



# Making sense of the doxastic approach to thought insertion

Pablo López-Silva<sup>1,2,3,4,5</sup> 

Received: 17 January 2024 / Accepted: 22 July 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

## Abstract

With a higher prevalence in schizophrenia, delusions of thought insertion (TI) are regarded as one of the most severe symptoms of psychosis. Patients suffering from TI report that external agents are able to place thoughts into their minds or skulls. A version of the doxastic approach characterizes delusions as abnormal beliefs rooted in anomalous experiences. Nonetheless, the exact role of these experiences in determining the content and the way in which delusional beliefs are fixated is still under debate. While *endorsement* models claim that the abnormal experience comprises the very content of the delusional beliefs, *explanationist* approaches claim that delusional beliefs emerge as explanations for abnormal experiences with less specific content. This paper combines conceptual analysis with phenomenological data to examine the merits of both endorsement and explanationist approaches to TI. I propose that potential solutions to the dispute could lay in finding a middle ground between the two approaches, and non-exhaustive ways in which hybrid doxastic approaches to TI could be formulated.

**Keywords** Delusions · Thought insertion · Psychosis · Doxastic approach · Endorsement · Explanationism

---

✉ Pablo López-Silva  
pablo.lopez.silva@gmail.com

<sup>1</sup> School of Psychology, Universidad de Valparaíso, Valparaíso, Chile

<sup>2</sup> Research Centre in Cognition and Language Development, Universidad de Valparaíso,, Valparaíso, Chile

<sup>3</sup> Institute of Complex Systems of Valparaíso, Valparaíso, Chile

<sup>4</sup> Millennium Institute for Research in Depression and Personality (MIDAP), Valparaíso, Chile

<sup>5</sup> LATAM Network Lead, The Collaborating Centre for Values-based Practice in Health and Social Care, St Catherine's College, University of Oxford, Oxford, UK

## 1 Introduction

With a higher prevalence in schizophrenia, delusions of thought insertion (TI) are regarded as one of the most severe symptoms of psychosis (Mullins & Spence, 2003; López-Silva, 2018; López-Silva & McClelland, 2023; APA, 2024)<sup>1</sup>. Patients suffering from TI report that external agents are able to place thoughts into their minds (Mellor, 1970; Mullins & Spence, 2003; López-Silva, 2018). In a commonly discussed case, a subject reports that “thoughts are put into my mind like ‘Kill God’, it’s just like my mind working, but it isn’t. They come from this chap, Chris. They are his thoughts” (Frith, 1992, p. 66). Another person with psychosis reports “I didn’t hear these words as literal sounds [...] Instead, the words just came into my head – they were ideas I was having. Yet I instinctively knew they were not my ideas. They belonged to the houses, and the houses had put them in my head” (Saks, 2007, p. 29). A case reported by Mellor (1970, p. 17) shows a patient claiming that Eammon Andrew (a popular TV presenter) was able to flash and screen thoughts such as like “the garden looks nice” and “the grass looks cool” into his mind.

Delusions such as TI – specially in schizophrenia – are preceded and accompanied by a number of transformations in the experience of the subjective, intersubjective and physical world (Mayer-Gross, 1932; Conrad, 1958; Payne, 2013, 2015; López-Silva, 2018; López-Silva et al., 2022b; Mishara et al., 2023). In the phenomenological tradition, ‘delusional atmosphere’ (*Wahnstimmung*) is the period lasting from days to months preceding the co-emergence of full-blown cases of TI and other positive psychotic symptoms (Conrad, 1958). During the months preceding an episode of TI, BS – a person with schizophrenia diagnosis – experienced a generalized sense of unreality, a general transformation of the social and subjective world, stating: “I had short periods of time in which I felt like I didn’t exist. I had other experiences in which I had to, for instance, touch a coffee table in front of me to make sure it was real” (López-Silva, 2018, p. 3). In light of its unique features, both psychiatrists and philosophers have found a formidable challenge when trying to come up with a plausible story about how TI and other delusions are formed and fixated in a subject’s mind (Coltheart et al., 2011; Coltheart, 2002, 2015; Sterzer et al., 2016, 2018; Mathieson, 2023).

Over the last years, the *doxastic approach* has become a popular view within analytic philosophy of mind, serving as the main conceptual framework for current neurocognitive models of delusional phenomena (Bayne & Pacherie, 2005; Bortolotti, 2010; Bayne, 2010; López-Silva, 2016; Clutton, 2018; Sterzer et al., 2018; Connors & Halligan, 2020). Neurocognitive models are crucial for the development of our understanding of the etiology of delusions and, therefore, their clinical treatment. A dominant formulation of the doxastic approach involves two claims. First, delusions are an abnormal type of belief, and, second, the proximal cause of delusions is “a highly unusual experience” (Bayne & Pacherie, 2004, p. 2; see also, Coltheart, 2002, 2015)<sup>2</sup>. Defenders of this *bottom-up* approach infer a causal relationship between

<sup>1</sup> In this paper, I will discuss delusions of TI in the context of schizophrenia.

<sup>2</sup> This claim – popular in the context of the analytic philosophy of mind – seems consistent with the picture of delusions in schizophrenia offered by the phenomenological tradition; delusions could be viewed as

experience and belief, going from the former to the latter<sup>3</sup>. However, there is contention around the specific role that these grounding experiences have in determining the content of the delusional belief reported by the patient (Hohwy & Rosenberg, 2005; Pacherie et al., 2006; Coltheart et al., 2011; Sollberger, 2014; Bongiorno, 2019).

Advocates of the *endorsement model* suggest that the relevant abnormal experience comprises the content of the delusional beliefs reported, so patients simply *believe* or *endorse* what they have experienced (Bayne & Pacherie, 2004; Sollberger, 2014). In contrast, advocates of *explanationist models* suggest that subjects adopt delusional beliefs as a way of explaining abnormal experiences that do not have the same content of the underlying doxastic state (Davies, Coltheart, Langdon & Breen, 2002). For explanationists, underlying experiential states in delusions formation do not comprise the entire content of the finally reported delusional belief. The idea is that, in explaining the abnormal experience, patients would add certain elements to the underlying experiential states as a way of making sense of them. Therefore, the delusional belief with the content, for example, ‘my bodily movements are controlled by aliens’ would arise as an explanatory way of making sense of bodily experiences with poorer or less specific content (for instance, bodily movements felt without any type of volitional control). Finally, as Langdon and Bayne (2010) claim, ‘the explanationist account of the route from experience to a reflective delusion locates most of the inferential processing downstream of experience, in the processes of hypothesis generation and evaluation, [...] the endorsement account locates it upstream of experience, before the received delusion is endorsed’ (p. 331).

Doxastic approaches have emphasized the analysis of delusional phenomena from a third-person perspective by focusing on the criteria that external observers typically apply when ascribing belief-like states to another person (López-Silva, Núñez de Prado-Gordillo, Fernández-Castro, 2024). These models - typically found where delusions have been approached within analytic philosophy of mind - have been criticized for not integrating the lived first-person perspective of the subject expe-

---

abnormal beliefs that are fixated in a subject’s mind in the context of those phenomenologically altered conditions (López-Silva, Núñez de Prado, Fernández-Castro, 2024). However, over the years a number of phenomenologists have criticized the doxastic approach for not grasping the transformation of reality that delusional phenomena involve (see Sass, 2004; Feysaerts et al., 2021). On this approach, delusions would be symptomatic of a general and all-enveloping transformation in the relationship between the self and reality (Sass & Pienkos, 2013, among others). It is not entirely clear why phenomenologists insist in this harsh separation. As Graham (2015) rightly suggests, delusional phenomena could be clearly understood as a stance towards reality composed by abnormal beliefs, existential feelings, phenomenological changes, and so on. As I have already suggested elsewhere, claiming that some delusions involve certain alterations of the structure of experience and consciousness is not incompatible with claiming that they are beliefs, taken as endorsements or explanations for such experiences (López-Silva, Núñez de Prado, Fernández-Castro, 2024). It is important to note that the doxastic approach is trying to come up with an answer to the question about the type of mental state that underlies reports we label as delusional; therefore, approaches to the subjective first-personal structure of delusional reports do not need to be inconsistent with those 3rd personal approaches (Clutton, 2018).

<sup>3</sup> Contrasting with bottom-up approaches, *top-down* views take delusional beliefs to be something like Wittgensteinian framework propositions. The claim is that top-down disturbances in fundamental beliefs of subjects would infuse their conscious experiences and actions with an abnormal character. In this sense, delusional experiences could not have their contents without top-down loading (Campbell, 2001, p. 96).

riencing the actual phenomenon (Lysaker, Buck & Lysaker, 2012; Fernández, 2019; Fusar-Poli et al., 2022; Ritunnano, Humpston & Broome, 2022; Faeyerts et al. 2021). Delusions are more than a mere conceptual artefact able to debunk certain theories about the human mind; they are conscious occurrences had and suffered by people. An exclusive focus on third-person analyses could lead to the hyper intellectualization of deeply embedded and embodied processes, obscuring the full breadth of TI phenomena. Neglecting the type of element that could only be retrieved by inspecting first-personal data could lead to the construction of an oversimplified *explanandum*, and therefore, faulty explanatory theories in applied fields such as neuropsychiatry and neuropsychology.

