qualitative similarity would suggest that, in principle, the two can come apart.[14] In the Capgras delusion there is recognition without identification (someone looks like a certain person but is taken not to be that person) so there ought in principle to be cases where there is identification, or at least *purported* identification, without recognition.

4.2.2. The Frégoli Delusion

We have seen that this can happen in dreams (with bearded man-sisters), but it can also happen, like with Capgras, following brain damage. The Frégoli delusion (that a known individual is taking on the different guises of surrounding strangers) involves precisely this: purported identification without recognition. Like in dreams, people are (wrongly) identified without the slightest overt justification. Feinberg et al. (1999), for example, report a brain-damaged patient, BJ, who :

…approached a severely disabled, wheelchair-bound patient in his early twenties

whom he had never met before, and claimed that the patient was his younger son. [...]

He maintained this misidentification despite clear differences in physical appearance

between the falsely identified patient and his son. Even when these distinctions were

---

[14] There are two readings of "identification": one is factive, so on this reading "identification" would mean successful identification. The other is non-factive, so identification could simply mean the activation of the identification mechanism, i.e. *purported* identification (which may or may not be correct). I use the factive reading and explicitly specify "purported" or "wrong", when I am referring to misidentification.

pointed out by staff, BJ maintained his original assertion that the patient was his son. (1999, p.378)

Both Capgras and Frégoli patients have no difficulty matching perceivable properties, namely, recognizing the *appearance* of thing. They just don't identify the people they misidentify on the basis of this matching. In the terminology I have introduced, both involve disorders of identification, not of recognition.

However, this raises the following question. If these patients were to base their judgments on appearances then they would get them right. The Capgras patient would judge that these people, who look just like her parents are her parents; the Frégoli patient would judge that this man who looks nothing like his younger son is not his younger son. So why aren't we designed to base our judgments of identification on appearances all the time? To answer this, we must consider the utility of that very same identification mechanism that wreaks such havoc in pathological cases.

4.3 Why the Bypassing?

The properties of an object that can be perceived (usually visually, of course) don't reliably singularize, as billiard balls, identical twins, or identical-looking animals abundantly demonstrate. In other words, you can be completely right about how something looks (or sounds), but very wrong about which individual it is. And this could be dangerous. Suppose I own two seemingly identical dogs, one is docile and the other is vicious: it is vitally important that I can *somehow* tell which is which. I may have developed the ability to do this, but am totally incapable of justifying (of having access to) how I do this.

What is inherently relevant is not whether this person looks a certain way, but whether they are the same individual as the one you encountered at a previous juncture. This will let you predict their personality, how they will behave (so you will know whether you should be relaxed or on edge) and will let you know what they know and what experiences they have shared with you (with loved ones, this can be rather a lot). The way an individual

18

looks or sounds is only indirectly relevant insofar as it is often a pretty reliable indication of identity. Most people, if their spouse went missing, wouldn't settle for someone who merely looked like (or even behaved like) their spouse. They would want their spouse: a certain individual with a certain past.

However, it is *prima facie* puzzling that the "property" of being an individual cannot physically trigger a tracking mechanism. It is not because your mother is a certain individual that she impacts on your nervous system in the way in which she does. It is because of her physical (e.g. visible, tactile, auditory, behavioral) properties. If someone were to recreate a duplicate of your mother, who looked, sounded, smelled, behaved etc, just like your mother, this duplicate would impact on your nervous system in the same way, in spite of being a different individual with a very different history.[15] You would, short of any evidence to the contrary, take this duplicate to be your mother.

Nevertheless, one must not confuse the causally efficacious properties that the tracking mechanism exploits, with the property tracked, which in this case isn't really a property, but an individual.[16] When tracking identity, our cognitive systems exploit all sorts of information (voice, mannerisms, appearance, smell etc.). But it can do so very automatically. When this happens we will judge that this person is a certain someone (or someone we've met, or someone we haven't met) without having consciously worked it out from her perceivable properties, and often without having post hoc access to why we have judged this.

