UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

Is the Capgras delusion an endorsement of experience?

Bongiorno, Federico

DOI:
10.1111/mila.12239

License:
Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Bongiorno, F 2019, 'Is the Capgras delusion an endorsement of experience?', *Mind & Language*.
https://doi.org/10.1111/mila.12239

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 25. Apr. 2024

Received: 17 March 2018 | Revised: 14 November 2018 | Accepted: 3 February 2019

DOI: 10.1111/mila.12239

ORIGINAL ARTICLE

WILEY

Is the Capgras delusion an endorsement of experience?

Federico Bongiorno

Department of Philosophy, University of Birmingham, Birmingham, UK

Correspondence
Federico Bongiorno, European Research Institute, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.
Email: fxb606@student.bham.ac.uk

Funding information
United Kingdom Arts and Humanities Research Council Doctoral Grant, Grant/Award Number: AH/L50385X/1

Abstract

There is evidence indicating that the Capgras delusion is grounded in some kind of anomalous experience. According to the endorsement model, the content of the delusion is already encoded in the Capgras subject's experience, and the delusion is formed simply by endorsing that content as veridical. Elisabeth Pacherie and Sam Wilkinson have in different ways attempted to articulate a comprehensive defence of this strategy, but here I argue that the endorsement model cannot be defended along the lines envisioned by either of them. I then offer a more promising way of spelling out the model, according to which the anomalous experience implicated in Capgras is metaphorical in character.

KEYWORDS

Capgras, delusion, endorsement model, high-level perception, metaphorical-as perception

1 | INTRODUCTION

A prominent line of inquiry in the philosophical literature on delusion relies on the assumption that delusions are beliefs, and tries to work out how such beliefs are formed.[1] The Capgras delusion is often described as the belief that loved ones (most commonly close friends or relatives) have been

[1] For arguments that delusions are beliefs, see Bortolotti, 2010 and Bayne & Pacherie, 2004. A distinction is commonly made between monothematic delusions, where people exhibit one or more beliefs concerning a single subject, and polythematic delusions, where people exhibit diverse beliefs encompassing a variety of subjects (see Coltheart, 2013, for discussion). By the term "delusions" I shall henceforth always mean monothematic delusions.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2019 The Author. Mind & Language published by John Wiley & Sons Ltd.

replaced by identical or near-identical others (these can take on a variety of forms, such as imposters, clones, aliens, robots). A more accurate characterisation would distinguish between two propositions that the Capgras subject believes (Aimola-Davies & Davies, 2009): the proposition that the perceived person is not (though she looks like) the person whom they love (e.g., <this man is not my father>), and the proposition that someone else has taken the loved one's place (e.g., <this man is an imposter>). Call the content of the former proposition MISIDENTIFICATION, and the latter REPLACEMENT.

On one influential view, known as the *endorsement account* (Bayne & Pacherie, 2004), the content of the delusion is the same or very close to the content of the Capgras subject's experience, and the delusion is formed when the content of experience is taken as veridical and endorsed in belief. This view has been expressed in different ways. Some take the endorsed content to be limited to MISIDENTIFICATION, for example, "This is someone who looks just like my close relative *but is not really her/him*" (Davies et al., 2001, p. 150), "That [currently perceived] woman is not that [remembered] woman" (Campbell, 2001, p. 90), others take it to be directly that of REPLACEMENT (Fine et al., 2005, p. 148, see also Pacherie, 2009, p. 110; Pacherie et al., 2006, p. 2).

Regardless of which description one prefers, there is the problem of whether the Capgras experience can acquire the sort of rich misidentification or replacement contents that the model says it does. Let us call this the *experiential encoding* problem (following Langdon & Bayne, 2010). Two related problems have been pointed out. The first is the problem of specifying the causal connection between the pathologic mechanism underlying the Capgras delusion and the content of the experience that the delusion is supposed to be an endorsement of. I shall call this the *aetiology problem.* The second is the problem of explaining how the experience in question could be had without its content being inherited from the delusional belief. I shall call this the *top-down determination* problem.

Very few have tried to flesh out an endorsement account of the Capgras delusion that specifically addresses the problems above. A first comprehensive attempt has come from Elisabeth Pacherie (2009) who argues that the endorsement account can be defended (and the problems overcome) if the delusion is interpreted as due to a mindreading disorder, and if the processes through which the Capgras experience is generated count as modular in the Fodorian sense (Fodor, 1983). Pacherie builds on William Hirstein's (2005, 2010) "representational theory" of misidentification, which is based on a distinction between two distinct ways of representing people: (external) representations of a person's physical properties and (internal) representations of their mental properties. According to this theory, in Capgras there is a disconnection between these two forms of representation, with the result that a known person may look the same but appear different on the inside.

More recently, Sam Wilkinson (2016) has presented a model of the Capgras delusion that he suggests can help form the basis for an endorsement account, and it can do so by addressing the experiential encoding problem. Wilkinson thinks the problem can be avoided once it is appreciated that the recognition of qualitative similarity (*recognition*) and the tracking of individuals (*identification*) correspond to separate cognitive functions. This claim is motivated by the idea that (a) the Capgras experience can be explained in terms of a failed tracking mechanism and that (b) the tracking of individuals does not involve any sort of rich representational content.

In this paper, my aim is twofold. First, I show that the endorsement account cannot be defended along the lines proposed by Pacherie and Wilkinson. Second, I explore a new model of Capgras delusion that addresses the above problems while lending itself to an endorsement interpretation. I shall rely here on a view presented by Christopher Peacocke (2009), which says that we should appreciate that perception sometimes involve both literal perceptual and metaphorical contents. Drawing on this view, I shall propose that the experience implicated in Capgras is made up of two elements: the literal

perceptual content that <this [perceived] individual looks just like that [familiar] individual> and the metaphorical content that <this [perceived] individual is a replacer of that [familiar] individual>. In the following, the notations *x* and *S* will be used to refer to the perceived individual and the familiar individual respectively. According to my proposal, Capgras subjects treat the metaphorical content in a literal way, with the result that they come to believe the delusional proposition that *x is a replacer of S*. This is conducive to the idea that the replacement content and not the misidentification claim is encoded in experience.

I shall proceed as follows: In Section 2 I contrast the endorsement account to the explanationist account, according to which delusions are explanations rather than endorsements of experience. After discussing both options, I focus on the three challenges that confront the endorsement model. In Section 3 I present Pacherie's attempt to overcome these challenges. I demonstrate that Pacherie's proposal is committed to the truth of two claims: (i) that there is a legitimate sense in which perceptual experience can represent a seen person as unfamiliar; (ii) that perceiving a loved person as unfamiliar in this sense is the same experience as looking at an imposter. I show that both (i) and (ii) do not hold. In Section 4 I argue that even supposing these claims are true, they are still incongruous with Pacherie's plea for the modularity of Capgras experience. In Section 5 I lay out Wilkinson's approach to delusional misidentification, as applied to Capgras. I argue that the approach's appeal to a failed tracking mechanism rather than misrepresented properties makes it unclear exactly how misidentification is to be involved in the experiential content and, consequently, also which content is to be endorsed in the form of belief. In Section 6 I briefly sketch what I see as a more promising way of spelling out an endorsement account of Capgras, which conceives of the endorsed content as a metaphorical content of perception.

2 | MODELS OF DELUSION FORMATION

2.1 | Top-down versus bottom-up approaches to understanding delusions

Because delusions are often accompanied by anomalous perceptual experiences, one of the main questions arising is that of determining whether, and in what way, such experiences constitute a source of evidence for the content of the delusion. In this respect, we can distinguish between two ways of accounting for delusion formation. On what have been called *top-down* approaches (which I put aside here), the delusion is a direct product of organic malfunction, and the anomalous experience is a consequence rather than cause of the delusion (Campbell, 2001; Eilan, 2001). In contrast, the so-called *bottom-up* approaches view the delusion as originating in experience.