In the remainder of this paper, I combine conceptual analysis with phenomenological data in order to examine the merits of both the endorsement and the explanationist approaches to TI. First, after clarifying some preliminary issues, I explore how these approaches can make sense of the level of subjective certainty and ambivalence characteristic of delusional reports. After, I focus on two specific problems. Bottom-up doxastic models of delusions of TI need to account for, at least, two things when it comes to the content of delusions, namely, (i) the content of the abnormal cognitive experiential state that grounds the delusional belief and, (ii) the content of the doxastic state (belief). I have called the former ‘the experiential encoding problem’ and the latter the ‘the doxastic encoding problem’. The challenge is, therefore, to explore whether the phenomenology of TI and other conceptual resources support the view that patients have a first-order experience of a thought *as* being inserted by an external agent (from which they go on to form the respective delusional belief) or whether the delusional belief arises as a way of making sense of cognitive experiences with abnormal phenomenal features. Finally, I show that both alternatives can partially resist the main objections that they face, and that potential solutions to the dispute is in finding a middle ground between the two approaches. In the final section, I explore non-exhaustive ways in which hybrid doxastic approaches to TI could be formulated.

## 2 Preliminary issues: room for comparison

Most of the arguments supporting explanationist models of delusions have been formulated by examining the role of perceptual states related to the content of certain delusions (Pacherie, Green & Bayne, 2005)<sup>4</sup>. This is why explanationist accounts have become very popular in the discussion of perception-based delusions, such as Capgras and Cotard delusions (Young & Leafhead, 1996; Bongiorno, 2019). However, Sollberger (2014) suggests that arguments cannot be directly applied to undermine a potential defense of an endorsement approach to TI, because this delusion is

<sup>4</sup> It is important to note that the distinction between *experiencing* and *believing* in the endorsement and explanationist models does not entail any distinction between conscious and non-conscious operations. It's not the case that subjects either *consciously endorses* or *consciously explain* such experiential states. The distinction refers to the way in which the architecture of the content of the reported delusion is organized. Delusions - as reported by patients - pop into a subject's stream of consciousness as an already manufactured and imposed mental state so conscious cognitive operations might not necessarily play a constitutive role in their early formation. Here I shall remain neutral about this issue.

grounded in *introspective-cognitive experiences* and not in perceptual experiences. There would be a fundamental difference in the nature of these two types of experience. The endorsement model claims that the belief that someone else is inserting a thought into the mind does not refer to experiences of mind-independent worldly objects and their properties as in the case of perception; rather, the “the subject matter of thought insertion experiences are conscious mental events that the subject is first-personally aware of” (Sollberger, 2014, p. 597). The difference is that “the awareness-relation at stake holds between the subject and her own mental state. It is an instance of consciously based self-awareness and not, as in perception, of world-awareness” (Sollberger, 2014, p. 598). The suggestion here is that, while perceptual experiences need direct contact with the objects of one’s awareness, cognitive experience do not. However, the issue does not concern the type of *relatedness* of these two types of mental states (although in some circumstances this issue might be an important one); rather, it concerns their *representational nature*, and given that both cognitive and perceptual states are representational states, it is plausible to think that room for comparison can be allowed.

The advocate of the endorsement model could suggest that *what it is like to perceive* is very different from *what it is like to think*, and, in virtue of this difference, one should maintain a clear distinction between perceptual and cognitive states when approaching our target debate. In this sense, the differences between perceptual and cognitive states become clearer, but again, one might be able to say that the shared representational nature of these two types of experiences seem to allow, although with some constraints, some parallelisms. At the same time, caution is needed if arguments supporting the discussion of the nature of cognitive experiences rely exclusively on features that are paradigmatic of perceptual states (for example, direct contact with the content of the perception). An example of this is the problematic parallelism between *mental* and *bodily* actions, when discussing the sense of agency understood as willful generation of thoughts or bodily movement (Frith, 1992; Gallagher, 2004, 2014)<sup>5</sup>. When applied to bodily movements, agency as wishful generation *seems* to make sense (De Hann & De Bruin, 2010). In fact, we might distinguish between non-wishful bodily movements (like spasms) and bodily actions in virtue of the existence of a sense of agency embedded in the very phenomenology of the latter. However, such notion of agency does not seem to match the common phenomenology of thinking: “the default mode of thinking is precisely not an explicit and willful generation of thoughts” (De Hann & De Bruin, 2010, p. 383; see also Mishara et al., 2023). It is not my intention to argue for the implausibility of this parallelism further here<sup>6</sup>; the point is that, although one might be able to apply some arguments from the discussion of other types of mental states to the explanationist-endorsement debate of TI, this move should be always justified. At times, comparisons would work; at some others, they might not.

<sup>5</sup> For an empirically and philosophically well-informed discussion of many issues related to this parallelism, see: Proust (2009).

<sup>6</sup> For a discussion of this idea, see Proust (2009), López-Silva et al. (2019), and Mathieson (2023). I have recently developed this issue in López-Silva (2024).

A second issue is that the defender of the endorsement model seems to require the phenomenology of thinking to be distinguishable from the phenomenology of perceptual states in a very distinctive way. However, clear phenomenological distinctions become problematic sometimes (especially when examining schizophrenia). Take the case of *synaesthesia*, i.e. the experience of ‘seeing sounds’ or ‘smelling colours as odours’, etc. (Blackemore et al., 2005). Research suggests that experiences of this kind are empirically possible and not pathological (Hubbard & Ramachandran, 2005). What this type experience shows is that it is possible that different experiential modalities can overlap under certain circumstances. Now, in the case of thinking the same seems to apply, at least, in the context of psychosis. As Kusters (2020, p. 83) points out:

Usually, for the unmad person, thoughts have no taste at all, but with madness thoughts become more sensual; they might appear in color, for instance. Thoughts take place in a thinking area; they are large or small, sharp, round, soft, or hard. They also move; they take leaps and leave lines and tracks. Some thoughts feel heavy, while others feel very light. Thoughts become more physical; they can drag you along, and you can feel them racing through your body. You can even get them to flow out of your head and through your hands or the top of your skull.

Similarly, Cahill and Frith (1996) refer to a patient that claimed to have physically felt the alien thoughts as they entered his head, and claimed that he could pinpoint the place of entry (p. 278), supporting the idea provided by the phenomenological tradition in psychiatry that the type of cognitive experience leading up to delusions such as TI should be characterized as multimodal events where thoughts become sensory (Mayer-Gross, 1932; Sterzer et al., 2016; Mishara et al., 2023)<sup>7</sup>. Although there seems to exist some relevant phenomenal differences, it can be also claimed that, in the context of psychosis, cognitive states like thoughts can become like the objects of sensory perceptions. In turn, this suggests that some comparisons might be allowed between these states. Finally, it is important to note that these reports offer *prima facie* evidence neither for an explanationist nor for an endorsement model of TI. What they do show is that we should not rule out so easily certain arguments – such as those coming from the endorsement-explanationist debate of perception-based delusions – purely based on sharp phenomenological distinctions unless such distinctions can be further argued. A more fruitful path is to examine whether particu-

<sup>7</sup> Over the last years, some authors within the phenomenological tradition in psychiatry have argued that doxastic accounts are insufficient to make sense of delusional phenomena because delusional phenomena would be symptomatic of a more profound and complex transformation in the relationship between the self and reality (see for example, Sass, 1994, 2004; Feysaerts et al., 2021). The idea is that a number of experiential (perceptual and quasi-perceptual) alterations registered at the onset of psychosis would alter the fabric of conscious experience and, in consequence, the experience of the basic sense of self. Although this perspective introduces fundamental elements to the debate about delusions, it is not clear if it can be characterized as “antidoxastic” at all. In fact, the reasons used to resist the doxastic notion of delusions are not entirely clear. After all, claiming that delusions are symptomatic of alterations of the structure of phenomenal experience is not incompatible with claiming that delusions are beliefs, taken as endorsements (Sass, 2004, p. 77; see also Bayne & Pacherie, 2004; see also López-Silva, 2021) or explanations (Sterzer et al., 2016, 2018; Corlett et al. 2020, 2020) of such experiences for example.

lar arguments coming from different discussions can work in different contexts rather than rule all of them out at once<sup>8</sup>.

## 2.1 Explanation or endorsement?

### 2.1.1 Subjective certainty

Endorsement models of delusions seem to be able to plausibly explain the *subjective certainty* that characterizes delusional reports (Sollberger, 2014). The idea is that, if we are asked to explain why one holds the belief ‘it’s raining outside’ with a high degree of subjective certainty, one might directly refer to the experience that grounds that belief. Such natural endorsement of the relevant experiential state would be the basis of the certainty with which the belief is reported. A faster onset of belief-fixation would be related to a clearly distinguishable experience which would produce high levels of certainty. In the case of TI, the subjective certainty shown in the reports could be explained by patients having the actual experience of a thought *as* being inserted by someone else, which in turn, would reinforce the belief and explain its incorrigibility over time without treatment. In contrast, explanationist models might not be able to fully accommodate this certainty because beliefs adopted *via* explanation would be associated with a slower onset of belief fixation, and therefore, a slower experience of subjective conviction. A subject might form the belief with the content ‘it’s raining outside’ by explaining and making sense of different experiential inputs available without having the actual experience of rain (some vague experience of the sound of the drops in the window, the sound of the wind, changes in sunlight, etc.). However, here subjective certainty would come later as the crystallization of a certain explanation that is found sufficiently plausible in light of the evidence. The claim of the defender of the endorsement model is that this characterization would not match the phenomenology of TI that seems to show that subjective conviction is a feature of the first full-blown appearance of the symptom.