Since taking an individual to be in your presence is (to pick up on the turn of phrase that Schwartz and Maquet use) a "multimodal associative" phenomenon, the modality that triggers the identification (or misidentification) needn't only be vision. It could be the

---

[15] Indeed, this is another perplexing thing about the Capgras delusion. Some patients admit that the impostor looks just like the replaced loved one (others confabulate slight differences). Surely then the patient would expect to be fooled by the impostor's perfect disguise!

[16] Unless you take "the property of being an individual" to be a property, but, as 3.1 aimed to show, this is merely a mirage set up by natural language. Indeed, both logically and metaphysically, properties and individuals (which are bearers of properties) are traditionally opposed.

dominant modality, or a modality that is sufficiently used to identify others.[17] This view therefore has scope for explaining reported cases of the Capgras delusion in blind patients (see Dalgalarrondo, Fujisawa and Banzato (2002) for a brief review), where the misidentification occurs on the basis of audition. In contradistinction, fully sighted patient DS (Hirstein and Ramachandran 1997) doesn't misidentify his father on the phone, but only visually (i.e. his dominant modality). Note also that the existence of one single blind Capgras patient is highly damaging to any theory that takes the delusion to necessarily arise from damage to neural pathways that are dedicated to the *visual* processing of facial information (e.g. Ellis and Young 1990).

Hirstein and Ramachandran (1997) and Hirstein (2010), rightly emphasize the difference between the perceivable properties of an individual, and our representations of them "from the inside". According to them, the Capgras delusion occurs when our "internal" representation of the person is damaged or inaccessible. This produces the impression of someone who looks right on the outside, but seems different on the inside (that is why Hirstein 2010 claims that these misidentification syndromes are mindreading disorders). It is illustrative to see how this differs from my view. On their view, we represent (Hirstein 2010 often refers to his view as "the representational hypothesis") two different kinds of *properties* of a person: their appearance, and what they are like "as a person." However, although this may well be true, for understanding delusional misidentification, I don't want to emphasize the difference between two different kinds of properties (or representations). I want to emphasize the difference between two *tasks*, namely, recognizing *properties* and tracking *individuals*. The latter is independent of any property, even the "internal" ones. I can, in principle, see someone as my mother, even though both her appearance and her personality have changed beyond all recognition. Tracking is just a logically, epistemically and psychologically different task. Hirstein's view also struggles to account for the delusional

---

[17] And indeed we needn't take "presence" too literally either. Consider what underpins your taking yourself to be talking to your mother on the phone. Or it may not even involve recognizing something uniquely produced by her, like the sound of her voice. Consider the warm feeling you get from receiving a text message from a loved one.

misidentification of inanimate objects (e.g. Alexander *et al.* 1979) (and there have been reports of Capgras-style misidentification for animals and places too). This seems to illustrate just the kind of distinction I'm getting at. These patients will say things like, "that's not my set of keys: it *looks* just like it, but it isn't it". They fail to re-identify (to track) previously encountered objects, just as Capgras patients fail to re-identify previously encountered people.

Let's recap. We have characterized IEA. We have looked at three problematic issues for IEA. Two are philosophical. One concerns logical differences between judgments of identification and recognition. The other concerns the epistemology of these judgments, namely, the sorts of things that can support them. Then there was an empirical issue, concerning, (i) the independence of the mechanisms underpinning these tasks and (ii) the non-inferential (i.e. non evidence-based) nature of some judgments of identification. Now I'd like to ask: what does this judgment actually amount to?

5 Introducing Identity Files

In order for a subject to identify an individual, the individual in question needs to have at least some *salience* for the subject *as an individual* (rather than, say, as a kind or a bearer of certain properties). Think of people who are not salient in the relevant sense, that you encounter once, casually and in passing. These people you might qualitatively categorize on the fly in terms of kinds and properties (girl, red hair etc.). But if I asked you to *identify* such a person, if I asked you "*Who* is this person?" you would rightly protest that the question makes no sense. Answering that question requires one to make the connection between the encountered person and somebody one has encountered previously, whether in person, on television, in the newspapers etc. When I ask you to identify someone who has no salience for you *qua* individual, I am asking you to perform an impossible task: I am asking you to make a connection between this person (under a perceptual, or perhaps short-term recollective, mode

of presentation) and a person at a previous encounter, but where there is no such previous encounter. On the other hand, walking past someone who looks like, say, John Cleese, it makes perfect sense for me to ask, "Is that John Cleese?" I'm asking, "Is the man we just walked past the very same man we are so familiar with from Monty Python?" Now what does this connection involve?