2.2 | The standard bottom-up account of Capgras delusion

Bottom-up views are the most prominent in the literature. The standard bottom-up account builds on Ellis and Young's (1990) suggestion that the Capgras delusion arises from a deficit in face processing. It is widely accepted that familiar face recognition correlates with heightened autonomic nervous system activity, which is measurable in terms of increased skin conductance. Ellis and Young propose that Capgras results because a neuropsychological deficit causes the face recognition system to become disconnected from the autonomic nervous system, such that one fails to discriminate autonomically between familiar and unknown faces (Ellis & Young, 1990). This hypothesis has been confirmed in a number of experiments (Brighetti et al., 2007; Ellis et al., 1997, 2000; Hirstein & Ramachandran, 1997). According to Ellis and Young, it is plausible that the abnormal absence of

autonomic response to the sight of a visually familiar face should generate some kind of unusual experience, based on which the person with Capgras forms the belief that the face seen is that of a stranger.

2.3 | Explanationist versus endorsement

There still remains uncertainty, however, as to the exact nature of the experience on which the delusion is grounded. On the explanationist option (Ellis & Young, 1990; Stone & Young, 1997; Maher, 2005), the content of the Capgras subject's experience is sparser than the content of the delusion (e.g., <*x* feels unfamiliar>), and the delusion is invoked as a potential explanation for the anomalous experience. On the endorsement option (Bayne & Pacherie, 2004; Fine et al., 2005; Pacherie, 2009), the experience directly represents <*x* is not *S*> or <*x* is a replacer of *S*>, and the delusion is acquired simply by endorsing the experience as veridical. Unless otherwise noted, in what follows "experience" will be used to mean, more strictly, literal perceptual experience.[2] For ease, I will call the explanationist model *ES*, and the endorsement model *EN*.

2.4 | Strengths and weaknesses

Both models come with characteristic strengths and weaknesses. One strength that is often claimed for EN is that it makes a plausible case for the degree of subjective certainty with which the delusion is held. If the delusion really is just a matter of endorsing the content of experience, that might explain why the Capgras subject believes in his delusion so firmly. On this account, the delusional conviction is onset at the moment the experience is unreflectively adopted into belief. By contrast, ES would seem to have a harder task explaining why the delusion is held with a strong sense of conviction. Some have noted that if the subject was aware of the delusion playing an explanatory role, she would also be aware of its need for justification. But, they argue, this would be inconsistent with the quality of self-evidence with which the delusion is maintained (Langdon & Connaughton, 2013).

Another suggested advantage of EN over ES is that it provides a rationale as to why the delusion has the particular content that it does. The specificity of content might be explained by the fact that the delusion is simply expressing the way things experientially seem to the subject to be. It follows that there is no (or little) gap to be filled here between the experiential content and the content of the delusion. By contrast, ES incurs the burden of explaining the relation between the coarse-grained content of experience and the fine-grained content of delusion. If the content of the Capgras subject's experience is simply that there is something unfamiliar about a currently perceived person *x*, where *x* looks exactly like a familiar person *S*, what makes them judge that *x* is not *S* or that *S* has been replaced by *x*? And also, why do they not explore a wider range of hypotheses before taking this route (Gold & Gold, 2014; Parrott, 2019)?

One point on which ES enjoys a clear advantage over EN is in specifying the content of the anomalous experience itself. Explanationist theorists are less demanding with respect to what is packed into the experiential content. All that ES requires is the awareness that there is something unusual or unfamiliar about the perceived person. Things are different when it comes to EN. To begin, if it is to be part of the representational content of a Capgras subject's experience that <*x* is

[2] Writers about EN typically individuate the anomalous experience by reference to its propositional representational content, and conceive of it in literal perceptual terms (see, e.g., Bayne & Pacherie, 2004; Davies et al., 2001; Pacherie, 2009). So the suggestion is that Capgras subjects have a perceptual experience with the literal content that <*x* is not *S*> or that <*x* is a replacer of *S*>.

not *S*> or that <*x* is a replacer of *S*>, it might seem that the experience would need to have a content into which *x* enters along with the property of *being numerically distinct from S* or the property of *being a replacer of S*, and this raises the problem of whether experiential contents can include such properties (Davies & Egan, 2013). The problem might be understood as EN helping itself to a controversial position on the admissible contents of experience, namely, that high-level properties of objects in our environment (which is to say, properties other than spatial location, colour, shape, motion, etc.) can be experientially represented (for detailed discussion, see Hawley & MacPherson, 2011).[3] Hence a central task for the endorsement theorist is to demonstrate that the Capgras experience can include representations with the kinds of high-level contents that their account would seem to require (Langdon & Bayne, 2010, p. 339).

It is also unclear how the finding suggestive of diminished autonomic response in face processing is supposed to relate to the endorsement framework. As John Campbell has pointed out (Campbell, 2001, p. 96), the mere diminishment of autonomic response to faces does not of itself constitute an experience with any particular content, let alone contents such as MISIDENTIFICATION or REPLACEMENT.

A different sort of objection to EN, also raised by Campbell (2001), is that the Capgras experience could represent high-level contents like MISIDENTIFICATION only as a result of a top-down loading from the delusional belief. If that were true, that would jeopardise the bottom-up commitment of EN, for which the experience is a causal antecedent of the delusional belief. For if the content of the subject's experience is inherited from the delusional belief, then the delusional hypothesis is not derived from experience, it is prior to experience. So, even on the assumption that the Capgras subject's experience can have the content that the account says it does, a further hurdle remains that must be overcome. Endorsement theorists need to explain how a perceptual state can have such a content without inheriting it from a belief with the same content (Bayne & Pacherie, 2004).

Two things should be noted. First, the significance of the objection is not tied to the viability of a top-down model. Whilst the truth of this model is incompatible with EN, recognition of its falsity is not logically sufficient for showing that the Capgras experience could deliver the content <*x* is not *S*> without inheriting it from the belief that the person before one is not really *S*. It may well be that the top-down model is false, but one still needs some principled basis for saying that experience can have *that* content without top-down loading from the corresponding belief. The top-down model and EN cannot be right at the same time, but they can both be wrong at the same time.

Second, endorsement theorists need not deny *any* top-down down influence on the experience. That is to say, they need not assume that the experience is insulated from any *other* cognitive features than the delusional belief nor from any other *beliefs* with different content. There is evidence that experience is the joint product of both top-down (or concept-driven) and bottom-up (or data-driven) processing, such that factors like priming, context, and memory-based expectation modulate incoming information from the level of sensory registration (Powers et al., 2017; Corlett, 2019). Therefore, it is possible that top-down and bottom-up processes are combined in various ways to give rise to the high-level content that EN requires. And considering that sensory registrations alone do not provide principles of numerical identity, and that the Capgras subject's sensory systems seem to be functioning properly, this is a likely possibility (but see Section 4).

To summarise, three main lines of objection have been urged against EN:

[3] In the philosophical literature on delusions, the problem is typically construed as a problem concerning the array of high-level properties that can be included in the literal contents of perception (although see Section 5). I will argue in Section 6 that the question about admissible contents of perception is less of a problem for endorsement theorists if the relevant content is thought of in metaphorical terms.

1. *Experiential Encoding Problem*: the problem of how experience can carry the sorts of contents that EN vindicates.
2. *Aetiology Problem*: the problem of linking the specific breakdown of Capgras to the relevant experience's having these contents.
3. *Top-down determination problem*: the problem of how experiences with such contents can be had without top-down loading from beliefs with the same contents.

In what follows, I consider Pacherie's endeavour to overcome concerns (1) to (3).

3 | PACHERIE'S PROPOSAL FOR AN ENDORSEMENT ACCOUNT OF CAPGRAS

3.1 | Capgras as a mindreading disorder

As we said, it is not obvious how a mere failure to exhibit autonomic discrimination between familiar and unfamiliar faces could plausibly be viewed as prompting the contents of experience that need accounting for in EN. Because of that, one might be inclined to think that Ellis and Young's model is not the best suited to accommodate EN. Pacherie tries to respond to this concern by surveying some alternative models for the cognitive and neural correlates of face recognition. She claims that at least one of these models makes a strong case for EN over ES, this being Hirstein's (2005, 2010) account of Capgras as a mindreading disorder (i.e., due to a malfunction of our mind-reading systems).