It could be tempting to tip the scale in favor of the endorsement model. However, such conclusion is too hasty, as subjective certainty is a matter of degrees, and delusional phenomena are more complex than non-pathological beliefs. Sometimes, the endorsement defense implies that all TI cases are reported with the same high degree of subjective certainty, or at least, with the same degree of certainty as those beliefs acquired *via* endorsement. However, although some cases could be reported in this way, this is not always the case. Different cases of the same delusions can be reported with variable degrees of subjective certainty, even by the same person over time (Parnas & Handest, 2003; Fusar-Poli et al., 2022). De Hann and De Bruin (2010) claim that, in some cases, patients report their delusional episodes ‘as if’ they were the case: “it is as if my girlfriend can read my thoughts [...] it is as if I am from another planet” (p. 385, note 16). In many cases, patients are not entirely sure of what is going on with their experiential world and, in that context, the delusional belief might arise as

<sup>8</sup> I have argued for a relevant phenomenological distinction between cases of auditory hallucinations and TI (López-Silva et al., 2022). Often, phenomenological reports seem to support this distinction (Wilkinson & Alderson-Day, 2016).



a way of making sense of the patients' confusing experience of their inner and outer world (Mishara & Corlett, 2009; Sterzer et al., 2016; Mishara et al., 2023). From this point of view, explanationism could be better equipped to make sense of such variable degrees of certainty as a feature of an unstable and longer-onset process of doxastic fixation. This would not imply that there is no relevant grounding experience in the formation of the delusion. Rather, it means that it is plausible to think that, in some cases, this state can be vague and confusing with respect to its representational content.

The explanationist suggestion seems also phenomenologically appealing. The adoption of a delusional belief can be compared to the moment one finds an explanation after a period of overwhelming doubting and enquiring (*see* Conrad, 1958). Mishara (2010, p. 2) suggests that some delusions arise as 'revelations' or as '*aha* experiences' in the context of schizophrenia (*see* also Sterzer et al., 2016; Kusters, 2020, p. 84). This phenomenological quality suggests that those delusions might be adopted *via* explanation, and that subjective certainty would be a variable feature that could appear later in the process of belief production and fixation. In this way, the explanationist could also explain the incorrigibility of delusions, perhaps, as either the incorrigibility of an explanation that is playing an important role in unifying the experiential world of the patients (López-Silva, 2023) or as the incorrigibility of a belief adopted for the role it plays in the patient's attempt to deal with a psychological conflict (Bell, 2003; Bortolotti, 2014).

Advocates of the endorsement model could reply by suggesting that their approach is consistent with variable degrees of subjective certainty. Persons with schizophrenia express a conflict between the high degree of certainty with which they experience their delusions and what they know about reality prior to the delusional onset (Cermolacce et al., 2007; Fusar-Poli et al., 2022). In schizophrenia, even though patients believe what they are reporting with high degrees of conviction, they also express the feeling that something 'is not right' (Sass & Pienkos, 2015; Parnas et al., 2021). This conflict is characteristic of 'double-bookkeeping', a phenomenon that describes the ability of people diagnosed with schizophrenia to simultaneously live in two worlds, a shared social reality and a private solipsistic world where delusions are harvested (Bleuler, 1955). Reporting his own case, Jensen (2022, p. 1) observes that: "it is through this other reality that delusions manifest. It brings an alternative sense of reality that can both challenge and sometimes even deafen the sense of 'ordinary' reality". At the same time, Jensen explains that "ordinary reality is here essentially the 'normal', mundane world. It is a shared world in the sense that others also experience it and interact with each other within this world." But, contrasting, for the psychotic, the co-existing delusional reality "seems like a parallel or other reality that holds as much validity as the shared, ordinary world, but which often is separated from it." Finally, Jensen (2022) claims that: "often, the experiences pertaining to the so-called 'private-solipsistic world' are revelatory in nature, that is, they seem to offer profound insights and modes of cognition in a way which cannot be presented as such in the ordinary world" (ie, the experiences qualitatively differ from experiences pertaining to the shared-social world and is often not comparable with these).

Endorsement advocates could suggest that the existence of these two co-occurring different realities could be the source of high degrees of certainty coming from the



delusional reality itself, and, on the other hand, doubts and feelings of uncertainty coming from the collision between the delusional and the socially shared reality. The variability in the degree of certainty associated with delusions could be explained by this conflict. This would make the endorsement model more consistent with the variable degrees of certainty that characterize delusional reports and, additionally, could also explain the incorrigibility of delusional beliefs by appealing to the psychological or phenomenological role they play in the subject's reality. As it stands, the existence of variable degrees of subjective certainty seems to posit a challenge to both the endorsement and the explanationist alternative, so it seems that none of the alternatives should be discarded on these grounds. More importantly, this issue opens a second problem related to the ambivalence observed in delusional reports.

## 2.2 The ambivalence problem: To be, or not to be?

Parrott (2017) claims that reports of TI are characterized by the existence of certain ambivalence in the patients. Endorsement models could have problems explaining this issue. Parrott quotes Frith's patient who claimed that in TI "it's just like my mind working, but it isn't" (1992, p. 66), and Allison-Bolger's case (1999, #68) when she says: "own thoughts might say the same thing. But the feeling isn't the same. The feeling is that it is somebody else's." Parrott's objection is that the endorsement alternative becomes implausible in the face of this phenomenon because it needs to pack all the abnormality of the symptom into the representational content of the subject's experience, then, it would be unclear how the ambivalence of the person's report could be represented in the content of her experiences. After all, "how *could* an experience be 'just like' my mind working and also, at the same time, not like my mind working?" (Parrot, 2017, p. 50). The problem here seems to be how contrary elements could be simultaneously represented in the content of the experience.

In Parrott's opinion, it is just implausible to think that experiences can have the type of representational content that the endorsement model requires to explain this ambivalence, namely, experience with the content of the type ( $P \& \neg P$ ). Here, it is crucial to distinguish between the *feeling of ambivalence* that pervades the patients' report and the *ambivalence in the representational content* of the grounding experience ( $P \& \neg P$ ). Parrot's stronger claim seems to be that, within the endorsement model, the former element is explained by the latter. It may be that patients with psychosis have an experience like this, but, if so, it is far from coherent. Thus, by supposing that every component of what a people with psychosis report is somehow represented in the content of their experiences, the endorsement account would lead to an obscure understanding of the phenomenon.

The defender of the endorsement model can reply to the ambivalence challenge in, at least, two ways. The first way is to deny Parrott's stronger claim: it is possible to explain the reported feeling of ambivalence not necessarily in terms of the experiential content of the grounding experience. Rather, the endorsement theorist might suggest that such ambivalence can be plausibly explained by appealing to other features of the phenomenon. Sollberger (2014), for example, claims that: "one such feature would be the patient's experiencing a dissociation between SoO [sense of ownership] and SoA [sense of agency]" (p. 607). In TI, Sollberger argues, the thought

in question would be experienced with a normal sense of ownership but without a sense of agency (see also, Gallagher, 2014). Patients would be ‘owner-aware’ but not ‘agent-aware’ of a thought, and this dissociation might account for the *feeling of ambivalence* present in the patient’s reports. This reply looks attractive, but it might have some problems. One could say that the alleged dissociation between sense of ownership and sense of agency is not itself an experiential feature of the phenomenon, but rather an interpretation of it; it’s just a way of making rational sense of the patient’s reports. When approaching delusions, one should distinguish between what the patient says from how we interpret those reports. Here, an interpretation “is concerned with the reasons the patients might have – or fail to have – for their delusional speech” (Billon, 2016, p. 6). The defender of the endorsement model can recognize this subtle issue and reply that, even if the ownership-agency dissociation is an interpretation of the delusional speech of the patients, such interpretation explains the feeling of ambivalence alleged by Parrott, as it targets what patients *feel* in TI cases. Here, one needs to keep in mind that the main aim of the endorsement reply at this moment is to explain the ambivalence referred to by Parrott without appealing to the representational content of the cognitive experience grounding the belief, and, in this sense, the reply seems to do its job.

An explanationist might suggest that the real problem with this reply is that this specific interpretation of TI has received a number of objections over the years (see for example; Martin & Pacherie, 2013; Mishara et al., 2023); so, in replying to the ambivalence challenge, the defender of the endorsement model assumes a problematic interpretation of TI. One of the most relevant problems of this interpretation of TI for the endorsement theorist is that a lack of a sense of agency with a retained sense of ownership does not secure that the external attribution aspect is part of the grounding cognitive experience of TI (López-Silva & Cavieres, 2023). There is an important difference between becoming aware of a thought merely *as not being intended* and becoming aware of a thought *as being inserted by a specific agent*. The endorsement alternative needs to pack the latter rather than just the former aspect into the grounding experience of the delusional belief.