Some philosophers of language, when theorizing about reference to individuals use the somewhat metaphorical notion of a "mental file".[18] Mental files are created for individuals (upon first encounter) and retrieved (upon subsequent encounters) and are filled with information (viz. properties, predicates) which is taken to apply to the individual in question, but it is the file, or rather the acquaintance relation (which can be perceptual or through various informational chains, such as the media) on which the opening of the file was grounded, which fixes the reference.[19] The point is that a mental file can be virtually empty, or filled with largely inaccurate information, and still achieve reference.

Suppose I meet someone, John (call him John-1), and I am subsequently told information that I wrongly take to pertain to him (e.g. that he has climbed Everest), by somebody who unwittingly thinks that I met a different John (John-2). No amount of false information, even though that information accurately pertains to John-2, will mean that my beliefs (and other thoughts) cease to refer to John-1 and start referring to John-2. My belief about a given individual, that he has climbed Everest, is obviously a false belief about John-1, not a true belief about John-2. Namely, to whom I refer is determined by who I actually encountered and not who happens to fit the description I take to apply to an individual.[20]

The content of a file for an individual constitutes our conception of that individual. This conception will involve a variety of different kinds of information that is taken to pertain

---

[18] Strawson (1959), Evans (1982), Bach (1994), Perry (2001), Recanati (1993) all use something like mental files.

[19] The mental files theory can be seen as an elaboration of direct or causal theories of reference (as opposed to descriptive or satisfactional theories). It is not what happens have certain properties that fit a description, that is the referent of a singular term, but rather a particular individual that bears a certain causal relation to the use of the term.

[20] Mental files resemble the PINs (Personal Identity Nodes) postulated in several models of face recognition. But I do not intend them to be entities in cognitive modelling; rather they are abstract, philosophical entities that function at the level of *thought*.

to that individual; what they have done, when you have encountered them in the past, character traits etc., as well as what they look like. This conception, we have seen, can be false in many ways, and yet still be about that individual, since it is the initial encounter that caused the opening of the file that determines the referent. The conception, the contents of the file, can also be updated. Suppose my school friend, Jez, unbeknownst to me, went on to become a bank robber, underwent radical plastic surgery and fled to Australia. If I think to myself, "I wonder what happened to Jez?" I am thinking about that individual who is now in Australia with a radically different appearance (but, because I don't know this, I am not thinking about him under that mode of presentation). When I discover this information from a friend, I update my file for him. Rather, what it is to discover this *just is* to update my conception of him. If I go and see him in person and recognize his voice or his eyes or his manner, I update my conception of him in a more fine-grained and perceptual way. I momentarily open a file for "this man here present" but that gets very swiftly merged with the file for "My old friend Jez". The perceptual information I then get goes into that file and updates it. But in order for this to happen, the file needs to be correctly retrieved.

Put generally, judging that someone in your presence is a salient previously encountered individual involves the triggered retrieval of the correct file; correct in the sense that it is in fact the file that you had initially opened or created for *that very individual*. And it will be incorrect if you retrieve a file that wasn't opened for that individual (or fail to retrieve the file and erroneously open a new one) and that will be *mis*identification. The file contains all kinds of information and physical appearance is, firstly, only one kind of property among many, and secondly it is neither an essential nor a singularizing property (viz. A can change her properties, and others can look just like A). This means that a file can be retrieved in spite of the person looking superficially very unlike the person for whom the file was opened. All that needs to happen is that this person is seen as someone previously encountered, either on the basis of perceptual evidence (either voice, appearance, mannerisms), or indirect evidence (spatiotemporal trajectory) or, as suggested by our reflections on dreams and the Frégoli delusion, on the basis of a non-inferential, sub-personal tracking mechanism. The file can

