

Spinozan Doxasticism About Delusions

Bongiorno, Federico

DOI:

[10.1111/papq.12373](https://doi.org/10.1111/papq.12373)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Bongiorno, F 2021, 'Spinozan Doxasticism About Delusions', *Pacific Philosophical Quarterly*.
<https://doi.org/10.1111/papq.12373>

[Link to publication on Research at Birmingham portal](#)

General rights

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

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

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

When citing, please reference the published version.

Take down policy

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

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

SPINOZAN DOXASTICISM ABOUT DELUSIONS

BY

FEDERICO BONGIORNO

Abstract: The Spinozan theory of belief fixation holds that mentally representing truth-apt propositions leads to immediately believing them. In this paper, I explore how the theory fares as a defence of doxasticism about delusions (the view that they are beliefs). Doxasticism has been criticised on the grounds that delusions typically do not abide by rational standards that we expect beliefs to conform to. If belief fixation is Spinozan, I argue, these deviations from rationality are not just compatible with, but supportive of, their status as beliefs.

1. Introduction

The notion that people can entertain a proposition without believing it has historically had and continues to have widespread intuitive appeal. It is intuitively natural to think that when presented with a proposition whose truth values are unknown, we have the ability to *merely* entertain it.¹ There are, however, some dissenters from this view. For some scholars (Gilbert, 1991; Gilbert *et al.*, 1993; Huebner, 2009; Mandelbaum, 2010, 2014; Levy and Mandelbaum, 2014; Mandelbaum and Quilty-Dunn, 2015), we automatically believe every truth-apt proposition that we entertain and only then perhaps take measures to revise our initial belief. I will call this view (following Gilbert, 1991) ‘The Spinozan Theory’, to be contrasted with ‘The Cartesian Theory’, the view that we first entertain a proposition, then subsequently believe it, disbelieve it or suspend judgment about it.²

According to the Spinozan theory, if one entertains the proposition ‘clouds are made of cotton candy’, one thus believes that ‘clouds are made

Pacific Philosophical Quarterly •• (2021) •••• DOI: 10.1111/papq.12373

© 2021 The Authors

Pacific Philosophical Quarterly published by University of Southern California and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

of cotton candy'. The theory has been defended at length first by Dan Gilbert (1991) and more recently by Eric Mandelbaum (2010, 2014). I should note at the outset that my aim here is not to argue that the theory is true, although I will briefly touch on some empirical evidence supporting it. Instead, my question is: assuming that the Spinozan theory is true, what consequences may there be for our understanding of delusions?

The term 'delusion' refers to a clinical symptom observed across a range of psychiatric disorders, including schizophrenia, dementia, schizoaffective disorder, bipolar disorder and major depression. Delusions can take various forms (Coltheart, 2013). Monothematic delusions concern a single topic. Much philosophical discussion has focused on *Capgras*, which presents as the conviction that a loved one has been replaced by an imposter, and *Cotard*, in which the person is convinced that she is dead, that she is missing internal organs or even that her body is rotting away. Many individuals with schizophrenia experience polythematic delusions, which are not restricted to a single topic but encompass a wide variety of subjects.

There has been much debate about whether delusions are beliefs. The position according to which delusions are beliefs is known as 'doxasticism about delusions' (Bayne and Pacherie, 2005; Bortolotti, 2010, 2012). The most common versions of anti-doxastic arguments rely on interpretationist assumptions about what is required for belief ascription. According to interpretationism, we ascribe beliefs in order to make sense of what other people do (Davidson, 1985; Dennett, 1987). Those who take this line posit a tight connection between being rational and being ascribed beliefs. At least three standards of rationality are taken to be essential to belief ascription (Bortolotti, 2010). Minimally, beliefs ought to be well supported by evidence (epistemic rationality), in line with other beliefs (procedural rationality) and action guiding in the appropriate circumstances (agential rationality). Suppose I tell you that Paula dislikes wine and my only evidence is that I once saw her drink a glass of punch. Suppose that later in the day I tell you that she is going to a wine-tasting class and that I have bought three bottles of wine for her dinner party tonight. These deviations from rationality raise questions about whether I really believe that Paula dislikes wine. Delusions are often like my claim about Paula: they are unsupported by evidence, they are badly integrated with other beliefs and they fail to manifest in action. This has led some philosophers to conclude that delusions are not beliefs but nondoxastic, or partially doxastic, states.³

In her defence of doxasticism, Lisa Bortolotti (2010, 2012) rejects the idea that rationality *per se* is individuating of beliefs. She justifies this by saying that if we deny belief status to delusions on grounds of irrationality, then we must conclude the same of many (apparent) beliefs that also do not live up to standards of rationality. On the positive side, she argues that the irrationality of delusions is no obstacle to their being classified as beliefs. Call this the 'standard' defence of doxasticism. Bortolotti is only concerned with

the notion of belief that underlies our ordinary mentalistic ascriptions, and she makes no commitment to what beliefs are outside the folk-psychological discourse. It may be wondered, however, whether the doxastic status of delusions can be genuinely secured within the explanatory framework of folk psychology. One might want more from a defence of doxasticism. One might, for example, want an analysis that helps explain why delusions, qua beliefs, fail to satisfy norms of rationality. This would require, at a minimum, a view of cognitive architecture that settles how an agent fixates her beliefs. If it could be shown that the failures of rationality found in delusions result naturally from belief-fixating processes, then the argument for doxasticism would be stronger.