It is important to note that a number of our thoughts are experienced as *unbidden* i.e. suddenly popping into our stream of conscious without an associated sense of agency (Martin & Pacherie, 2013); as Frankfurt (1976) points out, some of our thoughts “strike us unexpectedly out of the blue” (p. 240). For the sake of the discussion, let’s assume that the distinction between sense of ownership and sense of agency is a plausible one in the case of thinking as well as in the case of bodily movements. In the case of an unbidden thought, ownership over the thought in question is retained, while its sense of agency is missing, just as in the case of thought insertion within the so-called standard approach. However, although we cannot identify any sense of agency for the unbidden thought, it is not externalized as in TI cases. *Prima facie*, this might suggest that, apart from this dissociation, TI patients would *explain* such dissociation in terms of external agency. Billon and Kriegel (2015) claim that an inserted thought *qua* inserted instantiates all the phenomenal properties of a normal thought but in addition, it has an extra phenomenal property: *it feels inserted* (p. 16). The problem for the endorsement alternative is that it needs to pack the whole content of the delusional belief in the experience that grounds so *feeling just inserted* it is not

sufficient for the endorsement theorist to make his case. In contrast, an explanationist might simply suggest that patients do feel inserted thought as inserted, but the external attribution aspect of the belief is an explanation for this experiential feature of the grounding thought.

There is a second way in which the endorsement theorist can reply to the ambivalence challenge. This reply shares aspects of the reply offered for the subjective certainty problem. Based on Bleuler's work, Sollberger (2014) appeals to the subject's background knowledge to account for the feeling of ambivalence identified in the delusional report. The idea is that persons with psychosis can distinguish between their delusional world and the real world and "this sort of double-awareness might arguably give rise to a competitive tension between, on the one hand, endorsing the delusional experience and, on the other, not endorsing it" (p. 607). One of the problems with this reply is that such *doubting* at the moment of endorsing - or not endorsing - a certain experience is not consistent with the degree of subjective certainty that the endorsement alternative attributes to delusional reports. This model puts subjective certainty as a feature of the early stages of the symptom's onset and this explanation of the feeling of ambivalence does not match this picture. It seems more sensible to say that, given this ambivalence, subjective certainty would come later in the development of the symptom, but such a view is more consistent with an explanationist approach, apparently. As it stands, it is not clear how these replies can explain the feeling of ambivalence identified by Parrott.

The endorsement theorist seems to be well aware of this issue and elaborates a second strategy that consists of accepting that the feeling of ambivalence can be grounded in the experiential content of the delusion, while denying that this issue is problematic. Sollberger (2014) appeals to the 'waterfall-illusion' to argue that it is possible that some perceptual states represent contradictory contents<sup>9</sup>. Usually, this illusion is taken to instantiate a case where a perceptual experience represents something as both *moving and not moving* (Crane, 1988). There are a few issues about this suggestion. First, one might resist such a way of characterizing the phenomenon (see for example, Mellor, 1988). In fact, the whole discussion about the waterfall illusion seems to be premised in an oversimplified phenomenological description of what happens in this case, for it is just not clear if one actually experiences the elements of the image as moving *and* not moving *simultaneously*. It might be the case that one experiences the illusion as moving *and* as not moving at different moments by quickly shifting the focus of visual attention and by increasing the inclusion of peripheral aspects of the field of visual awareness into the different perceptual inputs of the image. A second option is to say that one can experience different aspects of the illusion *as moving* and *as having moved* (Mellor, 1988). Thus, one might be able to resist the claim about the simultaneity of the two aspects allegedly implicated in this visual illusion. Certainly, a further discussion of this issue would take us too far from the main debate, but the point is that the acceptance of the fact that the feeling of ambivalence can be grounded in the representational content of experience of TI is controversial. Perhaps, such a skepticism about Crane's interpretation of this percep-

<sup>9</sup> For a description, example, and discussion of this illusion, check the Illusions Index in [www.illusion-index.org](http://www.illusion-index.org).

tual state is not needed, as the explanationist might simply reply by saying that even if this type of experience is possible, they can only occur in cases of perception. Perhaps this type of experience can be instantiated exactly because of the mind-independent nature of the stimuli they represent (perpetual experiences require direct contact with its objects, unlike cases of thinking). The point is that it is not obvious that one can experience such ‘conjunctive’ states in cases of thinking. What it is not controversial is to say that all that the ‘waterfall-illusion’ shows is that one can experience something that is not actually moving as moving. However, given that simultaneity is a requirement for the endorsement reply in terms of the possibility of contradictory experiences to work, appealing to the waterfall illusion does not entirely help.

There is a final issue that remains unresolved. Even if cognitive experiences with contradictory representational content exist, the reply does not make sense of why patients would endorse such experiences. Even if these experiences were possible, they may not be quickly endorsed. Again, the issue of subjective certainty plays a crucial role here. One of the main problems is that the general endorsement defense is rooted in the idea that deluded persons trust their experiences just the way non-deluded subjects do. Such an idea is not unproblematic. In an interview with Roberta Payne about her episodes of TI, she claims:

No, I did not trust my experiences in a normal way in the period of time leading up to this delusional period of aliens and rules. In the months preceding it, I experienced unreality many times [...] I had other experiences in which I had to, for instance, touch a coffee table in front of me to make sure it was real. I had short times in which nothing outside myself seemed to exist: I told my psychologist that it felt like the world I was observing was like wallpaper (Payne, 2015).

The endorsement assumption seems to be in clear conflict with the way in which, at least some people, approach their own pre-psychotic episodes. In this sense, an explanationist account seems to be in a better position to make sense of the way some persons approach their experiential states in periods that precede the adoption of delusions of TI. This does not mean that an endorsement model cannot reply to the ambivalence challenge. Perhaps the endorsement advocate could appeal again to the phenomena of double-bookkeeping. Ambivalence could be explained by appealing to experiential elements outside the representational content of the delusion. The idea is that the awareness of two co-existing realities could explain the feeling of ambivalence characteristic of delusional reports in schizophrenia. So, again, even though patients believe what they are reporting with high degrees of conviction, they also express a feeling of ambivalence coming from the collision of those two realities in the subject’s stream of consciousness. This reply might not be entirely satisfactory, but it allows the endorsement advocate to deal with the ambivalence problem on experiential grounds. Another potential reply offered by the endorsement advocate could claim that the feeling of ambivalence identified by Parrott is not characteristic of all cases of TI, which is supported in the literature (Mullins & Spence, 2003; López-Silva, 2018; Mishara et al., 2023). While some patients do show ambivalence in their reports, it is not part of all cases. Just as in the case of the discussion about the variable degrees of subjective certainty of different reports of the same delusion, one might be tempted to suggest that delusions reported with feelings of ambivalence might be endorsed *via* explanations of vague grounding experiences, while

those reported without – or with low degrees of – feelings of ambivalence might be endorsed via endorsement processes. In this sense, Roberta’s case would be an example of the former modality of delusions of TI fixation.

## 2.3 The experiential encoding problem

Bottom-up doxastic models of delusions of TI need to account for at least two things: (i) the content of the abnormal cognitive experiential state that grounds the delusional belief and, (ii) the content of the doxastic state (belief). In doing so, both endorsement and explanationist models face different empirical and conceptual challenges. Here are two final issues:

### 2.3.1 The incredibility objection

Explanationist models have an easier job in explaining the content of the cognitive experience grounding TI because the less one packs into the content of this experience, the easier it is to account for how such experiential state acquires its content (Pacherie et al., 2006, p. 567). This has been called ‘*the experiential encoding problem*’. Parrott’s objection seems to be just one of the many formulations of this problem available in the literature. The idea is that, when making sense of TI, it is unproblematic to think that certain thoughts might be able to convey the vague content “this thought is not mine.” One can clearly agree that this type of cognitive experience is actually quite possible, even in non-pathological cases. A problem here, an explanationist would say, is that the endorsement theorist needs to appeal to the existence of cognitive experiences of the specific type “this thought has been inserted”, and this is taken to be a very implausible move (see Vosgerau & Voss, 2014)<sup>10</sup>.

There are two more ways in which the experiential encoding problem can be formulated. The first consists in claiming that the type of cognitive experience with the content “this thought has been inserted” is just impossible, period. This has been called the ‘*incredibility-objection*’. One should be careful with objections of this kind since psychotic experiences seem to challenge our most basic notions about the nature of conscious experience<sup>11</sup>. Now, it is not clear why exactly the type of experiences required by the endorsement model should be considered impossible in principle. Sollberger (2014) seems to suggest that this objection relies on a certain lack of empathy when he asks: “Is it only because people who are lucky enough not to suffer from inserted thoughts fail to imagine what such experiences could be like?” (p. 598). If this is the point, the defender of the endorsement model might reply that there are several ways in which we can make rational sense of psychotic symptoms and, therefore, make sense of how certain types of - *prima facie* implausible - expe-

<sup>10</sup> This problem is slightly different from the one pointed out by Parrott (2017), which focuses on how the endorsement model would struggle to explain experiences with the content of the form (P&¬P).

<sup>11</sup> As Sollberger (2014) points out: ‘The incredibility-objection does not seem to have much bite, though. It seems to rely on a crude form of *experiential chauvinism*. For why exactly should such introspective-cognitive experiences not be possible? Is it only because people who are lucky enough not to suffer from inserted thoughts fail to imagine what such experiences could be like?’ (p.598).

riences can occur. Moreover, the whole discussion about the experiential routes to delusional beliefs is premised on this basic idea.

