Received: 5 June 2021 | Revised: 2 October 2021 | Accepted: 22 March 2022

DOI: 10.1111/mila.12441

SUBMITTED ARTICLE

WILEY

A vessel without a pilot: Bodily and affective experience in the Cotard delusion of inexistence

Philip Gerrans

Department of Philosophy, University of Adelaide, Adelaide, South Australia, Australia

Correspondence
Philip Gerrans, Department of Philosophy, University of Adelaide, Napier Building, North Terrace, Adelaide, SA 5005 Australia.
Email: philip.gerrans@adelaide.edu.au

Funding information
Australian Research Council, Grant/Award Number: DP 180101323

The initial cause of Cotard delusion is pervasive dyshomeostasis (dysregulation of basic bodily function). This explanation draws on interoceptive active inference account of self-representation. In this framework, the self is an hierarchical predictive model made by the brain to facilitate homeostatic regulation. The account I provide is an alternative to two factor accounts of the Cotard delusion that treat depersonalisation experience as the first factor in genesis of the Cotard delusion. I argue that depersonalisation experience and the Cotard delusion are produced by different breakdowns in the process of self-modelling.

KEYWORDS
active inference, avatar, Cotard delusion, depersonalisation, self awareness, self-model

1 | INTRODUCTION

A 48-year-old man with no medical history, apart from a previous short depressive illness, was seen by a psychiatrist after a self-electrocution attempt. Eight months later, he first told his general practitioner that his brain had died. He further explained that "I am coming to prove that I am dead", that he no longer needed to eat or sleep and was condemned to a kind of half-life, with a dead brain in a living body. He acknowledged that his abilities to see, hear, think, remember and communicate proved that his mind must be alive: He could not explain how his

mind could be alive if his brain was dead, but he was certain that this was the case. Psychotropic treatment had little therapeutic effect and his delusion receded only to return (Charland-Verville et al., 2013, p. 1997).

The Cotard delusion of death or inexistence was first described by Jules Cotard in 1882 as the "Delire des Negations". The delusion is often preceded by or associated with beliefs about distortion, derealisation or disappearance of body parts or functions and/or of the world itself (Billon, 2016; Cotard, 1882; Debruyne et al., 2009; Dieguez, 2018).The delusion forms part of a wider Cotard syndrome whose symptoms reflect dysregulation of basic bodily functions and a loss of the sense of reality of self and world. Cotard delusion has been found in cases of affective disorders and schizophrenia, neurosyphilis, acute encephalitis, sub-dural hemorrhage, arteriovenous malformations, brain neoplasms, migraine, Parkinson's disease, semantic dementia, cerebrovascular disease and epilepsy, among others (Restrepo-Martínez et al., 2019; Ramirez Bermúdez et al., 2021).

The philosophical interest of the delusion is obvious. Its paradoxical content simultaneously reports and denies the reality of the self. It also provides an avenue for the development of theories of self-representation and self-awareness within the field of cognitive neuropsychiatry. The most comprehensive philosophical account to date is due to Alexandre Billon (2016), who integrated the insights of the early psychiatrists and phenomenologists with recent structural philosophical accounts of subjectivity. He describes Cotard delusion as belief formed to explain loss of the sense of "mineness". Mineness is a philosophical term of art that refers to the feeling that experiences belong to the self, understood as a unified persisting entity (Billon (2017); García-Carpintero & Guillot, forthcoming). Loss of mineness produces the detachment that forms the core of depersonalisation experience. Billon integrates this account with explanations of depersonalisation experience and delusion developed by cognitive neuropsychiatrists that argue that delusions have two factors: anomalous experience (the first factor) and a belief that explains that experience (the second factor). Billon's (2016) two factor account of Cotard delusion makes depersonalisation experience the first factor and the delusional metacognitive response to that experience the second factor.

Billon's (2016) account abstracts from neural mechanisms but the two factor framework has been used to interpret imaging studies of the delusion and integrate them with studies that implicate insula cortex dysfunction as the basis of depersonalisation experience. Integrating Billon's account with these neuropsychiatric accounts provides a combined account in which insular malfunction produces the experience of depersonalisation, and abnormal reasoning based on prefrontal cortical dysfunction produces the delusional belief as a metacognitive response to that experience. The structure of this combined account is given in this diagram borrowed from Restrepo Martinez et al. (2019):

Such models require further cognitive theorising to explain why and how the neural correlates produce the phenomenology. I argue that the best theory of the relationship between insula functioning, "mineness" and depersonalisation suggests two things. Firstly, that the sense of mineness lost in classic "dissociative" depersonalisation experience is essentially a form of affective experience. This is a view that Billon rejects, preferring a "structural" account of mineness. Secondly, that the anomalous feeling state that comprises the first factor in many typical cases of Cotard delusion is not depersonalisation but the experience of pervasive and intractable *dyshomeostasis*. That is to say inability of the brain to regulate very basic bodily and related cognitive functions (things like sleep/wake cycle, arousal, appetite and basic psychomotor and visceromotor function) (Stephan et al., 2016).

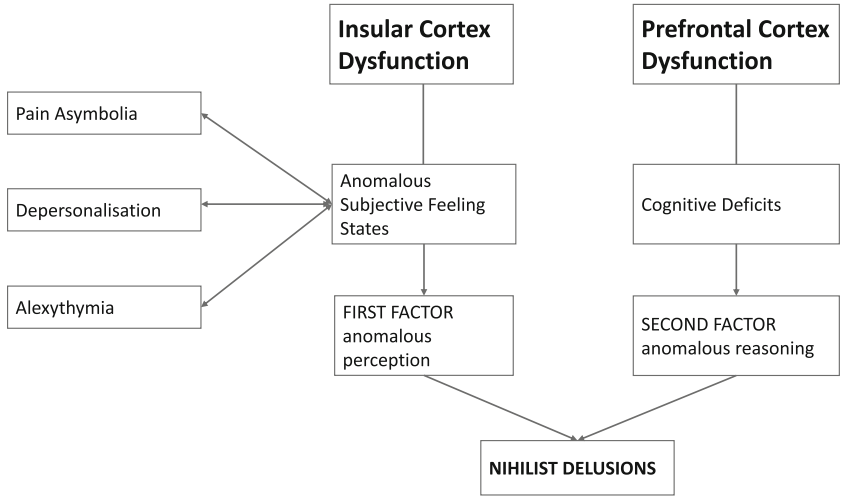

FIGURE 1 A two-factor model of the Cotard delusion. Factor 1 is depersonalisation experience based in insula dysfunction. Factor 2 metacognitive deficit based on prefrontal dysfunction

These conclusions are products of an interoceptive active inference theory of self-representation (Gerrans, 2020; Pezzulo et al., 2015; Seth & Friston, 2016). On that account the experience of mineness is the result of hierarchical self-modelling, an integrative processes that attributes signals generated in the process of homeostatic regulation to a unified persisting entity. The result is interoceptive experience of what has been called a core self, bodily self, or "material me" (Hohwy & Michael, 2017; Hohwy & Seth, 2020; Kirchhoff et al., 2018; Limanowski & Friston, 2018; Seth, 2013; Seth & Tsakiris, 2018; Tsakiris & Critchley, 2016; Tsakiris & De Preester, 2018). This basic, bodily, sense of self subtends an affective sense of self (the subject of emotional states and feelings) as well as sensorimotor and agential senses of self *qua* originator and controller of action. The so-called narrative self is an explicit conceptual representation of the self as the protagonist of an autobiographical trajectory comprised of episodes of first-person experience. Thus, the narrative I reports and communicates experience generated at lower levels of self-representation (Goldie, 2011).

Within this framework, depersonalisation experience involves a selective failure to annex a class of experience to a largely intact hierarchical self-model. The Cotard delusion is a result of the destruction or degradation of that model. It is consistent with this view that there can be overlap between symptoms of depersonalisation disorder and Cotard syndrome, especially while the delusion develops because of nature of self-modelling. However the delusion reports the experience of human life without an intact regulatory self-model whereas in depersonalisation the self-model is largely intact.