The paper unfolds as follows. I first outline the Spinozan theory in more detail, with particular attention to the way it characterises belief. I then consider the implications of applying this theory to delusions. This generates a new version of doxasticism, which has two advantages over the standard defence: (i) it puts pressure on the very notion that one can be deluded that p without believing that p , and (ii) it accommodates the deviations from rationality attributed to delusion within a plausible theory of belief fixation. Finally, I address whether delusions really fit the Spinozan characterisation of beliefs, and why this is something all doxasticists should want to accept.⁴

2. *I think, therefore I believe*

2.1. CARTESIAN VERSUS SPINOZAN CONCEPTION OF BELIEF

In our ordinary thinking about ourselves, it seems obvious to us that we are able to evaluate propositions as candidates for belief, before believing them. This simple notion is the centrepiece of the Cartesian view and has been standard in both philosophy and cognitive science (Quine, 1960; Fodor, 1983; Dennett, 1987; Pylyshyn, 1989). The Cartesian view, when fleshed out more fully, is defined by three following claims (Mandelbaum, 2014). First, agents are able to assess the truth value of a proposition they entertain before believing it. In this sense, belief fixation consists in a sequential process with multiple stages, one involving the entertaining of the proposition and the other involving going ahead and believing or disbelieving it. Second, believing and rejecting a proposition are alternative outcomes of a single mental process (Gilbert, 1991) and thus should show the same pattern of effects in situations where cognitive resources are depleted. Third, both believing and rejecting a proposition involve effortful mental activity.⁵ That is, in order for a proposition to be believed or rejected, a person needs to have sufficient cognitive capacity to evaluate that proposition.

The Spinozan takes an alternative view, arguing for the temporal precedence of belief fixation over rejection, and the unity of entertaining some

proposition and believing it (Gilbert, 1991). This view can be stated in three broad claims as follows (Mandelbaum, 2014). First, agents are unable to entertain propositions without believing them. In other words, they have no choice but to believe whatever proposition they happen to entertain. Second, believing a proposition and rejecting it as false are functionally different cognitive operations, which suggests that they should be affected differently when people's cognitive resources are limited or depleted. Third, forming a belief is an automatic and effortless process. By contrast, replacing the attitude of belief with that of disbelief is an effortful mental task, which can only be carried out once the belief has been acquired. This is why belief fixation can occur effortlessly under cognitive load, while the task of rejecting previously held beliefs is made more difficult by increased cognitive load.

2.2. MOTIVATION FOR THE SPINOZAN THEORY

The Cartesian and Spinozan theories make different predictions about the ways in which the belief-fixating process should break down in the wake of resource depletion (Gilbert, 1991; Mandelbaum, 2014). The Cartesian theory predicts that depleting cognitive resources should disable the system from deciding about the truth of the propositions it entertains. Because cognitive load depletes cognitive resources, propositions should neither be believed nor rejected when individuals are placed under cognitive load. This is because believing and rejecting a proposition are products of one and the same mental process and so should be affected by cognitive load to an equal extent (Mandelbaum, 2014).

Conversely, if the Spinozan theory were correct, resource depletion should prevent the system from disbelieving a proposition, but not from believing a proposition. The Spinozan theory treats believing as a reflex. Because cognitive load interferes only with effortful and cognitively taxing processes, if believing truly is reflexive and undemanding of cognitive resources, it should remain unaffected by concurrent load. However, because disbelieving is particularly onerous in terms of cognitive resources, the prediction is that it should be shut off by concurrent cognitive load (Mandelbaum, 2014).

Gilbert and colleagues marshalled empirical evidence that they claim is best accommodated by a Spinozan theory of belief fixation (see Mandelbaum, 2014, for a comprehensive review and discussion). Such evidence comes from experiments in which subjects under cognitive load are presented with statements tagged as true or false and observed to treat all statements as true. In Gilbert *et al.* (1990), for example, participants were shown a sequence of fictitious Hopi word definitions, such as 'a twyrin is a doctor', shortly followed by a feedback that the preceding statement was either true or false. On some trials, feedback encoding was interrupted by a tone-detection task, which was meant to deplete participants' cognitive resources. The outcome was in line with the Spinozan theory: depleting

cognitive resources selectively impaired participants' ability to remember false statements as false, while leaving intact their ability to remember true statements as true (Gilbert *et al.*, 1990, pp. 603–604).

Additional evidence from a variety of experimental settings appears to support the Spinozan theory. Studies on persuasion through fiction suggested that people change their real-world beliefs upon engaging with fictional stories (Green and Brock, 2002; Marsh and Fazio, 2006; Appel and Richter, 2007). Other studies found that repeated warnings about false consumer claims foster the remembering of those claims as true (Skurnik *et al.*, 2005). Still other studies have shown that repetition increases statements' perceived truthfulness, even when those statements contradict well-known facts (Fazio *et al.*, 2019). I note, however, that the status of the evidence for the Spinozan theory has not gone uncontested. Nadarevic and Erdfelder (2019), for example, performed experiments similar to Gilbert and colleagues' and got very different results (see also Street and Kingstone, 2016, but see Asp *et al.*, n.d., unpublished manuscript, for a counter-defence).⁶

To repeat, my goal is not to argue for the truth of the Spinozan theory. My goal is more modest and narrow: to evaluate its capacity to improve the prospects of doxasticism about delusions. The evidence that I briefly surveyed here does not conclusively support the Spinozan theory, but it makes the hypothesis plausible enough that it seems worthwhile to begin considering its implications for whether delusions are beliefs. Before we get to that, however, we need to take stock of the theory's understanding of belief.