also, as in the Capgras case, fail to be retrieved even though the person in question *does* look like the person for whom the file was opened. On this view, the delusional misidentification in the Capgras case consists in an inability to associate the loved one, and with this, her biography, with the person now present (who is *de facto* the loved one). This is made nicely explicit in Lucchelli and Spinnler (2007, p.189), who tell of the case of Fred and Wilma (both fictional names, of course). When Fred denies that Wilma is his wife, he cites that he "knew [her] very well as his sons' mother." Obviously, "mother of my sons" is not a perceivable property, but it is in the "Wilma file", which is failing to be correctly retrieved, and which he cannot associate with the woman here present.

Whereas the Capgras delusion illustrates a failure to retrieve a file (and, one might say, the erroneous opening of a new file) in the presence of someone who looks like the person for whom the file was opened (because they are that person!), the Frégoli delusion involves the erroneous retrieval of a file upon contact with someone whose appearance is – and, strikingly, is recognized by the subject to be – nothing like the person for whom the file was opened. It is worth noting that IEA, although *prima facie* plausible for the Capgras delusion, is less plausible for the Frégoli delusion. Presumably, since it appeals to diminished affect in the presence of loved ones to explain Capgras, IEA would need to appeal to *heightened* affect in the presence of strangers to explain Frégoli. But why should heightened affect lead to the judgment that one is in the presence of a *specific* person (as is the case in Frégoli delusion)? One would instead expect a certain "haven't I met you before?" response (a response we are all familiar with). On the file-retrieval model, the patient judges that it is that specific person because their cognitive system has retrieved or activated the file for that person.

6. Consequences of the Non-inferential File Retrieval View

What does the view that the delusional misidentification is non-inferential rule out? As I said, when I claim that delusional misidentification is non-inferential, I mean personal-level inference; the sort of inference that a person, and not a part of the person, does. To avoid a merely terminological dispute, let me outline what I take to be the concrete consequences of this view, and how it differs from existing views.

(i) Since the delusion isn't an inference on the basis of experiential evidence, it makes no sense to speak of subjects who have "the same experience" but who don't go on to form the delusion (Recall that this is where I part company with Bayne and Pacherie 2004, and Pacherie 2009).

(ii) These delusional patients will, at least potentially, perform normally in personal-level reasoning tasks. Personal-level reasoning biases aren't needed to explain the presence of the delusion.

(iii) The delusion will be something that the patient will be certain of when the perceived impostor/stranger is present. The subject will have no actual grounds for their belief, but may confabulate some *post hoc*, e.g. minor changes in appearance. As Hirstein 2010 points out, these are treated "as *confirmatory* of a preexisting belief" (p.246).

6.1 File retrieval versus Two-factor IEA

IEA claims that the patient adopts a delusional hypothesis on the evidential basis of an affectively anomalous experience. We have seen that such views, as they stand, plausibly fail (*pace* Maher) to account for the tenacity of the delusion. On the explanationist models, since delusions involve an inference to the best explanation of the anomalous experience, one would expect the delusional hypothesis pretty swiftly to cease to be the best explanation (viz. a *rational* person *ought* to abandon such a hypothesis). Similarly, on the endorsement model, one would expect the patient to reject what his visual experience is telling him. For example, on either explanationist or endorsement models, one would expect the following assurance on the part of the doctor to provide defeating evidence and hence cure the Capgras patient: "The

reason why you believe that this person isn't your mother is because you have had an accident that has damaged your brain and will make her feel unfamiliar/seem to you like she isn't your mother."[21] But it doesn't. As we saw, this has led the majority of IEA theorists to postulate a "second factor" in the form of a reasoning deficit.

There is a lot of truth behind the two-factor theories, in particular the claim that organic delusions are associated with two loci of damage (one temporal and one frontal). Suppose you had a Capgras patient, with the frontal and temporal damage, and you had another patient with the same temporal damage, but not the frontal damage. I am not claiming that the latter patient would have the Capgras delusion. I am not denying that frontal damage is necessary for the delusion. For this reason I am not unsympathetic to two-factor theories *per se*, but only to those that seem to imply IEA, or take it for granted. I will call such views "two-factor IEA". I oppose what two-factor IEA suggests, which is that the temporal damage should be thought of as determining the experiential evidence (factor 1), and the frontal damage as providing a reasoning deficit that operates upon that evidence (factor 2).