The sense of the self as the entity that sustains affective states plays a crucial integrative role in this hierarchy. The affective sense of self is produced by the metarepresentation, integration and interpretation of homeostatic signals according to emotional context. This is a process of what Clark and Karmiloff-Smith (1993) called representational redescription and I call "affective transcription". This interpretation of information about basic body state in terms of emotional salience produces a qualitative change in experience, from purely bodily to affective. An example is raised respiratory and heart rate during exercise and an episode of panic. The feeling of uncontrollable anxiety is an affective state, the feeling of racing heartbeat is an interoceptive

state. The increase in arousal is the same but the experience is qualitatively different when affectively transcribed in terms of inability of the subject to cope with uncertain threat.

This article endorses a thesis that depersonalisation experience is the result of hypoactivity in the insula (Medford, 2012; Medford et al., 2016; Medford & Critchley, 2010; Phillips et al., 2001; Sierra et al., 2012). This hypoactivity is the result of spontaneous inhibition typically arising as a dissociative response to stress or trauma. As a result, affective transcription is disabled leading to loss of the affective sense of self while other levels of self-modelling remain (relatively) spared. So, on this view, depersonalisation experience is the result of loss of the affective level of hierarchical self-modelling. Cotard syndrome is the result of a catastrophic breakdown of self-modelling at all levels caused by cascading effects of intractable basic bodily dysregulation or *dyshomeostasis* as it has been baptised (Stephan et al., 2016). If this is correct, then depersonalisation experience is not the first factor in the Cotard delusion although it may be a part of the syndrome of bodily affective and cognitive problems caused by dyshomeostasis.

In the rest of this article, I develop and defend this idea and compare it to current two factor models of the Cotard delusion (Figure 1). I first present Alexandre Billon's phenomenological account of the Cotard delusion and show how the structure of his two-factor account is consistent with the recent neurocognitive account of Ramirez Bermúdez et al. (2021) and collaborators. I then introduce the interoceptive active inference self-modelling account and (i) show how it supports Sierra's (2009) "affective" account of the sense of mineness and its loss in depersonalisation experience; (ii) consider objections to this affective account of insula function deriving from two sources and argue that they mischaracterise the nature of the affective problem in depersonalisation experience and (iii) argue that depersonalisation and the Cotard delusion are produced by different breakdowns in the process of interoceptive self-modelling. As well as refining the explanation of the Cotard delusion (Gerrans, 2001), the account provides case study of the interoceptive active inference account of self-representation. Ultimately, on that account, subjective experience produced by interoceptive self-modelling produces an *avatar*, a representation of a simple unified entity that allows the subject to integrate myriad interacting subsystemic functions whose operations are inaccessible to explicit cognition. The feeling of being a unified self plays the same regulative role as an airplane icon on the control screen of an airplane. The pilot who banks the plane by manipulating the icon has no direct access to the systems that keep the plane in the air and on course. The icon is a high-level abstract avatar whose purpose is to provide a single, simple, target of systemic regulation.

Similarly, the feeling that we are unified persisting entities is the result of an adaptive cognitive strategy of simplification and abstraction that enables us to predict and manage our organismic trajectory through life by managing the experience of selfhood (Metzinger, 2011).

2 | THE SENSE OF MINENESS

Below is a description of a case of Cotard delusion arising in the context of anti-NMDAR encephalitis. This patient was the subject of one of the recent and rare imaging studies of this disorder. I quote in detail because it gives a sense of the pervasive and disturbing nature of the symptoms of Cotard *syndrome*. A focus on the propositional content of the delusion, while understandable in the context of cognitive theorising, shifts the explanatory focus away from the precipitating experience of the syndrome, which is certainly not confined to depersonalisation.

> After claiming to be dead, Mr. A, a 19-year-old man, was referred to the National Institute of Neurology and Neurosurgery of Mexico, because of psychotic and catatonic symptoms. According to his mother, one month before admission, his appetite had diminished, he barely slept at night and, during the day, he stayed withdrawn and silent. Mr. A would complain about his body, saying that his tendons had dried up and that his organs, particularly his heart, were getting bigger. He would speak of an internal haemorrhage: all his organs had been torn apart. Mr. A began to insist that he was dead: "I can no longer feel the blood flowing through my veins". This assertion did not deprive him of looking for sharp objects to cut his neck and forearms, he tried on multiple occasions to commit suicide. Gradually, his behaviour became disorganised: he began to urinate on his clothes, refused to eat, and his verbal production diminished to mutism...
>
> Four days after he received his fifth session of plasma exchange, his alertness improved, but psychotic symptoms remained. When asked how he felt, he answered: "*I do not have feelings because I am dead... All human beings will be dead*." He added that he had a "right-sided" heart and that it had stopped.
>
> The patient was transferred to the neuropsychiatric unit. Over the following days, catatonic symptoms continued to improve (BFCRS score of 8/21), but nihilistic delusions persisted unchanged: "I am dead, among other human beings. This is like a program where I am dead". The patient claimed to be a dead person among the corpses under the earthquake debris. Doctors and nurses were part of a game whose objective was to cover up his death. (Restrepo Martinez et al., 2019, pp. 471–472)

Alexandre Billon has proposed that the Cotard delusion is a response to the experience of depersonalisation. Depersonalisation is a condition in which subjects feel detached from experience as though it does not belong to them. For example:

> It almost *feels like* [as if] *I have died*, but no one has thought to tell me. So, I'm left living in a shell that I don't recognize any more (Sierra & David, 2011; my italics).

On the combined two-factor account the depersonalised person feels "as if" they do not exist but because they retain a capacity for rational insight they do not develop a delusion. As Alexandre Billon says:

> The Cotard syndrome can thus be characterised as the delusional form of depersonalisation, depersonalisation being, conversely, the "as if" form of the Cotard syndrome. (Billon, 2016, p. 357)

On this view depersonalisation experience is the essence of Cotard delusion (the first factor) and the second factor consists in a metacognitive abnormality that means the subject *endorses* her experience. As he says, Cotard's nihilistic delusions, I argue, simply *consist in taking this altered phenomenology at face value* (Billon, 2016, p. 357). Restrepo-Martínez et al. (2019), adopt the same approach in their interpretation of their imaging studies implicating insular cortex dysfunction.

> [I]nsular cortex dysfunction would elicit *anomalous subjective feeling states*, as well as well-known psychopathological constructs as pain asymbolia, depersonalisation, and alexithymia. These *anomalous perceptions* would correspond to the first factor necessary for delusions formation. Prefrontal cortex dysfunction would explain the second factor in delusion formation: *an abnormal way of reasoning that explains how anomalous perception cannot be denied* (Restrepo-Martínez et al., 2019, p. 472; my italics)

Combining the accounts of Bermudez et al. and Billon we have the idea that insular cortical dysfunction produces the experience of depersonalisation, and the delusional belief endorses the content of that experience.

Billons' (2016) account is not a mechanistic account, so it is not proposed as an explanation of the role of neural correlates. If, however, we accept a broad consensus that hypoactivity of the insula, especially the right anterior insula, is a neural correlate of depersonalization experience then we can ask why and how insula hypoactivity produces the feeling that experiences do not belong to the subject. This requires a neurocomputational account of the functioning of circuitry implicated in Cotard syndrome and depersonalization. That account can be derived from overlapping themes in predictive processing theories of interoception, emotion and self-awareness.

3 | INTEROCEPTIVE ACTIVE INFERENCE

Predictive processing is a neurocomputational implementation of the free energy principle (FEP). The FEP starts from two ideas. Firstly, that organisms act to maintain their integrity in the face of entropic forces. The most basic form of optimization is homeostatic regulation. Secondly, in so doing they embody a representation of themselves as probabilistic model that predicts their trajectory in homeostatic state space. Organisms act to reduce "surprisal" or Shannon entropy, a measure of amount of information unpredicted by that model (Allen & Friston, 2018; Corcoran & Hohwy, 2017; Friston, 2009; Kirchhoff et al., 2018; Limanowski & Blankenburg, 2013). The predictive processing neurocomputational implementation of this idea treats surprisal as a "prediction error" a signal that the sensory consequences of a regulatory action are not predicted by a model. The minimization of prediction error is an iterative hierarchical process in which higher-level models of the organism and its sensory domain predict the sensory consequences of action and the organism acts to realise that prediction. Unpredicted sensory consequences generate error signals that propagate up the hierarchy leading to model revision and/or recalibration of action (Friston & Kiebel, 2009; Hohwy, 2013; Sel, 2014; Seth et al., 2011; Van de Cruys, 2017). This minimization of predictive error produces a model of the structure of the external world as the inferred cause of perceptual experience and of the self as the inferred cause of interoceptive (representation of the internal *milieu*) experience.