According to Hirstein, there are two neural systems devoted to understanding the behaviour(s) of others (Haxby et al., 2000). The first (medial temporal) pathway is responsible for producing what he calls "external representations" (representations of a person's facial and bodily features). The second (lateral temporal) pathway subserves the processing of "internal representations" (representations of that person's mental states).

Hirstein claims that the Capgras delusion could result from a dissociation between external and internal representations of a particular person, where the former are intact and the latter are damaged, inaccessible, or badly replaced (Hirstein, 2005, 2010). Subjects would be able to recognise the seen face of a loved one, but would present a deficit in activating the correct representation of their loved one's mind (e.g., characteristic emotions, moods, reasons for action, and beliefs). A representation would still be in play, but it would be other than the one subjects have used until then. The resulting state would then be the experience of a person retaining the same look but having a different inner world. As Hirstein puts it, "for the Capgras' subject, the familiar face is present, but the person is not" (Hirstein, 2010, p. 248).

According to Ellis and Young, what is primarily defective is the autonomic processing of familiar faces. By contrast, Hirstein thinks that the primary anomaly stems from the fact that the internal representation of a close person is inaccessible and replaced with a new one.[4] This would create the impression of "looking at an imposter," and the reduced autonomic responsiveness would be an effect rather than a cause: "The Capgras' subject is looking at someone who visually resembles his father, but who appears to have a different mental life, a different personality, with different dispositions to do different things. This is exactly what an impostor is, and this is exactly the experience one would have looking at an impostor" (Hirstein, 2005, p. 133).

[4] For Hirstein, the internal representation system would break as a result of damage to the orbitofrontal cortex and the temporal lobe, which are both thought to be implicated in mindreading (Hirstein, 2005, pp. 101–134).

Pacherie takes Hirstein's proposal to provide support for the plausibility of EN. The reason is that, she argues, in Hirstein's interpretation the subject's experience is not simply that of a coarse-grained feeling of unfamiliarity about a given person, but directly that of a seen person as "unfamiliar" in the sense of "different on the inside". That being so, the delusional belief could be interpreted not as an explanation of the experience, but as an endorsement of it (Pacherie, 2009, p. 116). There are two claims being made here: (i) experience can represent that a seen person is unfamiliar (or familiar); (ii) seeing someone who visually resembles a close person but has a different mind is experientially the same as looking at an imposter. I will argue that both (i) and (ii) are unjustified. Let me begin with (i). We saw earlier that the question of which properties are represented in experience is crucial for the plausibility of EN. Therefore, the proponent of EN could not assume (i) without further argument, as this is exactly what needs explaining. I take it that Pacherie sees (i) as following logically from Hirstein's approach (or at least from one of its possible readings). But, I argue, this interpretation is not correct.

3.2 | Unfamiliarity

According to Hirstein, we perform mindreading by simulating the mental states of another and thus by using our mind to put ourselves into their situations. He indeed concedes that in doing so we may take ourselves to directly perceive such mental states, when we are in fact only modelling them in our mind (Hirstein, 2010, p. 245). It might therefore be that people with mindreading disorders "misperceive" the mental state another is in, for instance by misconstruing their facial expressions of emotions. If the person whose mental states are being misconstrued is a familiar one, patterns of "misperception" recurring over time may well give rise to an experience of unfamiliarity or to the inference that the person in question has changed.

But from this claim it does not follow, at least not in any obvious way, that the person in question is being perceptually represented *as unfamiliar*. Neither can one exclude the possibility that the subject is simply having a non-perceptual feeling of unfamiliarity along with the perception of a person who looks just like herself. It is consistent with Hirstein's account that familiarity is an associated but distinct state that accompanies the perceptual experience without being part of what determines its content.

Pacherie might respond that a deficit in mindreading would not only affect the capacity to understand other people's mental states, but it could also interfere with the ability to extract the "dynamic signature" of a face (Pacherie, 2009, p. 113)—a term she uses to denote the distinctive movements that a face makes in expressing a particular mental state (e.g., surprise). She might argue that normally recognising the distinctive dynamics of a person's face is the same thing as perceiving that person literally as familiar. It would follow that failing to recognise the characteristic dynamics of a familiar person's face (e.g., your father's ordinary way of facially expressing surprise) would be the same as perceiving that person as unfamiliar. However, by assimilating familiarity to the dynamic signature of a face, Pacherie would overlook one important pattern that is found in Capgras delusion. As we know, subjects typically insist that the alleged imposter looks just like the replaced person, suggesting that the visual representation might have retained the same content as before the onset of the delusion (see, e.g., Hirstein & Ramachandran, 1997; Ramachandran & Blakeslee, 1998). If the identification of the dynamic signature was disrupted, we would expect subjects to give more details about how the imposter's way of animating her face would differ from the replaced person's one. This, however, is not the case.[5]

[5] Capgras subjects might sometimes refer to minor physical discrepancies between the purported imposter and the original person. A woman, for instance, claimed that she could tell her son had been replaced in that her real son "had different-coloured eyes, was not as big and brawny, and […] would not kiss her" (Frazer & Roberts, 1994, p. 557). Such reports, however, can be plausibly interpreted as post-hoc rationalisations for beliefs which are already held rather than proper perceptual reports (Bortolotti, 2010).

3.3 | Imposters

The second step of Pacherie's argument in support of EN is the claim that perceiving a familiar person as unfamiliar would equate to the experience one would have if one were looking at an imposter. Let us assume for the sake of argument that one could indeed perceive a person strongly resembling one's father as unfamiliar. Recall that by "unfamiliar" we mean here a person who is represented as having different personality traits (namely: different emotional states, moods, motives, desires, intentions, and beliefs). Would one thereby perceive that person as an imposter? This line of reasoning rests on a basic misconception of the distinction between qualitative and numerical identity. One can remain numerically the same individual despite the sometimes dramatic qualitative changes in body, character, and personality one has gone through.

This point is overlooked by Hirstein and Pacherie because they conflate the sight of a person resembling your father but with a different mind, and the sight of a person resembling your father but with a different identity (see also Wilkinson, 2015, p. 211; 2016, p. 393). The conflation is nicely epitomised by the following passage: "[…] The patient does not merely see someone familiar as unfamiliar, he perceives that person as having a different identity. […] This is because he sees them as no longer having the same mind, the same motives, moods, and emotions" (Hirstein, 2010, p. 244).

Now, one might speculate that if you were to perceive a person who looks like your father and claims to be your father as numerically distinct (i.e., having a different identity) from your father, you would indeed be looking at an imposter. Yet all that Hirstein's account shows is that Capgras subjects may suffer from mindreading-related perceptual failures; it does not show how this should lead them to perceive their familiars as being numerically different from themselves. One might then try to argue that perceiving a person simply as having a different mind from the person you originally knew would be enough for you to perceive that person as an imposter. But this is wrong; an imposter of *x* is not someone that merely looks like *x* and has a different personality from *x*, but rather someone who is numerically distinct from *x* and intentionally deceives you into thinking that she is.

If the above is correct, then neither of the claims Pacherie makes in defence of EN are warranted by Hirstein's account. It is not clear that such an account justifies the claim that perceptual experience might be able to represent a seen person as unfamiliar; and it is even less clear why allegedly perceiving some familiar person as unfamiliar should give you the experience of seeing that person as an imposter. If so, then Hirstein's mindreading deficit is not better suited than Ellis and Young's affective deficit to generate the sort of experiential content envisaged by endorsement theorists. Neither does Hirstein's account offer any clue as to how such content could represent high-level properties like that of *being an imposter*. As such, both the aetiology and the experiential encoding problem are left unsolved.

As we will see in the following section, the strategy Pacherie puts forth to address the top-down determination problem is even less promising. This is to hold that the processing through which the experience of familiarity is generated qualifies as modular in the sense advocated by Jerry Fodor (1983, 1985). For Fodor (1983), the essential criterion of modularity is informational encapsulation (p. 37). I will argue that the experience of unfamiliarity, as Pacherie conceives it, cannot be informationally encapsulated.