3. *Spinozan beliefs*

3.1. ARATIONAL BELIEF FORMATION

The main thrust of the Spinozan view is that belief fixation operates below the psychological level of explanation, in a way that resists being cast as rational or irrational. Levy and Mandelbaum (2014) make the point through the following example. Imagine a brick falls on your head, causing you to believe that there are only ten planets in the universe. Now of course, the process that led to the belief would be less than epistemically ideal. However, it would be just as strange and misleading to say that your being moved to belief resulted from an irrational inference (or in fact any inference at all). After all, the belief was acquired by means of brute-causal, triggering processes, and not as a result, say, of guessing or wishful thinking. Levy and Mandelbaum (2014) propose to call these types of belief-forming processes 'arational' – neither rational nor irrational – because there are no facts about the agent's prior mental states from which to assess their (ir)rationality (Wedgwood, 2017). Just as it is arational to believe that the universe has

ten planets as a result of being hit by a brick, so too it is arational to believe any arbitrary proposition we happen to entertain. In both cases, beliefs are acquired in a brute causal way, such that there is no room even for reasoning to go awry.

3.2. INCONSISTENCY

The Spinozan theory is committed to the view that people's beliefs can be and often are inconsistent.⁷ Inconsistent propositions are pervasive in our daily lives, and because we are most of the time under cognitive load, we often lack the capacity to reject all such propositions after believing them (Mandelbaum, 2014). The view that people believe inconsistencies needs some qualifications and some filling in of details. First, the claim does not imply that people who believe inconsistencies will *assert* those inconsistencies (Mandelbaum, 2014, p. 64, fn. 24). Asserting that p entails overtly showing one's commitment to p and, on this picture, beliefs that p carry no such commitment. One can believe p and believe $\neg p$ without consciously endorsing p and $\neg p$ (for more on this, see next subsection). Second, what explains the existence of inconsistent beliefs is the fact our total set of beliefs (henceforth 'belief storage') is 'fragmented' into separate belief systems (Bendana and Mandelbaum, forthcoming). Belief storage is often described by the metaphor of a *web of belief* (Quine and Ullian, 1970). According to this metaphor, all of our beliefs are arranged into a single web-like network, which (i) is consistent and closed under entailment and which (ii) guides action in all contexts. If our beliefs are connected in a web-like way, altering one belief will have consequences for the entire belief set, such that when an agent changes any of her beliefs, all other beliefs will adjust in the light of that. Some have argued, however, that actual human cognition falls short of this idealisation. Rather than being stored in a single coherent database, our beliefs are stored into multiple clusters or fragments that may be inconsistent with one another (Lewis, 1982; Stalnaker, 1984; Egan, 2008b). This means that an agent may believe that p in one fragment (f_1) and that $\neg p$ in a separate fragment (f_2) without being prompted to eliminate the conflict between those beliefs. Indeed, if an agent holds inconsistent beliefs in different fragments, it is possible that p may be true according to f_1 , but false according to f_2 . Adjustment would be required only if a belief that p and a belief that $\neg p$ were stored in the same fragment, or if they were stored in separate fragments activated at the same time.

A related point is that separate and perhaps inconsistent fragments may be activated in different contexts to produce different patterns of behaviour. This can have various consequences. Agents may believe that p and that *if p then q* , but fail to act appropriately for q , because the beliefs that p and that *if p then q* are in separate fragments (Egan, 2008b). It can also happen that certain actions or aspects of behaviour are guided by a fragment that

includes p , while others are guided by a fragment that includes $\neg p$. For some Spinozans, this is reflected in cases where there is a trade-off between implicit racism and the antiracist attitudes that one explicitly reports. Consider someone (e.g. a white, liberal law professor) who sincerely professes the equality of all races but whose low-level, non-verbal behaviours are consistently racist (e.g. when a black person gets on the bus, she looks down to avoid eye contact). Assuming fragmentation, one could suggest that this person's racist and antiracist beliefs reside in separate fragments and that while the non-racist fragment is accessed for overt behaviours (i.e. speech), the racist belief fragment is accessed for spontaneous, unguarded reactions.⁸

3.3. OBJECTIONS AND FURTHER CLARIFICATIONS

There are at least two preliminary objections that could disqualify the Spinozan theory from the very outset (Mandelbaum, 2014). Before concluding this subsection, it is worth considering how the Spinozan might respond to these objections, as doing so will help clarify matters further.

The first objection is based on intuition: the intuition is that we can know introspectively what our current attitude is towards a proposition, such as whether we believe it or not. In general, we tend to think of belief as something that is accessible through introspection. If I ask you whether you believe that clouds are made of cotton candy, you would be confident that you do not, and that is because you think you know what you believe. If your intuition is right, then the Spinozan theory is false, for the Spinozan theory excludes that you can entertain a proposition without believing it.