Within this framework we need to add the concepts of allostasis and active inference. Allostasis extends the concept of homeostasis to incorporate:

> [A] smooth continuum of adaptive action selection, ranging from the primitive drives that work (for instance) to sate appetite via exploitation of the immediate environment, to the complex deliberative activities serving various motivations extending well beyond the basic requirements of internal milieu (Corcoran & Hohwy, 2017, p. 10).

Active inference refers to the role of action in optimising a model by sampling or exploration to reduce uncertainty by gaining new sensory information. A metaphor for active inference is cognitive foraging, searching in solution space to maximise fit between predictive models and sensory information by accumulating new sensory evidence (Allen & Friston, 2018; Friston et al., 2015; Gerrans, 2014; Mirza et al., 2016). In perception, active inference is *epistemic*, yielding knowledge about the structure of the external environment. Perceptual active inference (such as visual saccades) verifies a detailed model of a world of objects and properties because higher-level models that structure signal processing predict that the world has that structure. In interoception active inference is *instrumental*, reflecting the role of interoception in homeostatic regulation (Seth & Tsakiris, 2018). Interoceptive experience represents the overall state of the system relative to a predicted state, given a model, in order to calibrate regulatory action (Joffily & Coricelli, 2013; Seth & Tsakiris, 2018; Van de Cruys, 2017; Weise, 2018).

In the abstract sense favoured by formal models of predictive processing, all of life—cognitive and behavioural—can be reframed as active interoceptive inference: taking action to optimise a model that predicts states of the organism that best maintain its functioning given interactions with the world. The example of fatigue applies the idea.

Fatigue is an interoceptive experience that represents systemic energy depletion. It integrates innumerable subsystemic signals to give an overall experiential readout of organismic state. This is important because the mechanisms of energy restoration are not available to introspection or direct regulation. We cannot directly change levels of adenosine or glycogen but we can rest when tired. So, managing fatigue at the systemic level is a proxy for the management of subsystemic function. The organism takes the action that reduces fatigue in context (sleep, drink coffee) and in the process consolidates a model of the self *qua* source of interoceptive experience (Allen & Friston, 2018; Friston et al., 2013; Kirchhoff et al., 2018; Limanowski & Friston, 2018; Lyon et al., 2021). Thus interoceptive regulation produces a basic sense of bodily selfhood. A material me, as Seth calls it. An extra layer of processing allows states of the bodily self to be experienced as affective states. And, as with interoceptive experience, affective experience serves as a regulatory proxy. Modulating affective states through action is a way of optimising bodily function (Fernandez Velasco & Loev, 2021; Gerrans, 2019).

When the mind loses the capacity to predict and influence the flow of affect (positive and negative) through action it loses a form of self-awareness. As Deane et al. (2020, p. 8) put it: "Losing a sense of allostatic control, the system ceases to posit itself as a causally efficacious controller of sensations". Depersonalisation is one such case, Cotard is another more drastic with a different aetiology.

4 | DEPERSONALISATION AS DEAFFECTUALISATION

In depersonalisation experience subjects report feeling detached from their experience, including bodily experience, as if they are not the subject of that experience (Baker et al., 2003; Ciaunica, Charlton, et al., 2021; Ciaunica, Roepstorff, et al., 2021; Jay et al., 2014; Medford et al., 2016; Phillips et al., 2001; Sierra, 2009; Sierra & David, 2011) This is the loss of "mineness" referred to by Billon (2017) and Billon (Forthcoming). However, interoception itself remains intact. As Michal et al. (2014) reported:

> While they feel detached from their body and report emotional or physical numbing, *actual subjective body perception is unimpaired*, and heartbeat detection [a task used as a measure of interoceptive ability] similar to normal, healthy volunteers. (Michal et al., 2014, p. 6).

This conclusion about preserved interoception fits with a standard characterisation of depersonalisation experience as one in which awareness of bodily state is preserved along with higher-level cognition and conceptual or narrative forms of representation. "In spite of the dramatic nature of the experience, patients 'remain aware of the unreality of the change. The sensorium is normal and the capacity for *emotional expression* intact'" (Sierra & David, 2011, p. 99). Thus the loss of the sense of mineness in depersonalisation does not derive from a failure of basic interoceptive self-modelling. Depersonalisation dramatizes the need for an account that allows us to distinguish purely bodily, or interoceptive experience, from affective experience and the sense of mineness, although they are very closely related and have overlapping neural substrates.

Depersonalisation experience is part of a syndrome that often has a progressive character in which the world or body parts feel increasingly unreal, a phenomenon called "derealisation". Ultimately, the subject reports feeling as if experience does not belong to her. The explanation I propose is that experience is not modelled as belonging to the self. This can happen for particular classes of experience, or globally, and at different levels of self-modelling.

My focus here is on global forms of depersonalisation experience that arise when bodily self-modelling is intact but affective experience cannot be generated and incorporated into the self-model. This type of depersonalisation is the subject of the most comprehensive neurocognitive accounts of depersonalisation and at the same time exemplifies the key point that depersonalisation results from failure to incorporate experience into a largely intact self -model. Sierra and David describe, depersonalisation as:

> a state of emotional numbing and disable the process by means of which perception (including that of one's own body) as well as cognition become emotionally colored. Such "decoloring" will result in a "qualitative change" of conscious awareness, which is then reported by the subject as "unreal or detached". (Sierra & David, 2011, p. 99)

The hypothesised mechanism is spontaneous inhibition of anterior insula cortex by the ventrolateral prefrontal cortex (Medford, 2012; Medford et al., 2016; Medford & Critchley, 2010; Sierra et al., 2012): "[D]epersonalization represents an anxiety-triggered 'hard-wired' inhibitory response intended to ensure the preservation of adaptive behavior during situations normally associated with overwhelming and potentially disorganizing anxiety" (Sierra, 2009, p. 19).

Sierra (2009) explains this inhibition as a dissociative response to negative experiences such as pain, fear, or anxiety. The mechanism is primarily inhibition of anterior insular activity by the ventrolateral prefrontal cortex and other frontal structures involved in emotional regulation and higher order evaluation. A number of studies identify hypoactivity of anterior insula cortex in a range of voluntary, involuntary and drug-induced depersonalisation experiences providing: "powerful evidence that a lack of anterior insula activity is related to the diminished emotional responsiveness seen in DPD, and that a "re-awakened" insula is seen when patients improve and de-affectualization symptoms (and DPD symptoms more generally) are ameliorated. (Medford et al., 2016, Discussion, para. 10).

This account does raise one question. Namely, why is deaffectualisation experienced as loss of mineness or detachment from experience? After all the subject could report feeling affectively flat rather than "as if" not present in the experience. The answer to this question is provided by the hierarchical predictive self-modelling account (Allen & Friston, 2018; Deane et al., 2020; Gerrans, 2014; Van de Cruys, 2017).

Affective regulation is a crucial form of higher order interoceptive active inference that depends on the ability to experience the body not just *qua* physical system but under an emotional mode of presentation as a result of affective transcription. And as with predictive processing in general active inference accumulates evidence for a model of the self, inferred to be the entity that sustains these affective states.

This is a substantive hypothesis about the relationship between interoceptive and affective experience because in principle there are other ways to account for the relationship between interoception and affect. For example, it could be the case that interoceptive representations are *associated* with emotional representations rather than interpreted by them. However, the most persuasive accounts of insula functioning treat the anterior insula as a hub of processes that metarepresent and interpret (or as I call it transcribe) bodily signals allowing them to be experienced with affective tone.