4 | THE MODULARITY OF FAMILIARITY

Fodor's modularity thesis is a claim about mental architecture. The idea is that the mind contains a set of domain-specific processing systems (*modules*), each of which is independent from one another

and devoted to performing specific tasks. Fodor requires modules be informationally encapsulated, where this means that the processing carried out within a module is insulated from any information stored elsewhere in the cognitive system. High-level cognitive functions such as reasoning or decision-making are not modules, whereas low-level peripheral systems such as the visual system are. One simple way to appreciate the difference is to note that whereas the latter take well-defined inputs and send well-defined outputs (in the case of vision, sensations of colours, shapes, edges, etc.), the former combines, elaborates, and syntheses the outputs from multiple sensory modalities as well as the information present in non-modular systems (e.g., information in memory).

As we saw, Pacherie conceives of unfamiliarity not as the mere absence of autonomic arousal in the presence of a particular person, but as the way in which that person is perceived. For Pacherie, the Capgras experience is not just that "an experience as of a person that looks like one's father but lacks the feeling of familiarity that normally accompanies this visual experience" but rather "an experience of the visually presented person as unfamiliar" (Pacherie, 2009, p. 116). For this reason, one might worry that the experience could have that content only as a result of a top-down effect from the delusion itself. As such, Pacherie's view is charged with addressing the *top-down determination* problem for EN. This motivates the appeal to modularity and informational encapsulation in turn. For to say that a system is modular in the sense of being informationally encapsulated is to say that it is impermeable to any top-down influence.

One striking example that Fodor cites in support of informational encapsulation is the persistence of perceptual illusions such as the Müller-Lyer illusion. In that illusion, two horizontal lines of equal length appear unequal because of the biasing effects of the arrowheads demarcating them. The bottom line still looks longer even if you believe the two lines to be equal in length.

Pacherie employs this very same example to illustrate her argument that the processes involved in producing the experience of familiarity or unfamiliarity are informationally encapsulated. In the same way as knowing that the two lines are equal will not make them look so, being assured that the person you are looking at is someone you know will not lead you to experience her as familiar if your first experience is of unfamiliarity (Pacherie, 2009, p. 117). This, Pacherie argues, may go towards explaining why the Capgras delusion is so steadily adhered to. If the Capgras experience was not informationally encapsulated, the experience of familiarity would be restored on the basis of background beliefs or the testimony of others, but that is exactly not the case (Pacherie, 2009, p. 120).

This line of argument has some weaknesses. Pacherie's claim about the modularity, and hence about the informational encapsulation, of the processes through which the experience of unfamiliarity is generated applies both to normal subjects, as well as to subjects with Capgras delusion. However, as far as normal subjects are concerned, informational encapsulation is not necessarily a good explanation of why the experience of unfamiliarity is sometimes not eliminated in the face of disconfirming information. An alternative explanation may be that some top-down loading of the experience is already in place, and that the disconfirming information simply is not strong enough to override it. Consistent with this, we can conceive scenarios in which an experience of unfamiliarity is not impermeable to disconfirming information.

Suppose someone approaches you on the train asking if you remember them. As hard as you try, you cannot remember and they are still unfamiliar to you. Seeing your perplexity, they offer you a hint. They tell you that you went to the same primary school together. Whilst you still cannot recognise them, they start to feel more familiar. You guess their name but you get it wrong. They give you a second hint: you played together on the same football team until you were 14. Now they have become fully familiar. You guess their name twice until you get it right. Here it seems we have a simple case where top-down cues can help to re-establish a feeling of familiarity that was not initially

there. If so, this provides a counterexample to Pacherie's idea that the experience of familiarity results from modular processes equally across normal subjects and subjects with Capgras delusion.

Yet it seems Pacherie could simply give up the claim about the modularity of familiarity in normal subjects, while retaining the view that this claim applies to subjects with Capgras. After all, the example above only shows that top-down influences can restore familiarity in normal subjects. It does not show that top-down influences can restore familiarity in people with Capgras delusion, for whom the absence of familiarity has a neurological origin. It is indeed possible that, where caused by a somatic condition, the experience of familiarity lies outside the control of top-down influences, in accordance with modularity and informational encapsulation.

However, there are still grounds for doubting that the processes responsible for the experience of unfamiliarity involved in the Capgras delusion, as Pacherie conceives it, can legitimately be understood in modular terms. Recall that Pacherie's conception of the Capgras experience is modelled upon Hirstein's view of mindreading. On this view, experiencing someone as familiar is a result of one's retrieving and employing an internal representation of that person's mental life. An internal representation aggregates a wide range of information that has been accumulated over time about who that person is "from the inside," which includes distinctive beliefs, desires, attitudes, and intentions. But this is at odds with the principle of encapsulation. The processing through which the experience of familiarity is generated depends heavily on stored memory and past interactions with the relevant person—and so is unlikely to be insulated from the influence of centrally accessible mental states. It follows that the difficulties raised by the top-down determination problem cannot be resolved by appeal to Fodorean modularity.

In the following section, I shall consider an alternative model of Capgras developed by Wilkinson (2016). Wilkinson's approach promises a straightforward way out of the problems associated with an endorsement interpretation of Capgras, but, as we will see, it is not clear that it helps support such an interpretation.

5 | WILKINSON'S MENTAL FILES APPROACH

5.1 | Recognition versus identification

Wilkinson's proposal rests on two principles: the distinction between recognition and identification and the notion that talk of "mental files" can be fruitfully used to understand how identification goes wrong in Capgras. Suppose you were asked to judge whether a currently perceived individual (x) is the same as one encountered in the past (y). When we speak of x and y being the *same* we might mean one of two things: that x is qualitatively identical to y, or that x and y are numerically identical.

Wilkinson argues that each of these readings reveal logically different tasks. If the former, then the question is whether you *recognise* that x is like y (e.g., that the car parked outside your house is like the one that caused a crash the day before). If the latter, then the question is whether you *identify* x as y (e.g., the car parked outside your house as one and the same with the car that caused the crash). For one to recognise that x is like y is, argues Wilkinson, is for one to judge that x has some property in which it resembles y. The judgment is of the form Fx (where F is the property being predicated and x is the individual of which F is predicated). In contrast, the act of identifying x as y simply involves making a connection of identity between x and y, where the judgment predicates no property at all and takes the form $x = y$.

For Wilkinson, this logical distinction between recognition and identification reflects two distinct cognitive functions: the perceiving of qualitative similarity and the tracking or perceiving of

numerical identity. Although judgments of identification are often arrived at by personal-level, abductive reasoning based on recognition of similarities or spatiotemporal continuity, this need not be the case. Indeed, for Wilkinson, there are routes to identification that are not evidence-based, and whose processing bypass both matching of qualitative similarity and spatiotemporal considerations (Wilkinson, 2015, p. 212).

For example, if you had two identical-looking dogs, Fido and Scotty, you might be able to keep track of which is which independently of the properties you perceive them as having, without tracing the spatiotemporal path of either dog, and while being unable to articulate how you do so. Note that Wilkinson is not denying that properties and spatiotemporal principles are efficacious in the process that yield judgments about numerical identity. Rather his point is that such properties do not need to constitute evidence for such judgments, in that they are not playing any personal-level inferential role in their formation.

5.2 | Mental files

In Wilkinson's approach, the concept of a "mental file" (Perry, 1980; Recanati, 1993, 2012) serves as a framework within which to understand what happens cognitively when tracking someone's identity. To a first approximation, mental files are conceptions we build up of others based on information gathered about them. Whenever we meet someone for the first time, a mental file is created and filled with the information available. At each new encounter, the same file is retrieved and updated with newly discovered information.[6] Mental files are "non-descriptive singular mental representations" (Wilkinson, 2016, p. 397). This is not to say that they do not contain any descriptive information about the individuals to which they refer. It is rather that the correct mental file *F* for an individual *x* is retrieved regardless of whatever descriptive match there is between the content of *F* and the qualities possessed by *x* at the time. Successful retrieval occurs upon encountering the very same individual for which the file was originally created, no matter the extent to which its properties or location have changed since the first encounter (Wilkinson, 2016).[7]

Wilkinson (2016) draws a distinction between "Demonstrative" and "Stable" files (p. 398). The former are context-specific files that ensue from the here-now of a perceptual encounter with a given individual and which take something like this form: <here now this [person]>. The latter are files which contain a "stable enough conception" of a person and which can be brought out by reflection outside the immediate perceptual context (e.g., "Mom").