A possible response to this line of thought is to question whether beliefs are introspectively accessible across the board. According to Mandelbaum (2014), the intuition that they are is a kind of cognitive illusion stemming from the fluency with which information is processed in cases of commonplace or affectively toned belief. In such cases, the belief has been so frequently activated that we need not do any inferential work. If I ask you whether the sky is blue, or whether you love your parents, you can answer immediately because you have close familiarity with the questions. However, when we turn to beliefs that are not so familiar or strongly value-laden, things look more problematic. If I ask you whether you are prone to altitude sickness, you may answer immediately (perhaps you are a mountaineer), but more likely you will infer what you believe based on recollections of past experiences and behaviours.⁹

Mandelbaum (2014) makes yet another important point, namely, that we tend to overstate the extent to which we can introspect beliefs because we too narrowly think of them as mental contents rather than functional states. Beliefs include contents as parts, but they are over and above their contents. They play a functional role in our mental economy. In this role, beliefs operate on contents, but they are not contents themselves (Mandelbaum, 2014,

p. 76). While belief contents are sometimes available to introspection and sometimes not (think, for instance, of the contents of racists beliefs that cause one to avoid eye contact with black people), functional roles (such as a disposition to guide X kinds of behaviours) are not introspectable. The bottom line is that we do not have introspective access to our beliefs (Mandelbaum, 2014, p. 77).

But what is it to introspect the content of beliefs? Mandelbaum's answer is that we introspect a judgment that *p*, from which we infer that we believe that *p*. So, the 'beliefs' that we report having are in fact judgments about beliefs. There are two possible scenarios where judgments become accessible. In the first, we have access to a judgment that accords with the belief content (i.e. we judge that *p* because we introspect the content that *p*). From this, we can rightly infer that we believe *p*. In the second, we form a judgment that *p* as a result of deliberation, whereby we examine our past behaviours, simulate what others would do in our situation, register what we have reason to believe and so forth. The point is that we go through some reasoning process and come up with a judgment about what we believe (Mandelbaum, 2014).

Either scenario rests on a distinction between judgments or belief reports and beliefs *per se*. The former are personal-level states of a conscious subject. The latter are unconscious propositional attitudes. For Mandelbaum (2014), we only report beliefs that we consciously *endorse*, these being the ones which we count as justifiable and socially acceptable. What we endorse is determined by host of practical (i.e. not truth-relevant) factors, including avoidance of blame, social anxiety and shame-avoiding mechanisms. This is why the beliefs that we report may not often match what we actually believe (e.g. it is why the implicit racist only reports her egalitarian beliefs).

A second objection to the Spinozan theory is to ask why we should accept that the states that the Spinozan calls 'beliefs' are in fact *beliefs* (Mandelbaum, 2014, p. 82). These states lack some of the properties that beliefs are standardly assumed to have (e.g. they are acquired arationally, they are opaque to introspection and we can harbour inconsistent ones). Mandelbaum's response on behalf of the Spinozan theory comes in two steps: the first step is to articulate a complex of minimal necessary conditions for belief; the second step is to show that the states under discussion meet these conditions.¹⁰

A mental state, *M*, qualifies as a belief when at least the following conditions are satisfied: (i) *M* is semantically evaluable (true/false), (ii) *M* is capable of being acquired through perception, (iii) *M* is capable of interacting with desires in such a way as to cause behaviour and (iv) *M* is 'inferentially promiscuous' (viz. available to be used as a premise in reasoning; see Stich, 1978).

It should be clear that the states discussed in the aforementioned experiments are consistent with both (i) and (ii); statements like 'a twyrin is a

doctor' are perceptually learned and have truth conditions. To show that these states also meet (iii) and (iv), Mandelbaum cites a further experiment designed to test the Spinozan theory. In this experiment, Gilbert *et al.* (1993) asked participants to read two crime reports containing a mix of true and false information (printed in black and red, respectively) and then to play the role of judges. Half of participants had cognitive load of performing a concurrent digit-search task while reading the reports (the interrupted condition). The results showed that the prison terms recommended by uninterrupted participants were only minimally affected by false information, while false information significantly affected the prison terms recommended by participants in the interrupted condition. Interrupted participants recommended a sentence twice as long when the false information was aggravating rather than extenuating for the perpetrator. This demonstrates that the attitudes based on such information featured as premises in the reasoning that informed their sentencing behaviour, in line with (iii) and (iv).

4. *Spinozan doxasticism: the basics*

I have now sketched the fundamentals of the Spinozan theory of cognitive architecture. My guiding question in what follows will not be 'is the Spinozan architecture accurate?' but rather 'what would the consequences of it be for the question of whether delusions are beliefs?' The primary focus will be on how a Spinoza-inspired doxasticism (hereafter simply Spinozan doxasticism) can explain the failures of rationality characteristic of delusions more fully than standard doxasticism, as defended by, for example, Bortolotti (2010, 2012).

If we accept the Spinozan framework, then a simple modus ponens is all we need to conclude that delusions are beliefs.

- 1 If one entertains the proposition p , then one believes p .
- 2 Being deluded that p involves entertaining p .

Being deluded that p involves believing p .

This is not as trivial as it may look initially, for as we will see, it can integrate key features of delusion in an explanatory fashion. But let us take things one step at a time.