More specifically, in replying to the incredibility objection, Sollberger (2014) again appeals to the distinction between sense of mental ownership and sense of mental agency. For example, there are rational ways of making sense of experiences of TI or, in other words, there are many ways in which experiences of TI can become *intelligible*. The real problem is that the endorsement theorist does not prove that the *type of experience* he requires can become intelligible by appealing to the standard model of TI, the tenability of which is the main issue at hand. Lack of sense of agency does not entail external agency (of the type required by the endorsement model) as part of the representational content of the experience that grounds TI. In fact, the dissociation between ownership and agency seems to favor an explanationist model by which the type of contents assigned to grounding experiences is less demanding (Vosgerau & Voss, 2014). The advocate of the endorsement model might reply by saying that the relationship between ownership and agency can also be formulated in endorsement terms (see Gallagher, 2007, 2014). However, if this is the case, all it shows is that the ownership-agency relationship can be formulated in either endorsement or explanationist terms, so it does not give the endorsement model any privileged position over explanationist views.

In formulating a second reply to the incredibility objection, the endorsement theorist appeals to cases of synaesthesia. For a long time, it was claimed that experiences of synaesthesia were not genuine cases of perception (perhaps, another case of the incredibility objection but in the context of perception). Synesthetic experiences were often understood as imaginings, quick verbal associations, or as the result of extreme uses of metaphors (Ward & Cytowic, 2006; Zawislawska, 2019). However, as mentioned before, research has shown that synaesthesia is a genuine perceptual experience (Brang & Ramachandran, 2011). The endorsement defense concludes that, just as in cases of synesthetic experiences, it is plausible to think that there is *something similar to the experience of a thought inserted by an external agent*. In clarifying his point, Sollberger (2014) compares a blind subject trying to understand what it is like to perceive color:

Likewise, people who have never experienced inserted thoughts may fail to know what it is like to be first-personally aware of a thought that is woven into one's consciousness *as* belonging to another cognitive agent. But this lack of first-personal knowledge does not imply the bare impossibility of such experiences. Much more needs to be said in order to reach that conclusion (p. 599).

All that synesthetic experiences show is that it is plausible that there is *something similar to seeing sounds as colours or to smell colours as odours*, and so on. The problem is that these experiences are not close to the type of experiential representational content the endorsement theorist needs to account for in cases of TI. If we establish this parallelism in the right way (just for the sake of the discussion), the type of perceptual experiences that Sollberger should be comparing would be something like: 'I am aware of *smelling colours as odours*, and this state has been inserted by an external agent'. However, the types of experience equated by Sollberger seem rather different in terms of their representational content. As it stands, the reply does seem

to help the endorsement model to make its case about the existence of the type of cognitive experiences it requires to make sense of TI.

### 2.3.2 The non-agency objection

The second way in which the *experiential encoding problem* is formulated can be referred as the *non-agency objection*. This objection is in strict connection with the reply offered by the advocate of the endorsement model to the incredibility objection. Perhaps it might be possible for a subject to experience their own thoughts as not being their own, but as someone else's. Synofzik et al. (2008) explain that, in TI, the sense of non-agency is coupled with strange feelings (mostly of egodystonic nature, see López-Silva, 2015). These feelings would explain the extra reasons that subjects have to explain the thought in question in terms of external agency (not as in cases of unbidden thoughts, for example). The claim is that in such an ambivalent situation, the subjects feel the urgent need to explain what is going on and external attribution arises as an explanation for such a stressful situation: *if this is not mine, it must be someone else's*.

The first reply coming from the endorsement camp is that 'it is not mandatory to explain alien agency in non-experiential terms' (Sollberger, 2014, p. 599). But this reply is not very helpful nor is it mandatory to explain it in experiential terms. The whole issue of the debate is to clarify whether external agency is part – or not – of the cognitive experience that grounds delusions of TI. The issue is that this reply removes all the importance from this dimension of the debate. Here's a perhaps better reply by the endorsement theorist: One does not need to characterize a sense of mental agency in a disjunctive fashion, namely, as either 'I'm the author of this thought' or 'I am not the author of this thought'. Sousa & Swiney's (2011) suggests that apart from the absence of self-agency, there is a sense of alien agency attached to the underlying experience of TI. However, it is not clear what a sense of 'alien agency' means. An explanationist might suggest that this is a vague experience, and that 'alien agency' involves having the experience of a thought as being inserted by the specific agent, as delusional subjects refer in their reports. The other possibility is to recall Billon and Kriegel's suggestion about a thought experienced as 'being inserted', but again, this is not enough for the endorsement theorist to make their case. The challenge here is to explain how this sense of external agency – whatever that is – transfers to any of the external agents identified by subjects as the responsible party of their inserted thoughts. This last issue, again, seems to point to the need to characterize the process of adoption of delusions of TI as the result of a number of endorsement and explanationist strategies within a single subject.

### 2.4 The doxastic encoding problem

The endorsement alternative may better explain how the content of delusional beliefs of TI is acquired. After all, experiences with more specific and richer content seem to posit more constraints in structuring the beliefs they ground. As Langdon and Bayne (2010) claim: "by contrast, advocates of a purely explanationist account need to work much harder in order to explain the generation of the delusional content, for on such



an account the gap between experiential content and delusional content is wider” (p. 333). Let’s call this the ‘*doxastic encoding problem*’. The problem for the explanationist is that if subjects have a cognitive experience of a thought that is not theirs, with some other additional features (irritating feelings, being inserted, etc.), why do they then form the belief that the thought has been inserted by an external agent? While the endorsement theorist may suggest that this issue can be explained by delusional subjects having a cognitive experience with the same content, the explanationist claims that delusional beliefs of TI emerge as the subjects’ attempt to explain an unusual experience. However, as Pacherie et al. (2006) point out, “the problem with this suggestion is that delusional beliefs are typically very poor explanations of the events that they are supposedly intended to explain” (p. 567)<sup>12</sup>. Allegedly, more plausible explanations of the same strange experience that patients are undergoing would be available, “some of which might be actively recommended to them by family and medical staff” (p. 567).

The explanationist could reply by asking: why exactly should we expect psychotic patients to offer the same type of explanations that non-delusional patients are expected to offer?, after all, a number of other problems (perceptual, motivational, or purely neurocognitive) can influence the type of explanations that patients entertain to make sense of their experiential states (López-Silva & Cavieres, 2021). It is not obvious that, at the moment of experiencing the abnormal cognitive experience, patients always have other – better – explanations available. A number of delusional beliefs are adopted within contexts of profound affective and perceptual transformations (Marwaha et al., 2013; López-Silva, 2015, 2018; Mishara et al., 2023; López-Silva et al., 2022b). Marwaha et al. (2013, p. 6) claim that “the sense that emotional experiences are out of one’s personal control may prompt a search for meaning that may find explanations in terms of external influence.”

In our interview, Roberta Payne comments that the period of time preceding her episodes of TI was characterized by a total transformation in the way she perceived her inner and outer reality. As already mentioned, during *delusional atmospheres* ‘patients feel uncanny and that there is something suspicious afoot’, and the world is becoming a place where ‘everything gets a new meaning’ (Jaspers, 1963, p. 98). During this period, everything (outer and inner reality) is experienced as strange (Sass, 1992), everything looks unreal (Gross & Huber, 1972), and the entire world loses its predictability (Fuchs, 2005; Mishara & Corlett, 2009). In this context, thoughts are experienced as coming out of the blue (Kusters, 2020). As a consequence of all these changes, pre-psychotic subjects stop trusting their own experiences and an increased need for explanation arises (Mckay et al., 2007)<sup>13</sup>. In midst of all these changes, it is not clear if patients have better explanations of their experiences available; an if better explanations are available, it is not clear whether patients can become aware of them, or if they have the cognitive resources available to discard and examine alternatives. As Martin and Pacherie (2013) comment: “extraordinary events call for extraordinary explanations, so to speak” (p. 121).

<sup>12</sup> This point has been formulated in the context of the examination of perception-based delusions. However, the suggestion also applies in cases of TI.

<sup>13</sup> This seems to be consistent with the explanationist account of TI formulated by Synofzik et al. (2008).

Apart from suggesting that the content of experiences is less specific in these situations, the explanationist could claim that psychotic subjects arrive at explanations in terms of external agency as a way of infusing some basic order to their already disorganized awareness of reality (see Mishara, 2009; Mishara & Corlett, 2009; Mishara et al., 2023; López-Silva, 2023). Subjects could create these explanations because it is all they elaborate under such overwhelming circumstances. Thus, the delusional explanation would imply a quick way out from this general state of experiential disorganization. However, the endorsement theorist can suggest that this reply still does not explain why this is the only explanation that patients can elaborate to render their field of awareness more experientially organized. Again, the explanationist can say that it is not clear whether we should expect from patients with psychosis the same type of explanation offered by the non-psychotic. Perhaps the bizarre nature of the delusional explanation by psychotic patients can be accounted for by different doxastic styles. Pacherie et al. (2006) point out that two normal subjects could reason from the same type of experiential state quite different doxastic states, depending upon their personal doxastic styles. It could be the case that only one of these subjects stresses abnormal explanations while the other does not. Similarly, the bizarre nature of the explanations formulated by patients can be accounted for in terms of exaggerated doxastic styles, in conjunction with their overwhelming experiential circumstances. However, a lot more needs to be said by the explanationist for this suggestion to become more challenging.