In order to transcribe interoceptive signals affective processes exploit extra layers of processing that metarepresent interoceptive signals integrate them with knowledge that contextualises them (Barrett, 2017; Barrett et al., 2016) and re-represents them. This relationship between interoceptive processing and affective processing is reflected in cytoarchitecture (Barrett & Simmons, 2015). Sections of the posterior insula cortex take primary interoceptive afferents and integrate those representations to coordinate basic regulatory functions (Garfinkel & Critchley, 2013; Medford & Critchley, 2010; Moayedi, 2014; Terasawa et al., 2013).

This information about body state is re-represented and integrated with other information to transform purely interoceptive information into a representation of emotional salience. Circuitry implicated as the substrate of emotional processes such as the limbic and paralimbic systems co-ordinates processing across the mind that detects and determines emotional significance and initiates and regulates appropriate response. For example to determine whether an external (e.g., perceived object) or internal (feeling of pain or discomfort) stimulus represents a danger the organism must deploy a suite of cognitive systems that represent the stimulus, its relation to the goals of the subject, the capacity to respond, and monitor the success of the response. No discrete system is responsible for all and only the information processing relevant to these processes of evaluation (or appraisal as they are sometimes known) because the nature and degree of emotional salience is context dependent and handled at different levels from perceptual-reflexive to conceptual-reflective. Thus, emotional processing reflects a general feature of the active inference framework: the configuration by the mind of its resources to optimise functioning at different timescales. This feature of the active inference account explains why the circuitry involved in emotion and self-representation is largely polymodal. The representational structure implemented by activity in these circuits depends on the way they are configured in context rather than being "wired in" (Allen, 2020; Allen & Friston, 2018; Betti & Aglioti, 2016; Clark, 2015).

On this view, how the organism deals with information about body state integrated primarily by the posterior insula (e.g., heart rate) depends on its constant evaluation (appraisal) of its prospects in the world. What we call emotional processes underpin those evaluations. As a result primary interoceptive representation in the posterior insula is metarepresented and interpreted by emotional processes in terms of its goal relevance. For example, elevated heartbeat in

an episode of exercise is experienced differently to the same heartbeat level in an episode of anger or panic because of the different emotional context. When transcribed by emotional processes interoceptive signals are experienced as affects.

The anterior insula is a relay station between primary interoceptive representation and the emotional processing that co-ordinates these appraisals and transcriptions. It is thus a crucial hub of the distributed processing that transcribes interoceptive representation of body state into an affective representation that allows us to feel the significance of interoceptive states as affects.

This role for the anterior insula helps explain its enigmatic functional promiscuity and ubiquity. As Allen says it is "implicated in nearly every imaginable cognitive, sensory, and affective domain" (Allen, 2020, p. 265). It coordinates bodily perceptual and cognitive processing that determines how information represented across the mind bears on the organism's ability to achieve its goals. For this reason, the posterior insula is sometimes treated as a locus of bodily representation and anterior as the basis of affective representation (the classic literature on pain processing takes this approach, for example, Danziger, 2006). However, the distinction is not sharp. Rather, patterns of activity across the insula tend to be graded and continuous with more anterior activity more closely associated with affective experience in virtue of its connections with systems such as the ACC and ventromedial prefrontal cortex that are implicated in self-referential processing.

> A continuous gradient from multisensory and embodied input-integration to complex behavioural regulation as one moves along the posterior-to-anterior axis ... posterior insula exhibits mostly multi-sensory responses and is broadly connected to thalamic and primary sensory regions, the anterior insula is instead responsive to attentional (salience, response inhibition) and affective (emotion regulation and awareness) conditions and sends projections to the parietal-frontal control regions and brainstem nuclei ... with both profiles being freely mixed in the middle insula. Interoceptive predictive coding thus argues that the insular cortex integrates low-level sensory prediction errors with interoceptive and attentional expectations to regulate affective salience and emotion (Allen & Friston, 2018, pp. 2468–2469).

It is for this reason that Bud Craig (2009) and Seth (2013) developed the idea that in virtue of performing this role activity in anterior insula cortex creates a sense of self. One way to put this is to say that the feeling that one's body is "mine" is initially generated by interoceptive representation dependent on posterior insular activity, consistent with research into the experience of bodily ownership and awareness (Baier & Karnath, 2008). However, that activity is not independent of activity in mid and anterior insula that is constantly informing the system, not only of body state, but the emotional significance of body state in order to transcribe bodily into affective experience.

Seth adds the idea that the process is not feedforward or bottom up but is an instance of predictive processing that uses higher predictive level models (Seth, 2013; Seth & Friston, 2016). This is important because on the predictive view the nature of high-level models determines whether and how a bodily signal becomes salient and the nature of resultant active inference. For this reason, it is actually quite rare to experience a body state untranscribed, because the mind is always modelling affective consequences of action to sculpt the interoceptive signal "perception and action are now fundamentally affective and embodied in nature, possessing a salience (epistemic) or inherent (pragmatic) value for the organism in homeostatic terms" (Allen & Friston, 2018, p. 2470).

Depersonalisation helps make the point. The mind predicts that sensory experience will have an affective signature and when, intractably and unpredictably, it does not, the world feels strange and unreal and the subject detached from her own experience. If the anterior insula cortex is not functioning normally to transcribe interoceptive signals, other dimensions of self-representation, bodily, agential, sensorimotor and narrative are intact, but the agent does not feel as if any of the resultant experiences *matter to her*. The important point for disorders of the self is that this loss of anterior insula activity is unpredicted. The subject is confronted by an intractable mental experience of loss of a sense of affective self awareness predicted by the context of action.

5 | STRUCTURAL AND AFFECTIVE ACCOUNTS OF MINENESS. ROLE OF THE INSULA

This affective transcription account makes the sense of mineness lost in depersonalization a phenomenon of deaffectualisation. It integrates accounts of insula function derived from predictive processing theories of emotion and self-awareness and clinical and imaging studies of depersonalization. However, we should note that it is not uncontested. There are two sources of disagreement: one phenomenological and one drawn from neuroscience. It is important to show how the self-modelling account can deal with these objections because these responses help explain the relationship between Cotard syndrome and depersonalization.

Alexandre Billon (2017) is very clear that mineness cannot be an affective phenomenon. He points to reports like that of Schilder who wrote of a patient

> The objective examination of such patients reveals not only an intact sensory apparatus, but also *an intact emotional apparatus*. All these patients exhibit *natural affective reactions* in their facial expressions, attitudes, etc.; so that it *is impossible to assume that they are incapable of emotional response*. (Schilder, 1935; in Billon, 2017, pp. 206–207; my italics)

Billon uses this example as part of his structural account of depersonalisation. On this account mineness is a constitutive feature of experience that locates experiences as episodes of a particular subject with a unique autobiographical trajectory and perspective on the world. Mineness is ubiquitous (that is, a structural feature of all experience) and tacitly expected. Thus, when this structural element of experience disappears, the subject experiences both a negative sense (of unpredicted loss of mineness) and a positive sense (of unfamiliarity) as the world and her experience of it are rendered unfamiliar because she is no longer automatically located in it.

The crucial point is that on Billon's account mineness is a structural feature of *all* experience including affective and bodily (interoceptive) experience. Consequently, on his account it must be possible for subjects to lose the sense of mineness, not only from bodily and sensory experience and cognition, but also affective experience.

We should note, however, that the self-modelling account does not reduce "mineness" to interoceptive signalling. Rather it says that interoception integrates disparate interoceptive signals by attributing them to a unified subject. If that integration fails for a particular channel, then signals from that channel will not be experienced as mine: One will be depersonalised, so to speak, for experience generated by that channel. The fact that there are forms of depersonalisation in which a subject can experience bodily states without feeling a sense of

mineness actually counts in favour of the self-modelling account. In such cases, something has gone wrong with the ability to integrate that bodily signal into the interoceptive/bodily self-model. A very clear case is pain asymbolia in which nociceptive signalling (of bodily damage or threat to bodily integrity) is intact but the subject does not feel as if the pain is "hers". For this reason, it is aptly described by Klein as "depersonalisation for pain". In such cases, the subject does not feel pain signals as distressing because, not being incorporated into the self-model, they are not affectively transcribed at higher levels (Betti & Aglioti, 2016; Danziger, 2006; Gerrans, 2020).