Consider the Capgras delusion, in which people report that an imposter has replaced a spouse or relative. From the point of view of the Spinozan theory, the proposition a person asserts in uttering that is not my wife' is automatically believed, simply by being entertained. It does not matter what

propositional attitude the person has. All that matters is that having a propositional attitude (or reporting about that attitude) involves entertaining a proposition, and entertaining a proposition (as per the Spinozan view) causes belief.¹¹

Critically, the vast majority of philosophers who embrace non-doxasticism assume that the cognitions involved in delusions are propositional attitudes (Matthews, 2013). That is to say, delusional reports are expressions of mental states consisting of a representation of a proposition (e.g. ‘my wife is an imposter’) and an attitude towards that proposition (‘I *imagine* that’, ‘I *accept* that’ etc.). On some more parsimonious versions of non-doxasticism, delusions are identified with standard everyday propositional attitudes (ones that we already countenance) such as imaginings, or perhaps combinations of these, such as *imagining that p* and *believing that one believes that p* (Currie, 2000; Currie and Jureidini, 2001; Currie and Ravenscroft, 2002). On other versions, delusions are identified with hitherto undiscovered propositional attitudes, for example, mental states intermediate between imagination and belief, what Andy Egan (2008a) calls *bimaginings*. Regardless of which type of non-doxasticism one embraces, the Spinozan architecture would render doxasticism inescapable, because whatever attitude one has towards a proposition, it would result in the proposition being believed. The more interesting considerations, to which I now turn, concern how the Spinozan architecture can accommodate those features of delusion that are seen as most indicative of their rationality failures and thus most problematic for doxasticism: unresponsiveness to evidence, circumscription and double-bookkeeping (Dub, 2017).

5. Paradigmatic features of delusions

5.1. UNRESPONSIVENESS TO EVIDENCE

There are two senses in which delusions are said to be unresponsive to evidence: they are formed on the basis of insufficient evidence, and they are maintained in the face of counterevidence. It is not clear that *all* delusions are equally unresponsive to evidence in the ways the definition suggests.¹² What is certain, however, is that *most* are. Consider a classic case of delusional jealousy, in which a man claims that his wife is unfaithful to him because the fifth lamp-post along on the left is unlit (Sims, 2003, p. 119). The ‘evidence’ the man adduces to prove that he is being cheated on is not merely insufficient to support his claim, but altogether irrelevant. There is clearly no intelligible link between his wife’s infidelity and the lamp-post being unlit, suggesting that the man’s delusion lacks evidential support. The clinical literature also abounds with examples of delusions that are resistant to change, even when counterevidence becomes available. For

instance, in one striking case of anosognosia for hemiplegia (i.e. the inability to recognise impairments resulting from brain damage), one patient reports that she can clap despite the fact that her left hand is visibly motionless and no sound is audible (Berti *et al.*, 1998, pp. 29–30).

5.2. CIRCUMSCRIPTION

Another feature of delusions that has been brought to bear against their status as beliefs is the circumscribed role they play in one's cognitive economy. Circumscription is especially evident in monothematic delusions like Capgras. Delusions are circumscribed in three respects, the first of which is *inferential* (Campbell, 2001; Hamilton, 2007). This is not to say that people never make inferences from their delusions. On the contrary, if you ask a person with Capgras how and why loved ones are not who they appear to be, she will often elaborate on the content of her delusional state. Such elaborations, however, are typically disconnected from the rest of the person's worldview (Stone and Young, 2007), suggesting that the delusion must play a weak inferential role in the person's reasoning. Monothematic delusions are generally only elaborated when solicited with why-questions, but many other delusions that occur in schizophrenia (so-called 'polythematic') are elaborated spontaneously in intricate ways. It has been noted that even much elaborated delusions may exhibit some degree of circumscription (Sass, 1994), as evidenced by the phenomenon of double-bookkeeping (see next subsection).

The second respect in which delusions are circumscribed is *behavioural*, where this means that they tend to be causally inert in respect of behaviour (Currie, 2000; Young, 2000; Currie and Ravenscroft, 2002; Egan, 2008a; Bortolotti, 2012). Subjects with delusions often do not act on them and display inconsistent verbal and non-verbal behaviours. Suppose I tell you that an imposter who looks just like my mother has taken her place. If I truly believe that my mother has been replaced, then you would expect me to act in ways that are consistent with and can be explained by my belief. For instance, you might expect me to file a missing person report, run away from the imposter, take some steps to locate my mother and try to figure out how the switch occurred. However, people who experience the Capgras delusion may fail to do any of these things, with some even carrying on friendly interactions with the imposter (Lucchelli and Spinnler, 2007). Similarly, paranoid patients who accuse the nursing staff of trying to poison them may eat their meals without complaint and readily take oral medications as prescribed (Sass, 1992, p. 274).