According to Wilkinson, whenever we encounter someone who is familiar to us, we create a new demonstrative file, retrieve the stable file we have for that person, and merge them together. The processes of creating, retrieving, and merging files are not conscious ones. What is conscious is only the outcome of this chain of processes, namely a state with the content <this person here present is the very same person as my mother> (Wilkinson, 2016, p. 398). Wilkinson suggests that misidentification (at least of the kind involved in Capgras) occurs when files are mismanaged such that

[6] Relevant information may come in different forms, such as biographical details, physical appearance, and personality characteristics.

[7] Wilkinson's view is similar but different from Hirstein's. According to the latter, the Capgras subject fails to activate the correct mental representation when looking at their loved one, and that gives the impression of a person being different on the inside. On this view, misidentification occurs because the wrong mental properties are attributed to the person. In contrast, on Wilkinson's view, it is not the content of a mental file that fixes the reference to the person, as the information stored in the file does not singularly identify—one can, in principle, identify someone as the same person over time despite major qualitative changes. Rather, reference is fixed causally, not descriptively by matching of qualitative similarity, and this is supposed to explain why misidentification can occur independently of attributed properties.

the correct stable file fails to be retrieved in the presence of the person for whom the file was first created. This would cause the subject to judge that the person who is perceptually present and looks like the one they have a stable file for is not *that* person.

5.3 | Endorsement of what?

Now, what has this got to do with an endorsement interpretation of Capgras? Wilkinson (2016) thinks of the experiential encoding problem specifically as the problem of whether a person's identity can be represented or misrepresented in experience, that is, the problem of whether one "can perceive (or fail to perceive) a person's identity directly, without inferring it" (p. 399).

However, Wilkinson argues, the problem arises for EN only if we assume that numerical identity and numerical distinctness are properties that can be given to us in experience. In the case of Capgras, that would involve not only explaining how experiences can convey information about numerical identity, but also how one can experience the numerical distinctiveness of qualitatively indistinguishable individuals. If I show you two indistinguishable faces before and after a certain interval of time, nothing in the properties you perceive will tell you whether you saw the same face twice or two qualitatively indistinguishable but numerically identical faces. So how can a Capgras subject perceive the numerical distinctiveness of someone who is perceptually indistinguishable from the familiar?

Wilkinson's strategy is to argue that the identity of an individual is neither worked out from their perceivable properties nor itself a perceivable property, but rather something that is tracked prior to, and independent of, any properties the individual may experientially appear to have. Against this background, Wilkinson (2016) writes, the experiential encoding problem no longer exists, since "there's nothing mysteriously rich about the content of the Capgras experience […]" (p. 400).

We might also think that the aetiological and the top-down determination problems no longer seem quite so onerous for the endorsement theorist. As for the former, the relation between reduced autonomic arousal and conscious experience may be explained by reference to the failed tracking mechanism. That is, the lack of autonomic response may compromise file-retrieval, which in turn may cause a content-specific misidentification experience. As for the latter, because tracking is the condition under which identification is possible, it seems reasonable to exclude that it could be itself contingent on judgments of identification.

While Wilkinson may well be right that identification does not necessarily pass through the representation of high-level properties, it is still not clear then how identity can enter the content of experience, and if so, what form such a content would take. For this reason, although Wilkinson's view has great potential as a self-standing account of Capgras, it seems to me that it does not offer much help in terms of working out an endorsement account. Let me explain.

As I understand it, Wilkinson's approach is compatible with either of two different interpretations. On option (1), the physical properties of a certain individual (who is *S*) impacts on one's nervous system, leading to a subpersonal tracking mechanism. Due to a tracking deficit, the information that the individual is not *S* is extracted and the subject has an experience as if that (the individual) is not *S*. On option (2), the information that the individual is not *S* is passed straight on to a misidentification belief, without being first parcelled up into experiential content. We have seen that endorsement theorists share a commitment to the idea that the Capgras experience has representational content, and that such a content can be given by a proposition which specifies the way things are represented as being (e.g., <*x* is not *S*>). An implication of this is obviously that features that enter experiential content are features of the way things appear to the conscious subject. So it seems that if one wants to

say that experience delivers a content in which <*x* is not *S*> features, then one needs to make reference to the way things appear from the subject's point of view. But Wilkinson rejects this on the grounds that misidentification can be had regardless, and even despite any observable and appearance properties that things may be experienced as having.

The question then arises in what way misidentification is experiential and also, more specifically, how it can be an experience whose content is available as something to be endorsed. Wilkinson's (2016) analysis is couched entirely in negative terms. In saying that the content of Capgras experience is not rich, he does not offer many specifics as to what an experience with misidentification content is like, nor about what form this content takes. Wilkinson is explicit that people are not consciously aware of their tracking mechanism, and also that people can keep track of individual things regardless of the ways things are for them experientially speaking (e.g., what and where they are). But if so, what constitutes an experiential way of tracking numerically distinct individuals? And more importantly, what makes a state have a content such that it can be endorsed in belief? Sometimes it seems that Wilkinson wants to appeal to a doxastic view of experience, according to which experiences are belief-like states that coerce one into believing rather than presenting one with a content that can be in principle be rejected (p. 402). Yet this at best explains the way in which the content of experience is entertained. It does not show what form it takes. What is the difference between the content of an experience in which a subject misidentifies *x* as ~ *S* and the content of one in which a subject correctly identifies *x* as *S*? Without such a clarification, it is hard to see how (or even that) the mental files approach does support an endorsement interpretation of Capgras. Indeed, as it stands, option (2) seems to me the only fully intelligible way to understand how misidentification can emerge into consciousness on Wilkinson's account.

6 | CAN YOU SEE SOMEONE AS AN IMPOSTER?

The aim of this section is to develop an alternative way of understanding what constitutes endorsement in Capgras. Endorsement theorists have tended to characterise the content of experience as including at a minimum MISIDENTIFICATION, namely the claim that <*x* is not *S*>. As we have seen, Wilkinson's strategy does not really address the central question of what kind of experiential content MISIDENTIFICATION is supposed to be, and this makes it unclear why we should suppose that there is any experience at all prior to the formation of the misidentification belief. The fact that conscious experience may play no role in the emergence of the misidentification belief does not threaten Wilkinson's (2016) overall account of delusion misidentification. As he himself concedes, "perhaps by the time we get to conscious experience the misidentification is already there" (pp. 402–403). But if MISIDENTIFICATION is not an experiential content to be endorsed, this leaves two possibilities for the endorsement theorist: either abandon the idea of endorsement altogether, or suppose that REPLACEMENT alone is endorsed. In what follows, I explore this latter possibility. I first explain why I think REPLACEMENT is not a perceptual content in the literal sense (a content the experience represents *literally* as such), and then consider in what sense it might be metaphorical in character (a content the experience represents *metaphorically* as such). For simplicity, I will use here "imposter" as a broad term to refer to any form of replacer.

6.1 | Seeing someone literally-as an imposter

One straightforward sense in which we experience *x* as *F* is that in which we perceive something literally as an instance of a kind, as when we see a squirrel *as a squirrel*; or, when the lighting

conditions are not optimal, the squirrel *as a rat.* Several philosophers have regarded this species of experience as involving the activation of conceptual capacities (e.g., Brewer, 1999; McDowell, 1994). In a recent paper Peter Carruthers (2015b) remarks that we can consciously see something as something of a certain kind only if "the concept that represents that kind is bound into the object file that nonconceptually represents its other properties" (p. 503).[8] As will become clear shortly, the reason for focusing on this conception is that it offers not only a neuroscientifically informed framework for establishing which properties can be literally represented in perception, but one which allows the property of being an imposter to be among them.