Proponents of two-factor IEA (e.g. Stone and Young 1997, Davies et al. 2001) tend to take two things to support their view. The first is that there are patients who appear to have a lack of affective response to familiar faces, hence that are *prima facie* similar to Capgras patients, but who do not go on to form the delusion (Tranel, Damasio and Damasio 1995). What is hypothesized is that these patients have the first factor ("the same experience"), but lack the second-factor reasoning deficit. In spite of this, these cases fail to offer good support, because, as I mentioned earlier, this sameness of experience is postulated on the basis of diminished SCR, which falls very short of being conclusive evidence for sameness of experience. Relatedly, their lesions are very different from those in Capgras (which tend to be associated with a combination of right lateral temporal lesions and dorsolateral prefrontal damage). These cases have ventromedial prefrontal damage, which is absent in Capgras

---

[21] As mentioned, this would be undercutting, rather than rebutting evidence.

patients.[22] In related fashion, such patients, although not delusional, are in many respects more globally disabled than Capgras patients. Unlike Capgras patients, they treat loved ones and strangers identically, and have severely deficient decision-making capabilities.

A second kind of support is more direct in claiming that delusional patients demonstrably exhibit reasoning biases. However, the cited experiments (Garety 1991, Huq et al. 1988, John and Dodgson 1994) all involve schizophrenia rather than brain-damaged patients. In patients with such a global psychiatric pathology as schizophrenia, one would expect general reasoning biases. As far as I know, there is no evidence that such deficits exist in Capgras cases that occur in the context of brain damage only.[23] Brain-damaged Capgras patients (i.e. ones that don't have schizophrenia) appear to be surprisingly reasonable and levelheaded. You can see it for yourself in interviews with them. That is part of what is so striking about the delusion.

On my proposal, since the identity file is retrieved (or fails to be retrieved) as the result of a non-inferential mechanism, the delusion is not the product of any personal level inference, biased or otherwise. This certainly fits with the accounts of Capgras patients. Every time they lay their eyes of their loved one, they cannot help judging that the perceived person is not the loved one in question. Now what the two-factor theorist may then be right about is that frontal damage (in particular to DLPFC) causes some kind of reality-testing deficit. This would play a role in explaining why the delusional belief, once non-inferentially formed, is not rejected. It is illuminating (see also footnote 15) to note that DLPFC is hypoactive during dreams, which may explain why we are so uncritical of our dream experiences. But, like I have suggested, this uncritical response can itself be conceived as part of the experience.

---

[22] One might, for example, hypothesize that although these patients have no difficulty making the correct identification, they lack the ability to associate the appropriate emotional salience to the episodic information that is telling them that they are in the presence of a loved one.

[23] Granted, McKay and Cipolotti (2007) document a case of an internalising attributional style in purely organic cases of Cotard delusion. This supports Young and Leafhead's (1996) hypothesis that Capgras and Cotard delusions may have the same 'first factor' (viz. an affective disruption) but a different second factor (viz. attributional style). Whereas Capgras involves an externalising attributional style ("she feels affectively flat because *she*'s not my mother"), Cotard involves an internalizing attributional style ("she feels affectively flat because *I* am dead"). As the vocabulary suggests, these are not *biases* in the sense of being misuses of evidence. Rather they are attributional *styles*, default (possibly mood-based) stances on the world.

6.2 Misidentification of Loved Ones Only?

It has often been noted that IEA has the attraction of explaining why Capgras patients only misidentify loved ones. Namely, the patient sees someone who looks like a loved one and expects that person to feel like a loved one (or to provoke an appropriate affective response). It is this frustrated expectation, this mismatch, that drives the delusional inference. This expectation is not present in the case of those who are not loved ones, and hence they are not misidentified. How does the file retrieval view account for this pattern?