The third respect in which delusions are circumscribed is *affective*, which is to say that they are often characterised by a lack of appropriate accompanying affect (e.g. the kind of emotional comportment we would expect from someone who genuinely believe the things they profess to believe; see

Sass, 1994; Stone and Young, 1997). Using again the example of Capgras delusion, we can suppose that if your mother went missing, you would probably be concerned about her whereabouts. If you believed a person to be an imposter, you would most likely respond with rage and angry outbursts. However, most Capgras subjects show little concern about the welfare of their loved one and seem not much troubled by the presence of the imposter (Alexander *et al.*, 1979).¹³

5.3. DOUBLE-BOOKKEEPING

The term ‘double-bookkeeping’ refers to the fact that people who sincerely avow their delusions operate on two parallel but separate clusters of representations and retain a roughly accurate sense of which is which (Sass, 1994). On the one hand, there are the true beliefs that they hold about objective reality. On the other, there are the delusional ‘beliefs’ they profess about some inner subjective reality. Double-bookkeeping applies equally well to psychotic breaks and delusional episodes in schizophrenia, some of which are elaborated and polythematic, as to monothematic delusions. An individual with schizophrenia might have a complex delusional system involving undercover FBI agents conspiring to kill her, yet at the same time take long walks at night by herself. Similarly, as we have seen, individuals with Capgras might confidently assert that imposters are substituting for their wife or husband, and yet fail to even search for their displaced spouses.

6. *Why Spinozan doxasticism?*

Many have found that the aforementioned features of delusions cast doubt on doxasticism. The focus is on the concept of belief as used in folk psychology, the everyday practice of ascribing mental states to self and others as a way of predicting and explaining behaviour. Anti-doxastic arguments tend to move from an ‘interpretationist’ understanding of folk psychology, according to which when we ascribe mental attitudes to others, we aim to explain their behaviour given those mental attitudes. For instance, if you see me take an umbrella and a raincoat when I leave home in the morning, you can legitimately ascribe to me the belief that it is raining. A key idea behind interpretationism is that we must assume that a system is rational if we wish to explain the behaviour of that system via attitude ascription (e.g. Davidson, 1982, 1985; Dennett, 1987). One implication of this is that we can ascribe to a system only rational beliefs. Delusions, goes the anti-doxasticist, fail to be rational in several respects: they are often unsupported by reasons, inconsistent and partially otiose. Therefore, it is concluded, delusions are not beliefs.

We have seen that the intuitive strength of anti-doxastic arguments stems from the interpretationist assumption that one needs to be rational in order to be ascribed beliefs. The standard defence argument, developed at book-length by Bortolotti (2010), is to reject the rationality constraint on belief ascription. Bortolotti (2010) is concerned to adjudicate between ‘the good and the bad of interpretationism’ (p. 261). She thinks it is good that we understand belief in terms of our belief-ascribing practices. This means there is a threshold of interpretability that marks the minimum for a mental state to qualify as belief. But she also thinks it is bad that we idealise belief ascription by identifying such a threshold with a general backdrop of rationality (Bortolotti, 2010, p. 261).

As such, her strategy is to offer for each delusion a belief with a comparable type if not degree of irrationality. If it can be shown that the attitude in question is one that we are happy to ascribe as a belief to others, then rationality should not be invoked as a constraint on ascription. This has two implications for the status of delusions as beliefs. First, if we refuse to impose rationality constraints on belief ascription, then we cannot legitimately deny belief status to delusions on the ground of irrationality. Second, if the belief in question is irrational in a way comparable with delusions, then there would be reason to conclude that delusions are beliefs. In support of this argument, Bortolotti (2010) cites numerous examples in which we, as interpreters, are willing to ascribe irrational beliefs. Many people express self-deceptive beliefs (e.g. beliefs motivated by desires and emotions) that can be very resistant to critical engagement. It is also common for people to maintain superstitious beliefs (e.g. ones involving supernatural causation) that are inconsistent with the other beliefs they have (e.g. their commitment to a scientific world view). Again, people often fail to act on their self-reported beliefs. For instance, they might express the belief that sex without a condom can lead to HIV infection, but admit to not using condoms (Aronson, 1999). These cases violate rational constraints no less than delusions do: they are unresponsive to evidence, badly integrated with one’s other beliefs, weakly behaviour guiding. And yet, we are comfortable ascribing them as beliefs. If this is true, the story goes, we can interpret delusions as beliefs too.

The standard defence thus makes a convincing case for the conclusion that delusions can be ascribed as beliefs on the grounds that they conform to folk-psychological standards about what counts as ‘belief’ (see Rose *et al.*, 2014, for evidence that folk do in fact construe subjects as believing their delusions). Still, the question can be asked whether doxasticism can be fully defended on folk-psychological grounds, with no consideration of what, if any, functions beliefs are poised to perform within the mind. After all, it may be thought that a full-scale defence of doxasticism should enable us to see why delusions, in virtue of being beliefs, deviate from rationality in the way they do. We seem to give up that explanatory component of

doxasticism if we focus on just the standards of folk belief ascription. For doxasticists to avoid this outcome, delusions' status as beliefs would need to be empirically accommodated from within a viable model of cognitive architecture. I will argue that the Spinozan theory provides such a model and in so doing has the explanatory component won: if belief fixation is Spinozan, unresponsiveness to evidence, circumscription and double-bookkeeping are natural – or at least predictable – results.

7. *Delusions as Spinozan beliefs*

7.1. EVIDENCELESS/RESISTANT BELIEFS

Spinozan belief fixation is arational and therefore, by definition, unresponsive to evidence. Because of our cognitive architecture, we are set up with dispositions to immediately believe any propositions we happen to token, without weighing evidence first.¹⁴ We acquire beliefs in a brute causal way, via processes that work below the psychological level and that, as such, are impossible to counteract psychologically (Levy and Mandelbaum, 2014).