There are two final suggestions that the endorsement theorist might make. First, it is possible to recognize that some cases of TI are in fact adopted in the context of a profound affective and perceptual alteration, but at the same time, the endorsement theorist might insist in suggesting that they are adopted because certain thoughts are experienced as being inserted by external agents. The problem is that this does not seem to be the case during delusional atmospheres where the information regarding the causes of the different mental states of a subject is mostly missing. In fact, during this period patients experience a pervading feeling of passivity, which, in turn, seems to be based on the lack of awareness of the agency of their different mental states (Sass, 1992; Wegner, 2002; Fusar-Poli et al., 2022). The final suggestion that the endorsement theorist can make is to say that, while it is true that the adoption of delusions of TI is preceded by a number of phenomenological transformations, this does not apply to all cases. However, this does not explain how experiences with the type of representational content required by the endorsement theorist can be possible at all. Finally, it does seem consistent with the phenomenological and neurocognitive data (Sterzer et al., 2016, 2018; Fusar-Poli et al., 2022; López-Silva & McClelland, 2023).

## 2.5 Thought insertion and hybrid belief-fixation

Both explanationist and endorsement models of TI face important explanatory challenges. As it stands, the debate is far from resolved, in part because it has been formulated as a dichotomy: while some people argue that TI should be exclusively explained in explanationist terms, others claim that this should be done so in exclusively endorsement terms. However, when considering the complexity of our men-

tal life, such dichotomies are not always helpful. Arguably, the two models capture important elements of our target phenomenon and, in light of this, the solution to the dispute could lay on finding a middle ground between the two main alternatives examined. In this section I shall explore how this could be done. However, the proposal should be considered merely suggestive, not exhaustive, as a number of issues exceeding the scope of this paper would need refining.

A first integrative option is to claim that, in some cases, TI could be adopted *via* endorsement while, in some others, it can be adopted *via* explanationist mechanisms.<sup>14</sup> Delusions adopted *via* endorsement could be those grounded in experiences that preserve phenomenological features closer to paradigmatic cases of cognitive experiences, while those adopted *via* explanationist mechanisms would be those grounded in cognitive experiences, having phenomenological features that overlap with other experiential modalities (as proposed by Mishara et al., 2023). Similarly, those delusions adopted in the presence of an external agent would be more likely to be adopted *via* endorsement (Saks, 2007), while those adopted in its absence might be more likely to be adopted *via* explanationist mechanisms (Payne, 2013). Additionally, delusions adopted *via* endorsement would be related to higher degrees of subjective certainty, while those adopted *via* explanation would be related to ambivalence and variable degrees of subjective certainty. However, this option is still argumentatively problematic because it does not overcome the particular objections against each view involved in the debate<sup>15</sup>.

Let me explore a second integrative alternative. As stated earlier, cases of TI vary considerably in terms of subjective certainty, ambivalence, heterogeneity of the alien inserter of the thought, phenomenological context, etc. A key issue that is often overlooked is that that delusions – as well as beliefs in general – differ in the degree of complexity of their propositional content. The belief with the content “it’s raining outside” is certainly less complex in terms of propositional content than “there is a dog hidden behind the table.” In this sense, delusions of TI would be complex beliefs and this distinction is crucial to understand the way in which they are adopted. A plausible option to explain the many ways in which delusions of TI are heterogeneous is by suggesting that the two modalities of adoption of delusions work together to form a *single belief*. This is not an implausible idea because, after all, these two modalities aim at the same task, namely, implementing a subject’s representation of

<sup>14</sup> This idea has been explored in the context of the discussion of the relationship between delusions and confabulation (Langdon & Payne, 2010), and in the debate about the adoption of delusional beliefs of Capgras (Davies et al., 2001). In the context of TI, a challenge would be to clarify the commonalities between different cases, however, such a task goes beyond the scope of this article.

<sup>15</sup> This is by no means to say that such a proposal cannot be applied to other less complex cases of belief. Take the case of a subject S who holds the belief that he has heard God’s voice. S could have adopted this belief by using either the endorsement or the explanationist modality. S could have experienced the sound of the wind or some other auditory input *as* being God’s voice. This would certainly depend on S’s background knowledge, set of – not only religious – beliefs, affective states, mood, expectations, and a number of other factors. Perhaps S has been having a difficult time and, being a religious man, S was expecting to receive a sign from God. The point is that a number of affective, cognitive, phenomenological, and even cultural factors might have led S to hear a certain sound *as* God’s voice. Now, S could have experienced just a rather unclear and ambiguous sound that could have been taken as God’s voice by him. Again, such an explanation would depend on the aforementioned factors.

a particular internal or external state of affairs as actual (McKay & Dennett, 2009; Connor & Halligan, 2020). From this point of view, delusions of TI are better understood as being adopted through the *conjoined* action of both endorsement and explanationist mechanisms of belief-adoption, just like in some cases of non-delusional complex beliefs. Consider the case of the belief “there is a dog hidden behind the table.” A subject might have endorsed the experience with the content *there is something behind the table*. However, the doxastic state “there is a dog hidden behind the table” might have emerged as the dog being an explanation for a poorer experiential state. As such, it is plausible to say that the final doxastic state that is produced is the result of the conjoined operation of endorsement and explanationist mechanisms at a single moment.

Taking this idea into account, we could maintain that single delusions of TI contain various representational contents that can be distinguished in the patients’ report (López-Silva, 2018). Some of these contents would be arrived at *via* an explanationist route whereas others would be arrived at *via* an endorsement route. This is to say that we should think of single cases of delusional beliefs of TI as a *doxastic hybrid state*. The claim here is not that one modality works for some cases while the other works for the rest, while this option could be also worth exploring. Rather, the suggestion is that, when patients come to adopt delusions of TI, they do so by way of these two modalities working in parallel. This hybrid option seems to explain the variable degrees of certainty of the way delusions are reported and, more importantly, avoids the objections of the experiential and doxastic encoding problem. At the same time, the main challenge of a potential hybrid model would be to specify which aspects of the delusional belief are acquired via endorsement and which ones of them are acquired via explanation.

Let’s briefly explore some ways in which a hybrid model could be formulated. Consider the case of a subject who claims that the radio has placed a thought into her mind. Before this is reported, the subject might experience random unconnected thoughts, memories, and perceptual states (see Saks, 2007; López-Silva, 2023). During this period, the subject might start experiencing a profound permeability in the distinction between inner and outer reality (Fuchs, 2005; Payne, 2015; Mishara et al., 2023). Gradually, the experience of self and the world become blurry. During this pre-delusional period of experiential confusion, the subject experiences a thought as not being theirs (and inserted) *and* a visual or auditory experience of a radio. A hybrid model could propose here that the subject endorses those two experiential states separately, but in trying to make sense of the dominant phenomenological confusion, a delusional explanation puts them together, leading up to the final content of the doxastic state reported by the patient (see for example, Mishara, 2010). In this specific case, both the thought and the external agent would be present in the subject’s field of awareness, but they would be tied together in a delusional explanationist move. A potential objection is that not all delusions of TI are adopted in the context of these more general alterations in the structure of consciousness. However, this objection fails to recognize the most constitutive features of our target phenomenon. TI is usually taken to be a psychotic delusion, and from this point of view, it is important to note that most psychotic patients adopt delusions in the context of a fragmented experience of themselves and the world (Silverstein & Uhlhaas, 2004; Uhlhaas & Silverstein, 2005; Mishara et al., 2023; López-Silva & Cavieres, 2023). In fact, first-personal

accounts of the context in which TI emerge seem to support this idea (Saks, 2007; Payne, 2013, 2015; Fusar-Poli et al., 2022). Finally, to deny this is to overlook one of the most crucial aspects of the occurrence of most psychotic delusions. In fact, Jaspers (1963) considers this altered phenomenological context in which delusions are adopted as the hallmark of psychotic delusions (see also Sass, 1994, Chap. 3 and Ratcliffe, 2013). This is one of the ways in which the adoption of a delusional belief of TI might be characterized as a *doxastic hybrid*. However, it is not the only way.

It is important to note that there are cases where the alien agent might be not present in the patient's field of awareness at the moment of the adoption of the delusional belief. Here is a way in which a hybrid model might deal with these cases. First, it is important to consider the complexity of our conscious life. While writing this paper, I am not only experiencing certain thoughts (that are the base on which I guide my writing) or certain visual experiences; I am also remembering, hearing, smelling, evocating, all at the same time. However, while some of these states catch my attention, most of them remain in the background, or *periphery*, of my conscious activity. The mental states that remain in the periphery can be perceptions, thoughts, feelings, or memories. In trying to make sense of a thought that is experienced as inserted (in a context of experiential confusion), patients explain it picking randomly some of the elements that have remained in the periphery of their conscious attention or even in their memory (for a further development of this proposal, see López-Silva, 2021, 2023). It could be that 'Chris' was not part of the main field of awareness; however, 'Chris' might have been retrieved from a thought, or a memory from the fringe of consciousness. Thus, the delusional belief would combine two components as a delusional explanation, in a similar way as the first hybrid option aforementioned. It is important to note that a hybrid model does not need to appeal to a direct perception of the potential alien agent. Rather, it might arise from the fringe consciousness or subject's memories. Thus, a hybrid model might be able to link the final delusional report with the patient's set of background knowledge. As previously said, my suggestions are not meant to be exhaustive. Rather, here I have shown hybrid alternatives, and, motivated further examinations.