The first thing to note is that many Capgras patients go on to misidentify far more than their loved ones. Ellis et al. (1997) mention a patient who was under the impression that almost everybody in a town had been replaced. Furthermore, IEA, as it stands, struggles to explain the cases where the duplicates themselves are duplicated (i.e. the impostors are replaced, as in the original case described by Capgras and Reboul-Lachaux (1923)). This cannot be the result of expected emotional response, since the subjects would not expect warmth from the impostor (indeed it is this lack of warmth that makes them judge it to be an impostor in the first place). In spite of these problems for IEA, the file retrieval view still needs to explain the fact that some patients seem to only misidentify their loved ones.

IEA cannot explain Frégoli delusion, and among Capgras cases can only explain ones where loved ones are misidentified, and only misidentified once. Thinking of delusional misidentification in terms of files and tracking mechanisms gives us the tools to explain a wider variety of cases. Patients who misidentify differing classes of people (e.g. only loved ones or all acquaintances) in either stable or reduplicative ways (the doubles are stable or reduplicated) will be understood in terms of differing damage to the mechanisms underpinning formation and retrieval of files. Thus the "classic" Capgras patient, who misidentifies one or two loved ones, and does so in a non-reduplicative way, can be understood as having difficulty in retrieving the file for, e.g., "dad", and therefore cannot judge that the person they encounter is that person, although the new "double" file stabilizes

and is retrieved whenever the father is seen (DS himself says, "He's a nice guy, doctor, but he's not my father" (Hirstein and Ramachandran 1997, p.438)). The patient who misidentifies a wider class of people, also in a non-reduplicative way, can be understood as having difficulties with the retrieval of more files, or perhaps with all previously created files. Reduplicative delusional misidentification, e.g. where even "the other dad" is judged to have been replaced, may be seen as a problem either with retrieving both old and newly created files, or with a deficit in the creation of stable files that can be retrieved. This is obviously speculative and needs to be fleshed out, but it shows that the framework is flexible enough to accommodate the existing cases.

Another case, that fits none of these profiles just mentioned, is another problematic case for IEA. Although patient DS (Hirstein and Ramachandran 1997), misidentifies his parents and correctly identifies mere acquaintances, he occasionally duplicates *newly encountered* people. This is problematic for IEA since, how could the expectation of positive affect drive the required inference, since, presumably, with new people there wouldn't be one?

The file retrieval view, on the other hand, can explain this. When DS meets someone for the first time, a file is created in the normal way for them and this gets filled with any associated information. But if this person leaves the room for 30 minutes and returns, instead of the old file being retrieved and added to, an entirely new file is created: DS behaves as though he is meeting somebody new. This is not the lack of recognition that you get with amnesiacs. On the contrary, he'll have no difficulty remembering that he met someone, and he'll remember what she looked like. He'll, for example, say: "I've just met your twin, and she was dressed just like you." But it seems that the timing has to be right for this unusual duplication to occur. Again, this is speculative, but we might postulate the following.

When you meet someone briefly and for the first time, your brain has a fast-learning mechanism that creates a very "thin" file (viz. one that contains very little information) and which enables you to re-identify that person. In DS's case, due to his brain damage, the file is either not created with sufficient stability, or is not retrieved: the result, either way, is that a

new file is erroneously created (he takes himself to be meeting this person for the first time). However, given enough time with that person, the file will become "thicker", and as a result will be more stable and/or easier to retrieve. In addition to this, however, there has been damage to his ability to retrieve the, extremely "thick", file for his parents. That is why, although DS duplicates both new people and his parents, he doesn't duplicate the intermediate category, i.e. when he sees people a) who are not newly encountered but b) who are not loved ones. And that is also why he doesn't duplicate, over and over again (i.e. the doubles themselves have doubles), everyone that he has met since his accident.[24] To illustrate, he doesn't duplicate those he has seen repeatedly or for a longer period of time, such as his neuropsychologist, whereas he might sometimes duplicate a cleaner or a nurse.

7. Conclusion

Let's take stock. We have characterized IEA, noted epistemological differences between judgments of identification and judgments of recognition. We have noted that a non-inferential route to identification, one that bypasses recognition (and spatiotemporal considerations), would be useful to any subject presented with the task of tracking individuals. We have seen from intuitions about meeting strangers, and from evidence from dreams and the Frégoli delusion, that such a route would not only be useful but actually seems to exist and be put to use in human cognition.