According to Carruthers, the binding of a concept *F* into your perception of a certain object *x* is what allows you to perceive *x* literally as *F.* This is illustrated by reference to the well-known image of a Dalmatian dog in a black-and-white speckled background (e.g., Marr, 1982). As Carruthers points out, even if you know that the image represents a Dalmatian dog (assuming that you have never encountered the image before), it will take you a little time to recognise the set of black blobs as a Dalmatian. What happens during that time interval is that you are having an experience that non-conceptually represents that there are such-and-such black patches while concurrently entertaining the thought that the image depicts a Dalmatian dog. Carruthers's (2015b) suggestion is that you are able to perceive the blobs as parts of a Dalmatian only when the concept *Dalmatian* and the non-conceptual representation of the blobs come to be bound together into a single object-file (p. 503).

This story is consistent with emerging insights from neuroscience. For instance, a study by Wyatte, Jilk and O'Reilly (2014) presents data illustrating that feedback from higher-order visual areas implicated in semantic representation—specifically in the inferotemporal cortex (IT)—begin having an influence on object recognition-related tasks as early as 100 ms after stimulus onset, which is well before slower top-down attentional processes come into play (at around 200 ms after stimulus presentation). Perhaps the best illustration of this comes from scenarios in which the missing parts of partially occluded objects are filled in by the visual system.

Wyatte and colleagues introduce, as an example, a case where IT neurons respond to an occluded bicycle stimulus (i.e., with nothing in sight but the wheels). In the circumstances only a small fraction of the neurons will fire—those associated with wheel-shaped features—whereas the remainder population will remain dormant. However, IT neurons will send feedback signals back to earlier visual areas, activating neurons that selectively respond to visual features associated with bicycle wheels (e.g., a bicycle's frame, seat post, and saddle) despite the absence of relevant physical stimuli. The consequence is that IT neurons respond as if the bicycle was fully visible, such that "there is little-to-no difference between the response to the partially occluded object and the complete object" (Wyatte et al., 2014, p. 4).

Carruthers (2015b) cites cases like the filling-in case above as evidence that conceptual information is used by perceptual systems to test the input signal from lower visual areas, so as to estimate the best conceptual fit for the non-conceptual content of an object-file (p. 502). Now, if Carruthers is right that concepts can be bound into the literal content of perception, then the question arises, what are the limits of abstractedness of the concepts that can be so bound.

According to Carruthers (2015b), the only restriction that applies is that the applicability of the concept in question must be processed quickly enough for binding to take place: "conceptual

[8] The idea of *object files* has been used in vision science to denote location-based indexes whose role is to track the identity of an object across space and time (e.g., Kahneman, Treisman & Gibbs, 1992). Carruthers (2015a) suggests that object files incorporate both conceptual and non-conceptual representations of the object's properties (p. 66). The content of the object-file <*that*> produced by, say, the perception of an approaching car might be non-conceptual to some extent (yellow, curvy, moving), but conceptual to another extent (new, fancy, car).

information will need to be processed within the window of a few hundred milliseconds that elapses between presentation of a stimulus and its subsequent global broadcast [i.e., its becoming available across a wide range of cognitive systems]" (p. 504). The upshot is that if the applicability of a concept *F* is processed with sufficient speed, then it must be possible for someone to perceive something literally as *F*.

Perhaps the endorsement theorist can draw on such a possibility to argue that the content <*x* is an imposter> can be part of the literal content of the Capgras subject's perception. Imagine the following situation: as you enter your father's room, you catch a stranger with your father's mask rolled up to his forehead. If we accept Carruthers's account, there is no principled reason to exclude that the concept *imposter* in the circumstances can be processed fast enough to become integrated into the perceptual state itself. However, this is not the scenario Capgras subjects find themselves in, and so no generalisation can be had. It is true that the appearance of the misidentified person's face is not inconsistent with her being an imposter in disguise (a good imposter would presumably try and look as close as possible to the replaced person). But as opposed to the case of the masked man, there is nothing in the behaviour of the misidentified person to indicate that she is an imposter.

Based on this, we could presume that the concept to be bound into the literal content of perception is the one corresponding to the identity of the person in question (say, *father*), and that the processing of the concept *imposter* occurs at a later stage (after the perceptual output is broadcast into the impaired affective appraisal stage). If I am right about this, Capgras subjects do not perceive the misidentified person literally as an imposter, and whatever content their perceptual experience may literally represent, it is not REPLACEMENT.

6.2 | Seeing someone metaphorically-as an imposter

The only necessary condition that must be fulfilled for an endorsement interpretation to be felicitous is that there be an experience whose content is being endorsed in belief. As Susanna Siegel has nicely put it, "an experience is endorsed when one forms a belief that *P* on the basis of an experience whose contents include *P*" (Siegel, 2016, p. 107). This need not entail any commitment as to what sort of mental state experience is. So, although the experience in question is most often regarded by endorsement theorists as being literally perceptual, this is not a requirement of EN per se. That means that while REPLACEMENT is not literally perceptual in character, it may still serve as a content of experience and be available to one as something to be endorsed.

One option would be to pursue the idea that there are perceptual experiences whose contents include *P* but where *P* is not literally represented. This would allow REPLACEMENT to become incorporated into one's perceptual state without itself being literally perceived, which is to say, regardless of whether imposters or the like are represented in the literal content of perception. Do such experiences exist? In what follows, I shall give the outline of an approach which could underpin an affirmative answer.

In a 2009 essay titled *The Perception of Music: Sources of Significance*, Christopher Peacocke sets himself the task of explaining how it is that we are able to experience what we might call expressive qualities of music: "We can experience music as sad, as exuberant, as sombre. We can experience it as expressing immensity, identification with the rest of humanity, or gratitude". His main thesis is that "when a piece of music is heard as expressing some property *F*, some feature of music is heard metaphorically-as *F*" (Peacocke, 2009, p. 257; Peacocke, 2010). Before dealing with the case of auditory perception, Peacocke refers to a painting of pottery jars by Francisco de Zurbarán, which he considers an illustration of his thesis. He argues that the painted pots are seen by many *as a*

group of people, and that this is a matter of seeing them metaphorically-as such. For convenience, I will limit the discussion to this visual case, though the points we are about to make should be projectable mutatis mutandis across all sensory modalities.

There are three things about Peacocke's view that are worth noting. First, Peacocke thinks that the phenomenon to which we refer when we say that the pots are seen as people is genuinely experiential—rather than, say, just entertaining a thought about people while looking at the painting. For, he argues, it seems phenomenologically that you could look at four items on your desk while imagining that they are four people without having the distinctive experience that you have when you look at Zurbarán's pots.

Second, the experience is acquired through the same mechanism underlying metaphorical cognition. This is a subpersonal detection of isomorphism between two domains, each made up of concepts, properties and relations. When the isomorphism is detected, items in a source domain (say, the domain of emotions) are mapped onto items in another domain, to generate contents such as <the sky is angry> or <this music is sad>. In the experience of seeing pots metaphorically-as people, the item *people* is copied from a metaphorically represented domain to a "special kind of storage," which binds it to any representation involved in perceiving the pots' properties (skinny, elongated, cone-shaped, fragile). Peacocke recognises that what is designated by the phrase "special kind of storage" remains open to empirical investigation. But he remarks that this storage has to be special, in the sense that it must be different from the sort of storage that would make one see the pots literally as people (Peacocke, 2009, p. 267).

Third, Peacocke proposes that the concept *people* enters into the perceptual experience of the pots, only not as part of the literally represented content. What something is seen metaphorically-as, say *people*, does not contribute to the correctness of the experience, namely it is partitioned off as anything entitling the subject to judge that the perceived object really is a group of people. So although there is a sense in which the depicted pots are subsumed under the concept *people*, it should not be treated as a predicative sense, for predicative subsumption is typically belief-yielding, whereas this kind of subsumption is not (Boghossian, 2010, p. 73). Rather, Peacocke says that in perceiving the painting we non-predicatively subsume the concept of those pots under the concept *people*, where by "non-predicative subsumption" he just means a mapping from one concept to another under the sort of isomorphism described above (Peacocke, 2010, p. 189).