When we turn to self-modelling of affective states in depersonalisation things are slightly different. Billon (2017) in fact argues that affective states are intact in depersonalisation disorder but subjects nonetheless experience loss of mineness. Consequently, mineness cannot be an affective phenomenon. So his structural account poses a clear challenge to the affective self-modelling account of depersonalisation. The structural account can also draw support from interpretations of studies that suggest affective experience and self-awareness are not affected by lesion to the insula. For example Philippi et al. tested patient R "on a range of standard tests of self awareness (SA)" (my italics) and found

> R is a conscious, sel—aware, and sentient human being despite the widespread destruction of cortical regions purported to play a critical role in SA, namely the insula, anterior cingulate cortex, and medial prefrontal cortex (Philippi et al., 2012, p. 11).

The experimenters concluded that that Roger was self-aware in virtue of his ability to integrate very basic bodily signals arising in the brainstem with his declarative and semantic knowledge about his life and personality traits. "[W]e find little support for the hypotheses that implicate the insular cortex as critical to *all aspects* of SA" (Philippi et al., 2012, p. 14).

So, if we combine the structural account with interpretation of these lesion studies we could conclude that insula functioning is not necessary for affective processing or self-awareness. However, the lesion studies and the case quoted by Billon both used behavioural/verbal indices of affect and self-awareness.

This is crucial because affective experience is a dissociable component of episodes of emotion. On any theory of emotion an episode of emotion comprises physiological responses (such as autonomic and visceromotor responses experienced in interoception), action tendencies (semiautomatic prototypical behavioural responses), characteristic patterns of cognition and attention, emotional expression *and* affective experience (Brosch & Sander, 2013; Frijda et al., 1989; Sander et al., 2005; Scherer, 2004). While these components normally synchronise smoothly, they also have independent (though interacting) neural substrates that make dissociations possible. Such dissociations are an adaptation for emotional regulation. For example, instinctive action tendencies can be inhibited by bomb disposal experts and lion tamers who nonetheless are fully aware of the danger they are in. But they inhibit a component of the fear response (action tendencies) in order to engage effectively with the situation. Such engagements can take the form of higher order active inference in which situations can be re-evaluated, ultimately leading to alternate emotional evaluations and responses. Inhibitory capacities are thus an essential part of emotional processing, releasing us from stimulus bound automatic response patterns (Kalisch, 2009; Kalisch et al., 2006).

The effective component of an emotional episode can also be inhibited, either as a type of analgesic response that down regulates distress (Lee et al., 2014; Sierra, 2009), or as part of

higher-level emotional active inference. Sierra (2009) showed that voluntary inhibition of insula cortex produced "deaffectualisation experience" and the same inhibitory mechanism (activity in the ventrolateral prefrontal cortex) is seen in depersonalisation. It is quite consistent with the affective transcription account that affective experience can be inhibited while cognition perception and primary forms of bodily representation are intact. Indeed as Sierra noted, "[t]he sensorium is normal and the capacity for *emotional expression* intact" (Sierra & David, 2011, p. 99). This is consistent with typical accounts of depersonalisation. For example, in the same passage quoted by Billon, Schilder goes on to say: "The emotions ... undergo marked alteration. Patients complain that that they are capable of *experiencing* neither pain or pleasure; love and hate have perished with them" (Schilder, 1935; in Billon, 2017, p. 206).

This is consistent with Sierra's idea that what is lost in depersonalisation is affective *experience*, while other forms of bodily and perceptual experience are intact. Thus, the behavioural, physiological and cognitive aspects of an emotional episode can be intact but the patient does not feel the characteristic affective response. This gels with descriptions of deaffectualisation in depersonalisation, including this by a patient of Dugas and Moutier who said "I only feel anger from the outside, *by its physiological reactions*" (Dugas & Moutier, 1911; in Billon, 2017, p. 207; my italics).

The same point applies to the interpretation of lesion studies of the insula that argue that affect and self-awareness are preserved. In a battery of tests, Philippi et al. (2012) probed R's core, extended and introspective self-awareness. For example, R completed tasks like mirror self-recognition, self-other discrimination and performance on a:

> Questionnaire ... designed to measure personal insight, self-focused attention, and the endorsement of behaviors and thought processes characteristic of different types of SA. Some example items include, "Before I leave my house, I check how I look" and "It's hard for me to work when someone is watching me". (Phillipi et al., 2012, p. 5)

Their conclusion that R is self-aware is based on his performance in these tasks. However, these tasks are behavioural and perceptual/cognitive rather than designed to directly probe the structure of experience. It is possible to pass without relying on the experience of affective feelings. Indeed this view is actually consistent with Phillippi et al.' s conclusion that R was able to pass these tests using his intact semantic knowledge and basic bodily forms of awareness mediated by intact neural structures.

When we look at R's performance on other tasks that would normally involve the ability to experience affective states and incorporate them into a model of the self, a different picture emerges. For example, R does not seem to habituate to a series of painful stimuli. His responses, verbal and behavioural, are like startle responses. This is consistent with the idea that R is not distressed by pain or incorporating that distress into a self-model that predicts negative affect. Roger is simply experiencing the automatic transition from nociception to behavioural reaction (flinching, vocalisation). Thus pain is painful for R, in the sense that nociception is intact, but not affectively transcribed. In this respect, R resembles people with pain asymbolia, a condition in which patients report intact ability to detect pain but flattened affective response and sense of detachment while at the same time exhibiting exaggerated startle and aversive behavioural responses. The authors' interpretation is actually consistent with this view. Noting that R's limbic system is also lesioned they say that:

> [D]ata suggests that the limbic structures commonly associated with pain may play a fundamental role in pain regulation. Under this view, the missing regions in Roger's brain would impair his ability to control and downregulate his pain responses. (Feinstein et al., 2016, p. 1509)

These limbic structures are the hubs of emotional active inference that contextualise pain signals and allow the anterior insula to provide the emotional colouring (to borrow Sierra's nice phrase). This lack of ability to affectively transcribe bodily signals is reflected in the loss of a sense of mineness despite preserved semantic knowledge and habitual behavioural response.

5.1 | The first factor in Cotard delusion

I now turn to Billon's (2016) idea that depersonalisation experience is the first factor in formation of the Cotard delusion. On this view depersonalisation experience is the "as if" version of the Cotard delusion.

The Cotard syndrome includes feelings of derealisation and dissociation from body and external world that are part of the profile of depersonalisation. It is also clear, however, that Cotard syndrome is characterised by a network of interrelated symptoms that typically centre on the body. It also seems clear that Cotard delusion is not *essentially* dissociative. Although affective transcription and self-modelling are impaired the aetiology is not selective inhibition as a dissociative response. Rather, in Cotard syndrome the self-modelling hierarchy disintegrates. By this I mean that the integrity of the hierarchical processing structure that remaps homeostatic prediction error to produce interoceptive and affective senses of self-awareness degrades in a chaotic way. The disintegration is produced from the bottom up as very basic homeostatic functions become essentially uncontrollable (*dyshomeostasis* as it is called by Stephan et al., 2016), producing a cascade of prediction errors that overwhelm and ultimately destroy the self-modelling hierarchy. No attempt to minimize these errors is successful and so the mind *learns* that the self, *qua* source and target of regulation, no longer exists. Unlike depersonalisation, in which the self-model is largely intact, in Cotard delusion the model is destroyed. The result is that the narrative "I" remains as a purely verbal construct detached from the sense of self as a unified persisting entity.

Hohwy and Michael, in their discussion of self-modelling, explain why in principle this type of intractable dysregulation can produce the experience of inexistence:

> [T]he belief that I exist is equivalent to the belief that I act to maintain myself in certain states ... The reasoning is that I cease to exist if I cannot maintain myself in a limited number of states, that is, if my body begins to disperse across many states. I must act to prevent dispersion: inaction leaves me open to the unfettered impact of the second law of thermodynamics (Hohwy & Michael, 2017, p. 373).