I would like to end by considering some general lessons from this. One lesson may be that we need to be careful by what we mean by inference. In particular, we need to distinguish the personal-level inferences that persons do, and that epistemology tends to deal with, and the "inferences" that a part of your brain does. In this latter sense, delusional misidentification is an "inferential" process. But it is not the person who is performing the inference, on the

---

[24] Interestingly, the first reported Capgras patient, Mme M. (Capgras and Reboul-Lachaux 1923) duplicated people over and over (i.e. the doubles themselves have doubles) – her husband as much as eighty times!

basis of an experience. In particular, it is unwarranted to make claims about patients (viz. VMPFC patients) having the "same experience" and not going on to form the delusion. As we saw, these claims are based on skin conductance response, which, physiologically, can be interfered with at various different places (indeed that these patients have different brain damage seems to support this) and, phenomenologically, tells us very little about the qualitative character of the experience.

Another lesson is methodological. Philosophical reflection can help us understand the nature of the judgment or task at hand. Here, for example, we see that judgments of identity (identification) are extremely different from judgments of predication (recognition). One would therefore expect the psychological routes to these judgments to be different. Furthermore, it is worth considering that, given that delusions involve tenacious belief that is not in keeping with the beliefs of others, one would expect the physiological disruption of a useful, automatic tracking mechanism that we all take for granted to be a pretty good source of delusional phenomena (although of course there are others). It also nicely accounts for, and provides a parsimonious tool for describing, the Capgras and Frégoli delusions. The former involves recognition without identification, whereas the latter involves (purported) identification without recognition.

Undoubtedly, more needs to be said about the mechanisms underpinning all this. My hope is that I have shown the starting point for an alternative to IEA that deserves serious consideration.

Word count:
10,067

References

Alexander, M. P., Stuss, D. T., & Benson, D. F. (1979). Capgras syndrome A reduplicative phenomenon. *Neurology*, 29(3), 334-334.

Bach, K. (1994). *Thought and Reference.* Oxford University Press.

Bauer, R.M. (1984). Autonomic recognition of names and faces in prosopagnosia: a

neuropsychological application of the guilty knowledge test, *Neuropsychologia* 22/4: 457-69.

Bayne, T., Pacherie, E. (2004). Bottom-up or top-down?: Campbell's rationalist account of monothematic delusions, *Philosophy, Psychiatry & Psychology*, 11/1: 1-11.

Breen, N., Caine, D., & Coltheart, M. (2000). Models of face recognition and delusional misidentification: a critical review. *Cognitive Neuropsychology, 17(*1–3), 55–71.

Capgras J., Reboul-Lachaux J. (1923). The illusion of "doubles" in a chronic systematized delusion (Illusion des « sosies » dans un délire systématisé chronique). *Bulletin de la Société Clinique de Médicine Mentale* 2: 6–16.

Clark, A. (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science" *Behavioural and Brain Sciences*. 36 (3)

Davies, M., Coltheart, M., Langdon, R. and Breen, N. (2001). Monothematic delusions: Towards a two- factor account. *Philosophy, Psychiatry & Psychology*, 8(2/3): 133–158.

Dalgalarrondo, P., Fujisawa, G., & Banzato, C.E.M. (2002). Capgras syndrome and blindness: Against the prosopagnosia hypothesis. *Canadian Journal of Psychiatry*, 47(4), 387-388.

Ellis, H. D., Young, A.W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157: 239-248.

Ellis, H. D., Young, A. W., Quayle, A. H., and de Pauw, K. W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society of London: Biological Sciences, B264,* 1085-1092

Evans, G., (1982). *The Varieties of Reference*, Oxford: Oxford University Press.

Feinberg, T., Eaton L. Roane D., Giacino J. (1999). Multiple Fregoli delusions after traumatic brain injury. *Cortex* 35 (3): 373–87

Garety, P. (1991). Reasoning and delusions. *British Journal of Psychiatry*, 159: 14–18.