In principle, we are free to reject our newly acquired beliefs. The Spinozan theory indicates that disbelief and suspending judgment are possible, albeit only as modifications of an initially untested belief. However, we have seen that our capacity to reject false beliefs can be overridden by an increase in cognitive load (i.e. extra mental processing imposed on the cognitive system). There are several factors that can increase load, most of which are woven into normal daily activities. Two notable examples are mind-wandering and distraction. In everyday life, we are exposed to a vast amount of information, while our focus of attention is continuously switching between external stimuli and internal thoughts. Merely dividing attention between features presented in the same or different modalities is a primary source of cognitive load. In the study by Gilbert *et al.* (1993) discussed earlier, for instance, subjects in the interrupted condition were told to read the text of the crime reports and concurrently search a string of digits for number 5. The aim of this experimental design was to create a condition of split attention between two visual cues. The result was that the prison terms recommended by interrupted subjects were significantly affected by propositions they knew to be false. Beliefs were not simply acquired in an 'evidence-less fashion' but also retained in the face of disconfirming evidence (Mandelbaum and Quilty-Dunn, 2015).

Another factor that has been used in different studies to induce cognitive load is time pressure. Interestingly enough, it has been suggested that increased time pressure can make it difficult to suppress stereotype-based beliefs, even in circumstances where we most want to inhibit them

(Huebner, 2009). For example, one study found that subjects who were admonished not to be sexist in completing sentences with a missing word (e.g. 'Women who go out with a lot of men are ...') were more likely to make sexist completions when load was induced by asking for immediate responses (Wegner *et al.*, 1993).

In sum, resistance to belief revision is to be expected in situations where an increase in cognitive load due, for example, to heightened attention or time pressure depletes cognitive resources that are needed to override false beliefs.

There is suggestive evidence that one such situation is encountered, in a particularly acute and exacerbated fashion, among delusional and delusion-prone individuals. Several studies have focused on content-specific attentional biases in delusions. The most commonly used paradigm to study attentional bias is the emotional Stroop test (e.g. Gotlib and McCann, 1984), in which participants are instructed to name as quickly as possible the colour in which affect-laden and neutral words are printed, while ignoring their semantic meaning. If subjects' performance on affect-laden words is observed to be markedly slower relative to neutral words, attentional bias is inferred; this is because delays in colour naming are taken as indicative of greater attentional capture by the meaning of words. Bentall and Kaney (1989) studied the Stroop performance of people with persecutory delusions and found that their subjects responded most slowly to threat-related words (e.g. 'kill', 'spy' and 'pain'), an effect that was not observed for psychiatric (depressed) and normal controls. This finding has been replicated subsequently (Fear *et al.*, 1996). Attentional bias for delusion-specific information was further demonstrated in the case of a woman, JK, who was convinced that she had died (Cotard delusion) and that members of her family were not who they seemed (Capgras delusion). JK showed disproportionately longer colour-naming times for tests lists containing death-related and duplicate-related words relative to sets of neutral words. After her delusions had disappeared, however, her colour-naming times did not vary across test and neutral lists (Leafhead *et al.*, 1996).

Abnormalities in selective attention – particularly in attentional inhibition – have also figured importantly in schizophrenia research. One of the most consistently observed features in early schizophrenia (i.e. prior to the onset of frank psychotic symptoms like delusions) is the inability to screen out irrelevant stimuli from the environment, resulting in sensory overload and increased distractibility. Patients say such things as 'I can't shut things out', 'Everything seems to go through me' and 'I am attending to everything at once and as a result I do not really attend to anything' (McGhie and Chapman, 1961, p. 104). Failure to inhibit distractors has been explained in terms of aberrant assignment of 'salience' to neutral stimuli, driven by an excess of dopamine release outside the proper context (Kapur, 2003, 2004). The effect of this is to make events attention-grabbing that would

otherwise be inconspicuous, creating a sense of uncertainty and unpredictability, as well as a sense of urgency and time pressure to resolve it.

Experiences of aberrant salience, so understood, are involved in the formation of some, but not all, delusions. For example, when salience is misattributed to unrelated or coincidental phenomena (e.g. a causal conversation between a couple of window-shoppers), these phenomena may be interpreted as pertaining specifically to oneself, giving rise to delusions of reference and persecution. Other delusions, however, are accompanied by highly specific experiences, which cannot simply be captured in terms of aberrant salience. To be sure, these experiences may well be said to be phenomenally salient, if this means that the subject feels her attention being drawn to their contents. But even so, they are not *just* experiences of salience – they are not just about imbuing meaning to any stimulus or event that is currently occurring.

For example, people with Capgras delusion have an abnormally reduced affective responsiveness to familiar faces (Ellis *et al.*, 1997; Hirstein and Ramachandran, 1997; Brighetti *et al.*, 2007), with the result that persons whose faces are familiar feel unfamiliar. In another misidentification delusion, the Fregoli delusion (the belief that strangers are in fact familiar persons in disguise), people have an abnormally heightened responsiveness to unknown faces and thus feel as if strangers are familiar to them (Langdon *et al.*, 2014). In Cotard, there is a general flattening of affective responsiveness to all perceptual inputs, which could lead to the delusion of being dead (Young and Leafhead, 1996). For our purposes here, it does not matter how exactly delusions arise from such unusual experiences. The relevant question for us is whether such unusual experiences expend many cognitive resources. And the answer is certainly yes, not only because they are surprising, and thus in need of explanation, but also because they are often persistent and distressing for the person who undergoes them.