## 2.6 Final remarks

According to bottom-up doxastic approaches, delusions are abnormal beliefs grounded in varied abnormal experiential states. However, the exact role of these states is still under discussion. In this paper, I have formulated this debate in the specific context of thought insertions (TI). Both the endorsement and explanationist approaches have different explanatory strengths and weaknesses when attempting to make sense of the different aspects of our target phenomenon. Thus, it seems plausible to claim that, as Pacherie et al. (2006) do, a comprehensive account for delusional phenomena may contain both endorsement and explanationist elements. Arguably, the attempt to explain all delusions by using a simple doxastic alternative is far too ambitious, and it seems reasonable to suggest that, in some cases, patients might adopt delusional belief by way of endorsement and explanationist mechanisms working conjointly. Delusions are complex in terms of representational content, and that these two doxastic modalities structure different elements of the final doxastic state reported by patients. In this sense, TI could be characterized as a hybrid doxastic state. While the advocates of explanationist views still struggle

to account for the doxastic content of the delusions, and endorsement theorists still need to make better sense of the type of experiential states that they posit, the view proposed here seems to overcome these problems. Certainly, the potential hybrid formulations discussed is open to a number of improvements and objections. However, the main aim of this paper is not to defend a full version of a hybrid model of TI, but rather to explore potential hybrid models that may explain the process of fixation of delusional beliefs, by overcoming the problems commonly associated with full endorsement and explanationist models. Although further formulation and research is needed, hybrid doxastic models seem to be a plausible way to deal with complex phenomena such as delusions of thought insertion.

**Acknowledgements** I would like to thank Thomas Fuchs, Thomas Uebel, Elisabeth Pacherie, Álvaro Cavieres, Tom McClelland, Rob Knowles, Leo Tarasov, and Geraldine Ellwanger for insightful discussions on earlier versions of this paper. My special thanks go to Joel Smith and Tim Bayne for their dedication and support while preparing the first drafts of this paper during my PhD and to Roberta Payne whose comments from a first-personal point of view of schizophrenia were fundamental to develop the final version of this paper.

**Funding** This work was funded by the Project FONDECYT Regular N°1221058 ‘The Architecture of Delusions’ granted by the National Agency for Research and Development (ANID) of the Government of Chile and the Renewing Phenomenological Psychopathology Project (Discretionary International Exchange Award / 223452/Z/21/Z) granted by the Wellcome Trust to Matthew Broome and Giovanni Stanghellini.

## References

- Allison-Bolger, V. Y. (1999). *Collection of Case histories*. Unpublished typescript.
- American Psychiatric Association [APA]. (2024). In DSM-V-TR (Ed.), *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>
- Bayne, T. (2010). Delusions as doxastic states: Contexts, compartments and commitments. *Philosophy Psychiatry and Psychology*, 17(4), 329–336.
- Bayne, T., & Pacherie, E. (2004). Bottom-up or Top-down. Campbell’s rationalist account of monotheistic delusions. *Philosophy Psychiatry & Psychology*, 11(1), 1–11.
- Bayne, T., & Pacherie, E. (2005). In defence of the doxastic conception of delusion. *Mind & Language*, 20(2), 163–188.
- Bell, D. (2003). *Paranoia*. Icon.
- Billon, A. (2016). Making sense of the Cotard Syndrome: Insights from the study of Depersonalisation. *Mind & Language*, 31, 356–391.
- Billon, A., & Kriegel, U. (2015). Jaspers’ Dilemma: The Psychopathological Challenge to subjectivities theories of consciousness. In R. Gennaro (Ed.), *Disturbed consciousness*. MIT Press.
- Blackmore, S. J., et al. (2005). Somatosensory activations during the observations of touch and a case of vision-touch synaesthesia. *Brain*, 128, 1571–1583.
- Bleuler, E. (1955). *Lehrbuch der Psychiatrie*, 9e Auflage. Berlin: Springer Verlag.
- Bongiorno, F. (2019). *Is the Capgras delusion an endorsement of experience?* Mind & Language.
- Bortolotti, L. (2010). *Delusions and other Irrational beliefs*. OUP.
- Bortolotti, L. (2014). The epistemic innocence of motivated delusions. *Consciousness & Cognition*. Online first: <http://www.sciencedirect.com/science/article/pii/S1053810014001937>
- Brang, D., & Ramachandran, V. S. (2011). Survival of the synesthesia gene: Why do people hear colors and taste words? *PLoS Biology*, 9(11), e1001205. <https://doi.org/10.1371/journal.pbio.1001205>
- Cahill, C., & Frith, C. D. (1996). A cognitive basis for the signs and symptoms of schizophrenia. In C. Pantelis, H. E. Nelson, & T. Barnes (Eds.), *Schizophrenia: A neuropsychological perspective*. Wiley.



- Campbell, J. (2001). Rationality, meaning and the analysis of delusion. *Philosophy Psychiatry & Psychology*, 8(2-3), 89–100.
- Cermolacce, M., Naudin, J., & Parnas, J. (2007). The “minimal self” in psychopathology: Re-examining the self-disorders in the schizophrenia spectrum. *Consciousness and Cognition*, 16(2007), 703–714.
- Clutton, P. (2018). A new defence of doxasticism about delusions: The cognitive phenomenological defence. *Mind & Language*, 33(2), 198–217. <https://doi.org/10.1111/mila.12164>
- Coltheart, M. (2002). Cognitive neuropsychology. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (Vol. 4): Methodology in experimental psychology, 3rd edition (pp. 139–74). London: Wiley.
- Coltheart, M. (2015). Delusions. In R. Scott, & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*. Wiley.
- Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology*, 62, 271–298.
- Connors, M. H., & Halligan, P. W. (2020). Delusions and theories of belief. *Consciousness and Cognition*, 81.
- Conrad, K. (1958). *Die beginnende Schizophrenie*. Thieme.
- Corlett, P. R., & Fletcher, P. C. (2021). Modelling delusions as temporally-evolving beliefs. *Cognitive Neuropsychiatry*, 26(4), 231–241. <https://doi.org/10.1080/13546805.2021.1938984>
- Corlett, P. R., Krystal, J. H., Taylor, J. R., & Fletcher, P. C. (2009). Why do delusions persist? *Front Hum Neurosci*. 2009;3:12. <https://doi.org/10.3389/neuro.09.012.2009>. PMID: 19636384; PMCID: PMC2713737.
- Crane, T. (1988). The waterfall illusion. *Analysis*, 48, 142–147.
- Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy Psychiatry and Psychology*, 8(2–3), 133–158.
- De Haan, S., & De Bruin, L. (2010). Reconstructing the minimal self, or how to make sense of agency and ownership. *Phenomenology and the Cognitive Sciences*, 9, 373–396.
- Feyaerts, J., Kusters, W., Van Duppen, Z., Vanheule, S., Myin-Germeys, I., & Sass, L. (2021). Uncovering the realities of delusional experience in schizophrenia: A qualitative phenomenological study.
- Frankfurt, H. (1976). Identification and externality. In A. O. Rorty (Ed.), *The identities of persons* (pp. 239–251). University of California Press.
- Frith, C. (1992). *The cognitive neuropsychology of schizophrenia*. Erlbaum.
- Fuchs, T. (2005). Delusional mood and delusional perception: A phenomenological analysis. *Psychopathology*, 38, 133–139. <https://doi.org/10.1159/000085843>
- Fusar-Poli, P., et al. (2022). The lived experience of psychosis: A bottom-up review co-written by experts by experience and academics. *World Psychiatry*, 21(2), 168–188.
- Gallagher, S. (2004). Neurocognitive models of schizophrenia. A neurophenomenological critique. *Psychopathology*, 37, 8–19.
- Gallagher, S. (2007). The natural philosophy of Agency. *Philosophy Compass*, 2(2), 347–357.
- Gallagher, S. (2014). Relations between agency and ownership in the case of schizophrenic thought insertion and delusions of control. *The Review of Philosophy and Psychology Online First*. <https://doi.org/10.1007/s13164-014-0222-3>
- Gross, G., & Huber, G. (1972). Sensorische Störungen Bei Schizophrenien. *Arch Psychiatr Nervenkr*, 216, 119–130.
- Hohwy, J., & Rosenberg, R. (2005). Unusual experiences, reality testing and delusions of alien control. *Mind and Language*, 20, 141–162.
- Hubbard, E., & Ramachandran, V. S. (2005). Neurocognitive mechanism of synaesthesia. *Neuron*, 48, 509–520.
- Jaspers, K. (1963). *General Psychopathology* (7th edition). Manchester: Manchester University Press.
- Jensen, G. (2022). Delusions and reason. *Schizophrenia Bulletin*, sbac185. <https://doi.org/10.1093/schbul/sbac185>
- Kusters, W. (2020). *A philosophy of madness*. MIT Press.
- Langdon, R., & Bayne, T. (2010). Delusions and confabulation: Mistakes of perceiving, remembering, and believing. *Cog Neuropsychiatry*, 15(1/2/3), 319–345.
- López-Silva, P. (2015). Schizophrenia and the place of egodystonic states in the aetiology of thought insertion. *The Review of Philosophy & Psychology*. <https://doi.org/10.1007/s13164-015-0272-1>
- López-Silva, P. (2016). The typology problem and the doxastic approach to delusions. *Filosofia Unisinos/Unisinos Journal of Philosophy*, 17(2), 202–211. <https://doi.org/10.4013/fsu.2016.172.15>