Gerrans, P. (2012) Dream experience and a revisionist account of delusions of misidentification. *Consciousness and Cognition*, 21: 217-227.

Hirstein, W., Ramachandran, V.S. (1997). Capgras syndrome: a novel probe for understanding the neural representation and familiarity of persons. *Proceedings of the Royal Society of London*, 264, 437-444.

Hirstein, W. (2010) The Misidentification Syndromes as Mindreading Disorders. *Cognitive Neuropsychiatry*, 15 (1-3): 233-260

Hobson, J. (1999). *Dreaming as delirium: How the brain goes out of its mind*. The MIT Press.

Huq, S., Garety, P., Hemsley, D. (1988). Probabilistic judgments in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology*, 40(A): 801–812.

John, C.H., Dodgson, G. (1994). Inductive reasoning in delusional thinking. *Journal of Mental Health*, 3:31-49.

Johnson, M. K., and Raye, C. L. (1998). False memories and confabulation *Trends in Cognitive Sciences, 2.* 137-145

Lucchelli, F. and Spinnler, H. (2007). The case of lost Wilma: a clinical report of Capgras delusion, *Neurological Science*, 28(4): 188–195.

Maher, B.A. (1974). Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30: 98–113.

Maher, B.A. (1988). *Anomalous experience and delusional thinking: The logic of explanations*, in T.F. Oltmann and B.A. Maher (eds.) Delusional Beliefs, New York: Wiley, 15–33.

Maher, B.A. (1999). Anomalous experience in everyday life: Its significance for psychopathology, *The Monist*, 82: 547–70.

McKay, R. & Cipolotti, L. (2007). Attributional style in a case of Cotard delusion. *Consciousness and Cognition, 16*, 349-359.

Pacherie (2009) Perception, Emotions and Delusions: Revisiting the Capgras Delusion in T. Bayne and J. Fernàndez (eds.), *Delusion and Self deception: Affective and Motivational Influences on Belief Formation*, Hove: Psychology Press

Palermo, R. and Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*. 2007 Jan 7;45(1):75-92.

Perry, J. (2001). *Reference and Reflexivity*. Stanford, CA: CSLI Publications.

Pitt, D. (2004). The Phenomenology of Cognition, or, What is It Like to Think That *P*? *Philosophy and Phenomenological Research* 69 (1):1-36

Pollock, J. (1986). *Contemporary Theories of Knowledge*. (Towota, NJ: Rowman And Littlefield Publishers). 1st edition.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs* 34 (4):517–549.

Ratcliffe, M. (2008). The Phenomenological Role of Affect in the Capgras Delusion. *Continental Philosophy Review* 41 (2):195-216.

Ratcliffe, M. (2004). Interpreting Delusions. *Phenomenology and the Cognitive Sciences* 3 (1):25-48.

Recanati, F. (1993). *Direct Reference*. Cambridge University Press.

Reimer, M. (2009) Is the Impostor Hypothesis Really so Preposterous? Understanding the Capgras Experience. *Philosophical Psychology* 22 (6): 669 – 686.

Schwartz, S., Maquet, P. (2002). Sleep imaging and the neuro-psychological assessment of dreams. *Trends in Cognitive Science*, 6(1), 23-30.

Staton, R.D., Brumback, R.A., and Wilson, H. (1982). Reduplicative paramnesia: A disconnection syndrome of memory. *Cortex. 18, 23-36*

Stone, T., Young, A.W. (1997). Delusions and brain injury: the philosophy and psychology of belief. *Mind & Language*, 12: 327–364

Strawson, P. (1959) *Individuals: An Essay in Descriptive Metaphysics*. Routledge.

Tranel D, Damasio H, Damasio A. (1995). Double dissociation between overt and covert face recognition. *Journal of Cognitive Neuroscience* 7: 425–432.

Young, A.W. and Leafhead, K. (1996). Betwixt life and death: Case studies of the Cotard delusion, in P. Halligan and J. Marshall (eds.) *Method in Madness: Case Studies in Cognitive Neuropsychiatry*, Hove: Psychology Press, chapter 8.