Given all this, we can read Peacocke's proposal as describing a class of perceptual experiences where an object *x* is *literally* represented as *Q* while being *metaphorically* represented as something else, and where both representations contribute their contents to the overall phenomenology of the experience. Applied to Capgras, this could introduce a new perspective on what constitutes an endorsement way of forming the delusion, and one in which the problems facing endorsement theorists may be satisfactorily settled. The central hypothesis would be that a person is in a state in which *x* looks to be *S*, but where *x* is experienced metaphorically-as an imposter. By this way of thinking, the imposter belief could still qualify as an endorsement of content, but that would be the content of a metaphorical rather than literal perceptual representation.[9]

More work would need to be done, but we could view the metaphorical content as a byproduct of misidentification. Rather than thinking that the failed tracking mechanism only generates the judgment that someone is not a certain known individual, we might hypothesise that this also produces a perceptual experience with the metaphorical content <this [perceived] person is an imposter of that [familiar] individual>. Perhaps it happens that when the subject is confronted with a person who

[9] Of course, this raises the question how it is that a metaphorical content is endorsed in belief. Although a full treatment of this issue is beyond the scope of this paper, a preliminary indication of how it might be addressed is provided below.

looks exactly like *S* but is misidentified as ~ *S*, there is a mapping from the domain *imposter* to the domain ~ *S* which occurs in a manner similar to that required in metaphorical experience. On this sort of view, the subpersonal processing responsible for the failed tracking mechanism has then two conscious outcomes: a misidentification belief that the encountered person is not a certain known individual and a metaphorical experience as of that person being an imposter. The idea would be that while the misidentification belief is acquired through a non-experiential source, the replacement belief is acquired because the metaphorical content is treated as literal and endorsed as such.

This proposal may hold the key for dissolving all the three problems that we have used to assess endorsement proposals, namely the experiential encoding problem, the aetiology problem, and the top-down determination problem. As for the first, there are no strict limits to the range of properties that something can be perceived metaphorically-as possessing, provided that there is a functioning isomorphism between the perceptual and the metaphorical representations. As for the second, one might view the dysfunctional tracking mechanism as the medium by which the absence of autonomic response at the sight of *x* could bring about a metaphorical experience with the content that *x* is an imposter. On this picture, the mapping from *imposter* to ~ *S* (which underpins such experience) would be produced in response to a situation in which some *S*-looking person is misidentified as ~ *S* (due to defective autonomic responsiveness). As for the third problem, it is clear that one need not believe that a perceived object is really *F* in order to perceive it metaphorically-as *F*. I do not have to believe the pots are people in order to perceive them metaphorically-as such. So it seems that Capgras subjects could have a metaphorical experience with REPLACEMENT as content without that needing to be inherited from prior beliefs.

Now, even if it is true that the proposal under consideration avoids all three of the above problems, why suppose that the experiences Capgras subjects are having are metaphorical in nature? Support for this comes from considerations about the differing ways in which Capgras subjects are disposed to act with respect their replacement claims. While in some cases they express hostile and even violent attitudes towards the putative imposters (for a gruesome example, see De Pauw & Szulecka, 1988; see also Christodolou, 1977), in other cases they are friendly or indifferent towards them (as with patient Fred and his wife Wilma, see Lucchelli & Spinnler, 2007). The metaphorical framework can make sense of both kinds of subject by offering an explanation of why some fail to act in a way consistent with their assertions. This would be explained by the fact that metaphorical contents of experience are typically make-believed, not believed. That is, it may be that a significant reason why some subjects do not feel hate or aversion towards the "imposters" is that they do not actually endorse the metaphorical content as true; rather, perhaps, they use it as a springboard to create stories about the identity of the misidentified people, which result in claims about these being as the metaphorical experience represents them to be.

But then why would some other subject really believe that their loved ones are as they are metaphorically represented to be? Given that typically metaphorical experience does not lead to belief, why would this happen in Capgras? Fully answering this question is a difficult task which cannot be undertaken here, but some speculations can be made. Perhaps one becomes prone to believe that things are as metaphorically represented when one cannot believe anything that counts as good evidence against it. Imagine somebody who perceives *x* as looking like her mother, but who cannot help believing that *x* is not her mother, such that she dismisses any counterevidence to what she believes. Since *x* really is her mother, all of the available evidence is consistent with the way things look and against her belief. In these circumstances, it is arguable that one might be at least somewhat inclined to endorse the metaphorical representation of *x* as an imposter, because this (contrary to the weight of evidence) is consistent not only with *x*'s looking like her mother, but also with her belief that *x* is not really her mother. If so, it may be that the difference between those who go on to endorse the

content and those who do not depends on local psychological dispositions that are not uniform across all Capgras subjects, such as perhaps paranoid personality traits like distrust of others and unjustified suspicions of being deceived.

Of course, one might point out that an endorsement account along these lines is far more modest than it is standardly taken to be. Indeed, the misidentification judgement is not formed from endorsement processes, which means that the endorsement hypothesis cannot explain the delusion in its initial formation. However, my proposal locates the explanatory power of the endorsement hypothesis elsewhere. That is, although appeal to endorsement does not explain why Capgras subjects adopt the initial misidentification belief it does explain the trade-off between those who act in a way consistent with their replacement claims, and those who do not. The Capgras delusion is equally importantly characterised by the misidentification of familiar people and their replacement by doubles. Both misidentification and replacement claims acquire prominence in the Capgras subject's mental life and are reported sincerely and with conviction. Where I differ from other writers on the subject is in the idea that the explanatory power of the endorsement hypothesis is relative to the latter feature rather than the former.

7 | CONCLUSION

Only two attempts, by Pacherie and Wilkinson, have been made to systematically articulate an endorsement interpretation of the Capgras delusion. The first part of the essay considered two such attempts and argued that neither is a viable option. Pacherie's strategy does not escape the problems it was intended to, whereas Wilkinson's fails to provide a principled basis for determining what the endorsed content is to be. The second part of the essay introduced a new way of thinking about the experiential endorsement in Capgras, according to which the content endorsed in the delusional belief is metaphorical rather than literal. The claim that metaphorical experience is what is occurring in actual cases of Capgras is very speculative and certainly liable to empirical refutation. Peacocke's model itself requires considerable development before we can decide whether it is empirically adequate. Also, more work would be needed to show how exactly one could form a belief that familiar people are imposters simply on the basis of perceiving them metaphorically-as such. Despite these limitations, postulating a role for metaphorical experience in Capgras is not incompatible with any of the facts we know to be true about this condition, and has the benefit of allowing a novel way in which an endorsement interpretation of Capgras might plausibly proceed.

ACKNOWLEDGEMENTS

I acknowledge the support of a United Kingdom Arts and Humanities Research Council Doctoral Grant. Many thanks to two anonymous referees for *Mind & Language* for helpful comments and suggestions. Many thanks to audiences at the Society for Philosophy and Psychology at Johns Hopkins University, and the European Society for Philosophy and Psychology at the University of Hertfordshire, where earlier versions of this paper were presented. Thank you also to Peter Carruthers and Christopher Peacocke for conversations that considerably improved Section 6 of the paper. Finally, I owe a special debt to Lisa Bortolotti, Ema Sullivan-Bissett, Craig French and Matthew Parrott for extremely valuable feedback on previous drafts. The paper is dedicated to my father, Francesco Bongiorno, with gratitude.

ORCID

Federico Bongiorno https://orcid.org/0000-0002-1961-5852

REFERENCES

Aimola-Davies, A. & Davies, M. (2009). Explaining pathologies of belief. In M. Broome & L. Bortolotti (Eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives* (pp. 285–323). Oxford: Oxford University Press.