- López-Silva, P. (2018). Mapping the psychotic mind: A review on thought insertion. *Psychiatric Quarterly*, 89(1), 957–968. <https://doi.org/10.1007/s11126-018-9593-4>
- López-Silva, P. (2019). Elementos para una teoría de las atribuciones de agencia mental. En P. López-Silva (Ed.), *Discusiones contemporáneas en filosofía de la mente: Voces locales*. Valparaíso: Serie Selección de Textos – Sello Editorial de la Universidad de Valparaíso.
- López-Silva, P. (2021). La Marca De La Psicosis: Hacia Una Breve Síntesis Del Problema Tipológico De Los Delirios. *Revista Colombiana De Psiquiatría*. <https://doi.org/10.1016/j.rcp.2021.11.002>
- López-Silva, P. (2023). Minimal Biological Adaptiveness and the Phenomenology of Delusions in Schizophrenia. In A. Falcato & J. Gonçalves (Eds.), *The Philosophy and Psychology of Delusions*. London: Routledge.
- López-Silva, P. (2024). Thinking in Schizophrenia and the Social Phenomenology of Thought insertion. *Philosophical Psychology*. <https://doi.org/10.1080/09515089.2024.2334720>
- López-Silva, P., & Cavieres, A. (2021). Salto a Conclusiones Y Formación De Delirios en Psicosis: Un Análisis Crítico. *Psiquiatría Biológica*, 28(1), 9–13. <https://doi.org/10.1016/j.psiq.2020.12.002>
- López-Silva, P., & Cavieres, A. (2023). Schizophrenia and the error-prediction model of thought insertion. In P. López-Silva, & T. McClelland (Eds.), *Intruders in the mind: Interdisciplinary perspectives on thought insertion*. Oxford University Press.
- López-Silva, P., & McClelland, T. (2023). *Intruders in the mind: Interdisciplinary perspectives on thought insertion*. Oxford University Press.
- López-Silva, P., et al. (2022a). *Are these my thoughts?: A 20-year prospective study of thought insertion, thought withdrawal and thought broadcasting and their relationship to auditory verbal hallucinations*. Schizophrenia Research. <https://doi.org/10.1016/j.schres.2022.07.005>
- López-Silva, P., Cavieres, A., & Humpston, C. (2022b). *The Phenomenology of Auditory Verbal Hallucinations in Schizophrenia and the challenge from Pseudohallucinations*. Frontiers in Psychiatry.
- López-Silva, P., Núñez de Prado, M., & Fernández, V. (2024). What are delusions? Examining the typology problem. *WIREs Cognitive Sciences*. <https://doi.org/10.1002/wcs.1674>. Online first.
- Martin, J. M., & Pacherie, E. (2013). Out of nowhere: Thought insertion, ownership and context-integration. *Consciousness and Cognition*, 22(1), 111–122.
- Marwaha, S., Broome, M., Bebbington, P., Kuipers, E., & Freeman, D. (2013). Mood Instability and Psychosis: Analyses of British National Survey Data. *Schizophrenia Bulletin*, Published online first: <http://schizophreniabulletin.oxfordjournals.org/content/early/2013/10/25/schbul.sbt149.short>
- Mathieson, D. Schizophrenic thought insertion and self-experience. Review of Philosophy, & Psychology (2023). <https://link.springer.com/article/10.1007/s13164-023-00680-2>
- Mayer-Gross, W. (1932). Psychopathologie und Klinik Der Trugwahrnehmungen. In O. Bumke (Ed.), *Handbuch Der Geisteskrankheiten. Band I. Allgemeiner Teil I*. Verlag von Julius Springer.
- McKay, R., & Dennett, D. (2009). The evolution of misbelief. *Behavioural and Brain Sciences*, 32(6), 493–561.
- Mckay, R., Langdon, R., & Coltheart, M. (2007). Jumping to conclusions? Paranoia, probabilistic reasoning, and need for closure. *Cognitive Neuropsychiatry*, 12(4), 362–376.
- Mellor, C. S. (1970). First Rank symptoms of schizophrenia. *The British Journal of Psychiatry*, 117, 15–23.
- Mellor, C. S. (1988). Crane's waterfall illusion. *Analysis*, 48(3), 147–150.
- Mishara, A. L. (2010). Klaus Conrad (1905–1961): Delusional mood, psychosis and beginning schizophrenia. *Schizophrenia Bulletin*, 36, 9–13. Clinical Concept Translation-Feature.
- Mishara, A. L., & Corlett, P. (2009). Are delusions biologically adaptive? Salvaging the doxastic shear pin. *Behavioural and Brain Sciences*, 32, 530–531.
- Mishara, A., López-Silva, P., Rosen, C., & Heinz, A. (2023). Self-Disturbances, perceptual anomalies, and physicality: Towards a Multimodal Model of Thought insertion. In P. López-Silva, & T. McClelland (Eds.), *Intruders in the mind: Interdisciplinary perspectives on thought insertion*. Oxford University Press.
- Mullins, S., & Spence, S. (2003). Re-examining thought insertion. *British Journal of Psychiatry*, 182, 293–329.
- Pacherie, E., Green, M., & Bayne, T. (2006). Phenomenology and delusions: Who put the 'alien' in delusions of alien control? *Consciousness and Cognition*, 15(3), 566–577.
- Parnas, J., & Handest, P. (2003). Phenomenology of anomalous self-experience in early Schizophrenia. *Comprehensive Psychiatry*, 44(2), 121–134.

- Parnas, J., Urfer-Parnas, A., & Stephensen, H. (2021). Double bookkeeping and schizophrenia spectrum: Divided unified phenomenal consciousness. *Eur Arch Psychiatry Clin Neurosci*, 271, 1513–1523. <https://doi.org/10.1007/s00406-020-01185-0>
- Parrott, M. (2012). Subjective misidentification and thought insertion. *Mind & Language*, 32(1), 39–64.
- Payne, E. (2013). *Speaking to my madness*. CreateSpace.
- Payne, R. (2015). *Pre-psychotic experiential alterations*. Unpublished interview.
- Proust, J. (2009). Is there a sense of Agency for thoughts? In L. O'Brien, & M. Soteriou (Eds.), *Mental actions*. OUP.
- Ratcliffe, M. (2013). Delusional atmosphere and the sense of unreality. In G. Stanghellini, & T. Fuchs (Eds.), *One Century of Karl Jaspers' General Psychopathology* (pp. 229–244). Oxford University Press.
- Saks, E. R. (2007). *The centre cannot hold. My journey through madness*. Hyperion.
- Sass, L. (1994). *The paradoxes of Delusion: Wittgenstein, Schreber, and the schizophrenic mind*. Cornell University Press.
- Sass, L. A., & Philosophy (2004). *Psychiatry & Psychology* 11(1), 71–80. <https://doi.org/10.1353/ppp.2004.0047>.
- Sass, L. A., & Pienkos, E. (2013). Delusion: The phenomenological approach. 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. 632–657). Oxford University Press.
- Silverstein, S. M., & Uhlhaas, P. J. (2004). Gestalt psychology: The forgotten paradigm in abnormal psychology. *The American Journal of Psychology*, 117(2), 259–277.
- Sollberger, M. (2014). Making sense of an endorsement model of thought insertion. *Mind & Language*, 29(5), 590–612.
- Sousa, P., & Swiney, L. (2011). Thought insertion: Abnormal sense of thought agency or thought endorsement? Phenomenology and Cognitive Science. <https://doi.org/10.1007/s11097-011-9225-z>
- Sterzer, P., Mishara, A. L., Voss, M., & Heinz, A. (2016). Thought insertion as a Self-Disturbance: An integration of predictive coding and phenomenological approaches. *Frontiers in Human Neuroscience*, 10, 502. <https://doi.org/10.3389/fnhum.2016.00502>
- Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, 84(9), 634–643. <https://doi.org/10.1016/j.biopsych.2018.05.015>
- Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, 17, 219–239.
- Uhlhaas, P. J., & Silverstein, S. M. (2005). Perceptual organization in schizophrenia spectrum disorders: Empirical research and theoretical implications. *Psychological Bulletin*, 131(4), 618–632.
- Vosgerau, G., & Voss, M. (2014). Authorship and control over thoughts. *Mind & Language*, 29(5), 534–565.
- Ward, J., & Cytowić, R. E. (2006). Synesthesia and language. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (Second Edition) (371–376). London: Elsevier.
- Wegner, D. (2002). *The illusion of conscious will*. Bradford Books.
- Wilkinson, S., & Alderson-Day, B. (2016). Voices and thoughts in psychosis: An introduction. *Rev Philos Psychol*, 7(3), 529–540.
- Young, A. W., & Leafhead, K. (1996). Betwixt life and death: Case studies of the Cotard delusion. In P. Halligan, & J. Marshall (Eds.), *Method in madness: Case studies in Cognitive Neuropsychiatry*. Psychology.
- Zawislawska, M. (2019). *Metaphors and senses*. Peter Lang.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.