The concept of dyshomeostasis developed by Stephan et al. (2016) applies this idea to explain the depression. This is relevant to the Cotard delusion because extreme depression even depressive psychosis, is part of its aetiology. On some accounts Cotard delusion is a result of self-accusatory thoughts of inefficacy and lack of worth characteristic of extreme depression. Cotard himself noted "le malade s'accuse lui meme". Such accounts emphasise the content of the delusion and its connection with a depressive pattern of thought rather than the experience

that generates it (Young & Leafhead, 1996). However, the regulatory or dyshomeostatic account provides a bottom up explanation in which thoughts of the depressive subject or Cotard subject *reflect* the loss of ability to anticipate or regulate trajectory in homestatic state space.

Stephan et al. (2016) use the example of sustained and intractable fatigue. When the subject cannot regulate basic biological functioning such as sleep control, actions cannot reduce interoceptive surprise and a metacognitive interpretation ensues that leads to the subjective feeling of fatigue. At higher levels of active inference, the subject tries cognitive and behavioural strategies to reduce fatigue. However all strategies fail, which teaches the subject that no matter what the subject does she cannot control her*self* at the most basic level. In other words:

> [A]n agent's experience of enduring dyshomeostasis signals a fundamental lack of mastery and control (over bodily states and thus survival) which may generalize, from the allostatic domain to other cognitive domains that are crucial for self-evaluation, planning and action selection. (Stephan et al., 2016 p. 17).

This idea postulates a self-modelling control hierarchy that links one form of allostasis, energy regulation, to depressive symptoms and patterns of thought ultimately reflected at the level of the narrative I. But there are numerous conditions that can induce pervasive and intractable dyshomeostasis, including cancers, chemotherapy and, tellingly, autoimmune diseases. In the latter cases, the dysregulation can be even more pervasive and resistant than fatigue. Patients discussed by Ramirez Bermúdez et al. (2021) had a form of encephalitis that attacks the NDMA system, a basic neural function. They describe a patient whose admission to hospital was preceded by feelings of physical and mental weakness following the Mexico City earthquake. His symptoms included:

> [H]eadaches, insomnia, decreased appetite, irritability, auditory and visual hallucinations, delusions, and prolonged periods of mutism, and aggressive behavior toward himself and others. He also became paranoid, believing that his family wanted to harm him...
>
> fluctuating levels of consciousness with periods of psychomotor agitation, inattention, and disorientation.
>
> autonomic instability and catatonic features, including motor excitement alternating with immobility, mutism, staring, and posturing. (Ramirez Bermúdez et al., 2021, p. 65)

In other words, he experienced a catastrophic and chaotic failure across the hierarchy of systemic regulation. In such a situation, there is *no* adequate bodily model that predicts the consequences of regulatory action at any level and no affective self-model that allows predictable transcription of such symptoms and can serve as a regulatory proxy. The subject cannot model herself as someone whose bodily and affective states are predicted by, and responsive to, her regulatory actions. She learns that she does not exist as an entity that underpins a predictable and controllable trajectory of experience. Of course, she continues to exist as a body, the source of chaotic degradation, but she no longer exists as the self who can respond or intervene. However, unlike depersonalisation, in which the self-model is largely intact, in Cotard delusion the model is destroyed. *Dyshomeostasis* in acute depression shows how this destruction can arise from the bottom up. But it can also arise from other causes (case studies here involved

electrocution and encephalitis). All these cases are unified by the idea that is that when the bodily is unregulable at the most basic level, organismic function cannot be modelled as a regulatory target. And, on the account presented here, it is activity in circuitry that implements the regulatory model that generates the experience of mineness.

Descartes (1984/1641, p. 56) famously said that interoceptive states such as hunger and thirst *teach* us that we are not "present in our bodies as a sailor is present in his ship, but rather that I am very closely joined and as it were intermingled with it, so *that I compose a single thing with it*" (my italics). Within the predictive processing framework we might say that the experience of being the subject of interoceptive and affective experiences produces the inference that those experiences belong to a simple unified entity (Constant, 2021; Hohwy & Michael, 2017). The regulation of those experiences confirms that inference. When I rest, I feel refreshed; when I eat, I am satiated; when I achieve a goal, I am happy; when I anticipate uncertain negative events, worried; and so on.

States of intractable dyshomeostasis teach us the reverse. In severe cases like NMDAR encephalitis the mind infers that there is no viable regulatory target because experience is chaotic and unpredictable. The subject is unable to keep entropy at bay.

Restrepo-Martinez et al. (2019, p. 5) noted imaging studies of Cotard syndrome in which a consistently reoccurring theme is insula hypoactivity "the neural basis of a first factor necessary for the development of nihilistic delusions". Their discussion mentions another imaging study of a case of Cotard delusion following extensive atrophy of insula cortex, "demonstrating the central role of insular cortex in self -perception and its ability to produce Délire des negations through dysfunction" (Chatterjee & Mitra, 2015, p. e53).

However, in the case of depersonalisation, the cause of anterior insula hypoactivity is relatively selective top down inhibition. In the Cotard delusion, the cause appears to be more bottom up. Rather than predicable and regulable interoceptive signals that can be affectively transcribed and regulated accordingly the self-modelling-hierarchy is bombarded by essentially chaotic signals. In such a situation, self-modelling progressively fails.

6 | A SECOND FACTOR IN COTARD DELUSION?

Both Billon (2016) and Ramirez Bermúdez et al. (2021), propose versions of two factor explanation of the Cotard delusion in which a metacognitive deficit leads to the formation of a delusional belief that endorses the content of depersonalisation experience. The main topic of this article is the precipitating experience in each case. However, the predictive processing framework also suggests ways to conceptualise the relationship between experience and delusion. For example, Corlett and others argue that in delusion high *precision* is assigned to signals of sensory prediction error with the result that higher level models are revised to fit the sensory evidence (Corlett, 2019; Corlett & Fletcher, 2015). Thus, we do not need to postulate a metacognitive deficit in a system of belief fixation since this revision consists in the consistent application of Bayesian principles across the modelling hierarchy. Belief fixation is conceived of in terms of precision weighting of higher order models relative to lower order perceptual models.

Another possibility that fits the framework is that higher order active inference, understood as the search for alternative evidence to verify explanatory models, is disabled in these cases because the inhibitory/executive resources required are unavailable (consistent with a role for hypoactivity of dorsolateral prefrontal structures) and or/the precision on the sensory evidence is too high. In either case, beliefs do reflect the content of experience.

Whatever the merits of these proposals as general theories of delusion, I propose an account of the Cotard delusion consistent with the idea that a metacognitive deficit is not essential to its genesis that also fits the slender evidence base about its neural substrate. Namely the Cotard person is *correctly* reporting an experience of inexistence. The self that has ceased to exist is the self-model. Other dimensions of self-representation (conceptual, agential, sensorimotor) may remain but the self-model that anchors cognition and provides its *telos* is gone.

Recall that, on the interoceptive active inference account, the brain learns that the self is the predictable target of regulation because the self-model subtends all activities from thermal regulation to explicit planning and inference. However, when the body cannot be controlled at the most basic level as a result of dyshomeostasis the self, understood as the target of regulation and anchor of cognition, disappears. *There is no self*. From the metacognitive point of view, she is *accurately* reporting the disappearance of a neurocognitive structure that is the substrate of the feeling of mineness.

It is important that the experience of the Cotard delusion is not equated directly with the sensory experience of, for example dyshomeostasis. Rather, it is the global nature of experience produced by *consequent destruction of the self-model*. In other words, the concept of the self-model explains why persistent and intractable dyshomeostasis leads to the loss of he the sense of mineness. Dyshomeostasis erodes the self-model from the bottom up.

The concept of a hierarchical self-model, however, suggests that pervasive destruction of the self-model introduced at any level could lead to the feeling of inexistence. In particular, it explains why imaging studies strongly implicate hypoactivity in the default mode network in the Cotard delusion. The best way to conceptualise the role of the default mode network is at the top level of the self-modelling hierarchy. It communicates with hubs of emotional processing and insular circuitry that construct lower levels of the self-model to underpin explicit autobiographical thought by "integrating information from multiple sources, including autobiographical memory and interoceptive processes, suggesting an active role of the default mode network in the neural construction of the self" (Ramirez Bermúdez et al., 2021, p. 68). Other theorists describe the DMN as the substrate of the narrative I (Davey & Harrison, 2018) Activity in the DMN ensures that the linguistic representation of the narrative "I" is anchored by episodes of personal experience.