Bayne, T. & Pacherie, E. (2004). Bottom-up or top-down? Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry, & Psychology*, *11*, 1–11. https://doi.org/10.1353/ppp.2004.0033

Boghossian, P. (2010). The perception of music: comments on Peacocke. *British Journal of Aesthetics*, *50*(1), 71–76. https://doi.org/10.1093/aesthj/ayp050

Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford: Oxford University Press.

Brewer, B. (1999). *Perception and reason*. Oxford: Oxford University Press.

Brighetti, G., Bonifacci, P., Borlimi, R. & Ottaviani, C. (2007). "Far from the heart far from the eye": Evidence from the Capgras delusion. *Cognitive Neuropsychiatry*, *12*, 189–197. https://doi.org/10.1080/13546800600892183

Campbell, J. (2001). Rationality, meaning, and the analysis of delusion. *Philosophy, Psychiatry, & Psychology*, *8*(2/3), 89–100. https://doi.org/10.1353/ppp.2001.0004

Carruthers, P. (2015a). *The centered mind: What the science of working memory shows us about the nature of human thought*. Oxford: Oxford University Press.

Carruthers, P. (2015b). Perceiving mental states. *Consciousness and Cognition*, *36*, 498–507. https://doi.org/10.1016/j.concog.2015.04.009

Christodolou, G. N. (1977). The syndrome of Capgras. *British Journal of Psychiatry*, *130*, 556–564. https://doi.org/10.1192/bjp.130.6.556

Coltheart, M. (2013). On the distinction between monothematic and polythematic delusions. *Mind and Language*, *28* (1), 2013–2112. https://doi.org/10.1111/mila.12011

Corlett, P. R. (2019). Factor one, familiarity and frontal cortex: A challenge to the two-factor theory of delusions. *Cognitive Neuropsychiatry*, *24*, 165–177. https://doi.org/10.1080/13546805.2019.1606706

Davies, M., Coltheart, M., Langdon, R. & Breen, N. (2001). Monothematic delusions: towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, *8*(2/3), 133–158. https://doi.org/10.1353/ppp.2001.0007

Davies, M. & Egan, A. (2013). Delusion: cognitive approaches–Bayesian inference and compartmentalization. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini & T. Thornton (Eds.), *The Oxford handbook of philosophy and psychiatry* (pp. 688–727). Oxford: Oxford University Press.

De Pauw, K. W. & Szulecka, T. K. (1988). Dangerous delusions: Violence and the misidentification syndromes. *British Journal of Psychiatry*, *152*, 91–96.

Eilan, N. (2001). On understanding schizophrenia. In D. Zahavi (Ed.), *Exploring the self* (pp. 97–113). Amsterdam: John Benjamins.

Ellis, H. D. & Young, A. W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry*, *157*, 239–248. https://doi.org/10.1192/bjp.157.2.239

Ellis, H.D., Young, A.W., Quayle, A., & de Pauw, K. W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society: Biological Sciences, B264*, 1085–1092. https://doi.org/10.1098/rspb.1997.0150

Ellis, H. D., Lewis, M. B., Moselhy, H. F. & Young, A. W. (2000). Automatic without autonomic responses to familiar faces: Differential components of covert face recognition in a case of Capgras delusion. *Cognitive Neuropsychiatry*, *5*(4), 255–269. https://doi.org/10.1080/13546800050199711

Fine, C., Craigie, J. & Gold, I. (2005). Damned if you do, damned if you don't: the impasse in cognitive accounts of the Capgras delusion. *Philosophy, Psychiatry, & Psychology*, *12*, 143–151.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1985). Precis of the modularity of mind. *Behavioural and Brain Sciences*, *8*, 1–42. https://doi.org/10.1017/S0140525X0001921X

Frazer, S. J. & Roberts, J. M. (1994). Three cases of Capgras' syndrome. *British Journal of Psychiatry*, *164*, 557–559. https://doi.org/10.1192/bjp.164.4.557

Gold, J. & Gold, I. (2014). *Suspicious minds: How culture shapes madness*. New York, NY: Free Press.

Hawley, K. & MacPherson, F. (Eds.) (2011). *The admissible contents of experience*. Malden, MA: Wiley-Blackwell.

Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

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

Hirstein, W. (2010). The misidentification syndromes as mindreading disorders. *Cognitive Neuropsychiatry*, *15*(1), 233–260. https://doi.org/10.1080/13546800903414891

Hirstein, W. & Ramachandran, V. S. (1997). Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society of London B: Biological Sciences*, *264*, 437–444. https://doi.org/10.1098/rspb.1997.0062

Kahneman, D., Treisman, A. & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 174–219. https://doi.org/10.1016/0010-0285(92)90007-O

Langdon, R. & Bayne, T. (2010). Delusion and confabulation: Mistakes of perceiving, remembering and believing. *Cognitive Neuropsychiatry*, *15*(1/2/3), 319–345. https://doi.org/10.1080/13546800903000229

Langdon, R. & Connaughton, R. (2013). The neuropsychology of belief formation. In F. Kreuger & J. Grafman (Eds.), *The neural basis of human belief systems* (pp. 19–42). Hove; New York, NY: Taylor and Francis.

Lucchelli, F. & Spinnler, H. (2007). The case of lost Wilma: A clinical report of Capgras delusion. *Neurological Science*, *28*(4), 188–195. https://doi.org/10.1007/s10072-007-0819-8

Maher, B. (2005). Delusional thinking and cognitive disorder. *Integrative Physiological & Behavioural Science*, *40*(3), 136–146.

Marr, D. (1982). *Vision. A computational investigation into human representation and processing of visual information*. San Francisco, CA: Freeman.

McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press.

Pacherie, E. (2009). Perception, emotions and delusions: Revisiting the Capgras delusion. In T. Bayne & J. Fernandez (Eds.), *Delusions and self-deception* (pp. 107–126). Hove: Psychology Press.

Pacherie, E., Green, M. & Bayne, T. (2006). Phenomenology and delusions: "Alien" in alien control? *Consciousness and Cognition*, *15*(3), 566–577. https://doi.org/10.1016/j.concog.2005.11.008

Parrott, M. (2019). Delusional predictions and explanations. *The British Journal for the Philosophy of Science*, 1–32. https://doi.org/10.1093/bjps/axz003

Peacocke, C. (2009). The perception of music: Sources of significance. *The British Journal of Aesthetics*, *49*(3), 293–297. https://doi.org/10.1093/aesthj/ayp016

Peacocke, C. (2010). Music and experiencing metaphorically-as: Further delineation. *The British Journal of Aesthetics*, *50*(2), 189–191. https://doi.org/10.1093/aesthj/ayq005

Perry, J. (1980). A problem about continued belief. *Pacific Philosophical Quarterly*, *61*, 317–332. https://doi.org/10.1111/j.1468-0114.1980.tb00023.x

Powers, A. R., Mathys, C. & Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science*, *357*(6351), 596–600. https://doi.org/10.1126/science.aan3458

Ramachandran, V. S. & Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind*. New York, NY: William Morrow.

Recanati, F. (1993). *Direct reference: From language to thought*. Oxford: Blackwell.

Recanati, F. (2012). *Mental Files*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199659982.001.0001

Siegel, S. (2016). *The rationality of perception*. New York, NY: Oxford University Press.

Stone, T. & Young, A. W. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language*, *12*, 327–364. https://doi.org/10.1111/j.1468-0017.1997.tb00077.x

Wilkinson, S. (2015). Delusions, dreams, and the nature of identification. *Philosophical Psychology*, *28*(2), 203–226. https://doi.org/10.1080/09515089.2013.830351

Wilkinson, S. (2016). A mental files approach to delusional misidentification. *Review of Philosophy and Psychology*, *7*, 389–404. https://doi.org/10.1007/s13164-015-0260-5

Wyatte, D., Jilk, D. & O'Reilly, R. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Frontiers in Psychology*, *5*(674), 16–25. https://doi.org/10.3389/fpsyg.2014.00674

How to cite this article: Bongiorno F. Is the Capgras delusion an endorsement of experience? *Mind Lang*. 2019;1–20. https://doi.org/10.1111/mila.12239