Once again, these consideration are consistent with the idea that the Cotard delusion reflects the reality of experience produced by loss of the self-model rather than a metacognitive deficit. Charland-Verville's patient is one of many whose metacognitive faculties appear intact.

> He acknowledged that his abilities to see, hear, think, remember and communicate proved that his mind must be alive: he could not explain how his mind could be alive if his brain was dead, but he was certain that this was the case (Charland-Verville et al., 2013, p. 1997).

Such cases suggest that the appearance of a metacognitive deficit is produced by the surface, linguistic, features of the delusion. The inconsistency of delusional content with common sense metaphysics and propositional psychology need not reflect a problem with metacognition. Rather, it might tell us that the experience of mineness is the experience of a functioning self-model. When that is destroyed or degraded, partially or pervasively, from the bottom up or from the top down the experience of the self is also destroyed or degraded.

7 | CONCLUSION

It is worth recalling at this point the fundamental idea behind self-modelling account. Namely, the mind models the self as the entity whose boundaries are maintained and internal states and structures stabilised in the face of entropic forces. To do so, it recruits distributed processing resources across the mind, coordinating their activity to minimise free energy. In so doing, the mind embodies a model that predicts the consequence of its activities "as means to the end of embodied self-preservation" (Allen & Friston, 2018, p. 2481). On this way of thinking, the self is a *model* that enables the brain to steer the organism through the world by minimising prediction error.

That model allows the brain to regulate the body by proxy, acting to regulate interoceptive and affective experience. In depersonalisation experience, affective experience is inhibited, leaving most of the rest of the self-modelling hierarchy and its interactions with other processing hubs intact. This deaffectualisation makes it the case that, as Billon (2017) and Billon (Forthcoming) says, she no longer feels present in the world as an entity whose experiences signal its significance to her.

In the Cotard delusion bodily and affective levels of self-modelling that the narrative "I" have disintegrated. Jean, a patient with Cotard delusion studied by Young and Leafhead, described herself as "just a voice and if that goes I won't be anything ... if my voice goes I will be lost and I won't know where I have gone" (Young & Leafhead, 1996, p. 157). Jean's only possible response to this intractable sense of bodily/affective entropy is verbal: the linguistic shadow of a narrative "I" now disconnected from the structure of bodily and affective experiences that normally anchor it.

In the Cotard delusion, the subject uses the language of first-person report now unmoored from the control hierarchy disintegrating beneath it. The "I" of the Cotard delusion is like the airplane icon on the control panel of a plane whose electronics and hydraulics have melted down. It represents the plane, but attempts to use the icon to regulate the trajectory of the plane get no traction. The Cotard subject is, to adapt Descartes phrase, a disintegrating vessel without a pilot.

ACKNOWLEDGMENT

Open access publishing facilitated by The University of Adelaide, as part of the Wiley - The University of Adelaide agreement via the Council of Australian University Librarians.

ORCID

Philip Gerrans https://orcid.org/0000-0002-1755-8727

REFERENCES

Allen, M. (2020). Unravelling the neurobiology of interoceptive inference. *Trends in Cognitive Sciences*, *24*(4), 265–266.

Allen, M., & Friston, K. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, *195*(6), 2459–2482.

Baier, B., & Karnath, H. O. (2008). Tight link between our sense of limb ownership and self-awareness of actions. *Stroke*, *39*(2), 486–488.

Baker, D., Hunter, E., Lawrence, E., Medford, N., Patel, M., Senior, C., Sierra, M., ... David, A. S. (2003). Depersonalisation disorder: Clinical features of 204 cases. *The British Journal of Psychiatry*, *182*(5), 428–433.

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23.

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society*, *B371*(1708), 20160011.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429.

Betti, V., & Aglioti, S. M. (2016). Dynamic construction of the neural networks underpinning empathy for pain. *Neuroscience & Biobehavioral Reviews*, *63*, 191–206.

Billon, A. (2016). Making sense of the Cotard syndrome: Insights from the study of depersonalisation. *Mind & Language*, *31*(3), 356–391.

Billon, A. (2017). Mineness first: Three challenges to the recent theories of the sense of bodily ownership. In F. de Vignemont & A. J. Alsmith (Eds.), *The subject's matter: Self-consciousness and the body*. MIT Press.

Billon, A. (Forthcoming). What is it like to lack mineness? Depersonalisation as a probe for the scope, nature and role of mineness. In M. Garcia-Carpintero & M. Guillot (Eds.), *Self-experience: Essays on inner awareness*. Oxford University Press.

Brosch, T., & Sander, D. (2013). Comment: The appraising brain: Towards a neuro-cognitive model of appraisal processes in emotion. *Emotion Review*, *5*(2), 163–168.

Charland-Verville, V., Bruno, M. A., Bahri, M. A., Demertzi, A., Desseilles, M., Chatelle, C., Vanhaudenhuyse, A., Hustinx, R., Bernard, C., Tshibanda, L., Laureys, S., & Zeman, A. (2013). Brain dead yet mind alive: A positron emission tomography case study of brain metabolism in Cotard's syndrome. *Cortex*, *49*(7), 1997–1999.

Chatterjee, S. S., & Mitra, S. (2015). "I do not exist"—Cotard syndrome in insular cortex atrophy. *Biological Psychiatry*, *77*(11), e52–e53.

Ciaunica, A., Charlton, J., & Farmer, H. (2021). When the window cracks: Transparency and the fractured self in depersonalisation. *Phenomenology and the Cognitive Sciences*, *20*(1), 1–19.

Ciaunica, A., Roepstorff, A., Fotopoulou, A. K., & Petreca, B. (2021). Whatever next and close to my self—The transparent senses and the "second skin": Implications for the case of depersonalization. *Frontiers in Psychology*, *12*, 1219.

Clark, A. (2015). Embodied prediction. In T. K. GERger & J. M. Windt (Eds.), *Open MIND*. MIND Group.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A philosophical and psychological perspective on the development of thought. *Mind & Language*, *8*(4), 487–519.

Constant, A. (2021). The free energy principle: It's not about what it takes, it's about what took you there. *Biology and Philosophy*, *36*(2), 1–17.

Corcoran, A. W., & Hohwy, J. (2017). Allostasis, interoception, and the free energy principle: Feeling our way forward. In M. Tsakiris & H. de Preester (Eds.), *The interoceptive mind: From homeostasis to awareness*. Oxford University Press.

Corlett, P. R. (2019). Factor one, familiarity and frontal cortex: A challenge to the two-factor theory of delusions. *Cognitive Neuropsychiatry*, *24*(3), 165–177.

Corlett, P. R., & Fletcher, P. C. (2015). Delusions and prediction error: Clarifying the roles of behavioural and brain responses. *Cognitive Neuropsychiatry*, *20*(2), 95–105.

Cotard, J. (1882). Du délire des négations. *Archives de Neurologie*, *4*(152–170), 282–295.

Craig, A. (2009). Emotional moments across time: A possible neural basis for time perception in the anterior insula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1933–1942.

Danziger, N. (2006). Bases neurologiques de l'affect douloureux. *Revue Neurologique*, *162*(3), 395–399.

Davey, C. G., & Harrison, B. (2018). The brain's center of gravity: How the default mode network helps us to understand the self. *World Psychiatry*, *17*(3), 278–279.

Deane, G., Miller, M. D., & Wilkinson, S. (2020). Losing ourselves: Active inference, depersonalization and meditation. *Frontiers in Psychology*, *11*, 2893.

Debruyne, H., Portzky, M., Van den Eynde, F., & Audenaert, K. (2009). Cotard's syndrome: A review. *Current Psychiatry Reports*, *11*(3), 197–202.

Descartes, R. (1984/1641). Meditations on first philosophy. In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical writings of Descartes* (Vol. 2, pp. 3–62). Cambridge University Press.

Dieguez, S. (2018). Cotard syndrome. In *Neurologic-psychiatric syndromes in focus—Part II* (Vol. 42, pp. 23–34). Karger.

Feinstein, J. S., Khalsa, S. S., Salomons, T. V., Prkachin, K. M., Frey-Law, L. A., Lee, J. E., Tranel, D., & Rudrauf, D. (2016). Preserved emotional awareness of pain in a patient with extensive bilateral damage to the insula, anterior cingulate, and amygdala. *Brain Structure and Function*, *221*(3), 1499–1511.

Fernandez Velasco, P. & Loev, S. (2021). Affective experience in the predictive mind: a review and new integrative account. *Synthese*, *198*(11), 10847–10882.

Frijda, N. H., Kuipers, P., & Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–228.

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1211–1221.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214.

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience*, *7*, 598.

Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: New insights link internal physiology to social behaviour. Commentary on: "Anterior insular cortex mediates bodily sensibility and social anxiety" by Terasawa et al. (2012). *Social Cognitive and Affective Neuroscience*, *8*(3), 231–234.

Gerrans, P. (2001). Refining the explanation of Cotard's delusion. *Mind & Language*, *15*(1), 111–122.

Gerrans, P. (2014). *The measure of madness: Philosophy of mind, cognitive neuroscience, and delusional thought*. MIT Press.

Gerrans, P. (2020). Pain asymbolia as depersonalization for pain experience: An interoceptive active inference account. *Frontiers in Psychology*, *11*, 2643.

Gerrans, P. (2019). Depersonalization disorder, affective processing and predictive coding. *Review of Philosophy and Psychology*, *10*(2), 401–418.

Goldie, P. (2011). Life, fiction, and narrative. In N. Carroll & J. Gibson (Eds.), *Narrative, emotion, and insight*. Pennsylvania State University Press.

García-Carpintero, M., & Guillot, M. (Forthcoming). *Self-experience. Essays on inner awareness*. Oxford University Press.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hohwy, J., & Michael, J. (2017). Why should any body have a self? In F. de Vignemont & A. J. Alsmith (Eds.), *The subject's matter: Self-consciousness and the body*. MIT Press.

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(2), 1–33.

Jay, E. L., Sierra, M., Van den Eynde, F., Rothwell, J. C., & David, A. S. (2014). Testing a neurobiological model of depersonalization disorder using repetitive transcranial magnetic stimulation. *Brain Stimulation*, *7*(2), 252–259.

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLOS Computational Biology*, *9*(6), e1003094.

Kalisch, R. (2009). The functional neuroanatomy of reappraisal: Time matters. *Neuroscience and Biobehavioral Reviews*, *33*(8), 1215–1226.

Kalisch, R., Wiech, K., Critchley, H. D., & Dolan, R. J. (2006). Levels of appraisal: A medial prefrontal role in high-level appraisal of emotional material. *NeuroImage*, *30*(4), 1458–1466.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, *15*(138), 20170792.

Lee, M. C., Wanigasekera, V., & Tracey, I. (2014). Imaging opioid analgesia in the human brain and its potential relevance for understanding opioid use in chronic pain. *Neuropharmacology*, *84*, 123–130.

Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, *7*, 547.

Limanowski, J., & Friston, K. (2018). "Seeing the dark": Grounding phenomenal transparency and opacity in precision estimation for active inference. *Frontiers in Psychology*, *9*, 643.

Lyon, P., Keijzer, F., Arendt, D., & Levin, M. (2021). Basal cognition: Multicellularity, neurons and the cognitive lens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1821), 132.

Medford, N. (2012). Emotion and the unreal self: Depersonalization disorder and de-affectualization. *Emotion Review*, *4*(2), 139–144.

Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: Awareness and response. *Brain Structure and Function*, *214*(5–6), 535–549.

Medford, N., Sierra, M., Stringaris, A., Giampietro, V., Brammer, M. J., & David, A. S. (2016). Emotional experience and awareness of self: Functional MRI studies of depersonalization disorder. *Frontiers in Psychology*, *7*, 432.

Metzinger, T. (2011). The no-self alternative. In S. Gallagher (Ed.), *The Oxford handbook of the self* (pp. 279–296). Oxford University Press.

Michal, M., Reuchlein, B., Adler, J., Reiner, I., Beutel, M. E., Vögele, C., Schächinger, H., & Schulz, A. (2014). Striking discrepancy of anomalous body experiences with normal interoceptive accuracy in depersonalization-derealization disorder. *PLoS One*, *9*(2), e89823.

Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2016). Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience*, *10*, 56.

Moayedi, M. (2014). All roads lead to the insula. *Pain*, *155*(10), 1920–1921.

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35.

Philippi, C. L., Feinstein, J. S., Khalsa, S. S., Damasio, A., Tranel, D., Landini, G., Williford, D., & Rudrauf, D. (2012). Preserved self-awareness following extensive bilateral brain damage to the insula, anterior cingulate, and medial prefrontal cortices. *PLoS One*, *7*(8), e38413. https://doi.org/10.1371/journal.pone.0038413

Phillips, M. L., Medford, N., Senior, C., Bullmore, E. T., Suckling, J., Brammer, M. J., Andrew, C., Sierra, M., Williams, S. C., & David, A. S. (2001). Depersonalization disorder: Thinking without feeling. *Psychiatry Research: Neuroimaging*, *108*(3), 145–160.

Ramirez Bermúdez, J., Bustamante-Gomez, P., Espínola-Nadurille, M., Kerik, N. E., Dias Meneses, I. E., Restrepo-Martinez, M., & Mendez, M. F. (2021). Cotard syndrome in anti-NMDAR encephalitis: Two patients and insights from molecular imaging. *Neurocase*, *27*(1), 64–71.

Restrepo-Martínez, M., Espinola-Nadurille, M., Bayliss, L., Díaz-Meneses, I., Kerik, N. E., Mendez, M. F., & Ramírez Bermúdez, J. (2019). FDG-PET in Cotard syndrome before and after treatment: Can functional brain imaging support a two-factor hypothesis of nihilistic delusions? *Cognitive Neuropsychiatry*, *24*(6), 470–480.

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18*(4), 317–352.

Scherer, K. R. (2004). Feelings integrate the central representation of appraisal-driven response organization in emotion. In A. S. R. Manstead, N. Frijda, & A. Fischer (Eds.), *Feelings and emotions: The Amsterdam symposium*. Cambridge University Press.

Schilder, P. (1935). Psychoanalysis and philosophy. *Psychoanalytic Review*, *22*(3), 274–287.

Sel, A. (2014). Predictive codes of interoception, emotion, and the self. *Frontiers in Psychology*, *5*, 189.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573.

Seth, A. K., & Friston, K. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*, 395.

Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, *22*(11), 969–981.

Sierra, M. (2009). *Depersonalization: A new look at a neglected syndrome*. Cambridge University Press.

Sierra, M., & David, A. S. (2011). Depersonalization: A selective impairment of self-awareness. *Consciousness and Cognition*, *20*(1), 99–108.

Sierra, M., Medford, N., Wyatt, G., & David, A. S. (2012). Depersonalization disorder and anxiety: A special relationship? *Psychiatry Research*, *197*(1–2), 123–127.

Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., Tittgemeyer, M., Fleming, S. M., Haker, H., Seth, A. K., & Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, *10*, 550.

Terasawa, Y., Shibata, M., Moriguchi, Y., & Umeda, S. (2013). Anterior insular cortex mediates bodily sensibility and social anxiety. *Social Cognitive and Affective Neuroscience*, *8*(3), 259–266.

Tsakiris, M., & Critchley, H. (2016). Interoception beyond homeostasis: Affect, cognition and mental health. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160002.

Tsakiris, M., & De Preester, H. (2018). *The interoceptive mind: From homeostasis to awareness*. Oxford University Press.

Van de Cruys, S. (2017). *Affective value in the predictive mind*. MIND Group.

Weise, W. (2018). *Experienced wholeness: Integrating insights from Gestalt theory, cognitive neuroscience, and predictive processing*. MIT Press.

Young, A. W., & Leafhead, K. M. (1996). Betwixt life and death: Case studies of the Cotard delusion. In *Method in madness*. (*Case studies in cognitive neuropsychiatry*) (pp. 147–171). Psychology Press.

How to cite this article: Gerrans, P. (2023). A vessel without a pilot: Bodily and affective experience in the Cotard delusion of inexistence. *Mind & Language*, *38*(4), 1059–1080. https://doi.org/10.1111/mila.12441