The take-away point is that delusional fixity – the fact that delusions are retained in the face of counterevidence – is predicted by the Spinozan architecture. If delusions are formed and maintained under conditions of mental load and time urgency, as seems the case, then we should expect these factors to interfere with the effortful processing needed for belief revision. This is because, for the Spinozan, effortful processing is resource dependent and so can be disrupted when people operate under cognitive load or with scarce cognitive resources available.

7.2. QUESTIONS AND TENTATIVE ANSWERS

Taken together, the aforementioned considerations also help us address two questions that are potentially worrisome for Spinozan doxasticism (see footnote 10). One is the question: (1) why does not everyone who entertains delusional propositions (e.g. people who live or are in close contact with the

deluded individuals) end up being believing them? For instance, why does not the doctor who hears her patient say 'I am dead' believe that she has actually died? The second question is: (2) why cannot someone stop believing a delusional proposition by simply entertaining its negation (for instance, as a result of hearing friends and family members say, 'You are alive and well')?

The answer to (1) should be clear by now. Not everyone who entertains a delusional proposition is doing the same amount of cognitive work, and arguably, delusional and delusion-prone subjects (e.g. first-episode patients with schizophrenia) have a much higher cognitive load than any healthy individual around them. Normally, when we consider outlandish propositions, we can immediately reject the correspondingly acquired beliefs. And that is because we have enough cognitive resources to allow for the rejection process to get off the ground.

But this is not always the case, as we have seen. Think, for example, of the attention biases towards delusion-salient stimuli shown by some delusional subjects. Not only are these subjects overwhelmed by attentional demands, and thus less apt to reject their initial beliefs, but they preferentially encode material that serves to maintain them (Bentall and Kaney, 1989). Or again, consider what happens in first-episode schizophrenia, where patients are unable to leave out irrelevant stimuli to the point where they feel flooded by information. These patients will be so distracted while entertaining a delusional proposition that they may utterly lack the requisite cognitive energy to reject the newly acquired belief.

Now let us complicate things a bit. Suppose that friends and family members are under cognitive load of a comparably high level to that undergone by the deluded individual. Does this mean that they will too believe the delusion upon entertaining the delusional proposition p ? This is a difficult question, and one which I cannot fully address. Still, I make some tentative remarks. One possibility is the following. The person in question gets a significant credence boost from entertaining p and yet the resulting credence remains somewhat weaker than an outright belief. Perhaps this person fails to believe p outright because she has the simultaneous belief that the deluded individual is not reliable, or because, absent experiential abnormalities, background beliefs still have some traction. Another possibility is that the person does in fact form the outright belief that p . This may happen as a consequence of the overall psychological situation the person is in. For instance, if the person has a paranoid predisposition, and the theme of the asserted proposition is paranoid (e.g. 'someone is listening to me through my phone'), her credence may arise more easily to an outright belief.

What about (2)? Why is it that one does not reject the delusional belief after entertaining the negated delusional proposition? The answer here is threefold. First, attentional biases, aberrant salience and anomalous experiences are cognitively burdensome enough to make the rejection of a proposition lot harder, and this is true even if one entertains its negation. Secondly,

persistent exposure to information carried in anomalous experiences (e.g. 'x feels unfamiliar') keeps bringing up delusional hypotheses anew (e.g. 'x is unfamiliar'), and, as already noted, evidence suggests that belief increases with repeated exposure to the believed proposition (Fazio *et al.*, 2019). So, while negating the delusional proposition p might transiently decrease one's credence in p , the fact that the thought ' p ' keeps occurring to one might cause the credence to rise back up again by way of repetition. Finally, as we also noted, the Spinozan view allows for the possibility of inconsistent beliefs, which means that one may believe the negation of p , without necessarily disbelieving p .

Still another problem is that (2) is a specific instance of a more general issue: why cannot we get deluded individuals to believe (and continue to believe) *any* proposition by having them entertain it? If it is abnormally enhanced cognitive load that makes delusions 'sticky' by disabling rejection processes, how come it does not globally inhibit the rejection of whatever thought happens to be passing through one's mind? Take the case of someone with Capgras delusion who cannot disbelieve that proposition 'this is not my wife' because cognitive resources are tapped by an anomalous experience. Should we expect the situation to be any different with other implausible propositions besides the delusional one, such as ones based on consciously experienced visual illusions? Presumably not. But then it seems that Spinozan doxasticism faces an objection vis-à-vis the *monothematicity* of some delusions (cf. Davies *et al.*, 2001). If an anomalous experience inhibits a subject's belief-rejection abilities, as might be the case in Capgras, we would predict that whenever they entertain a bizarrely false proposition, they will believe it. And yet, Capgras, like all other monothematic delusions, involves a single delusion or at most a small set of delusions woven together by a common theme (Coltheart *et al.*, 2011).¹⁵

How to explain this? Consider again Capgras. One possible explanation is that the range of propositions subjects represent to themselves at time t_1 (the time at which the loved one is in sight) is limited to those relevant to explaining the anomalous experience being had. This seems intuitive enough. It is plausible that someone who is attending to an unusually perplexing experience signalling a wife-looking person as unfamiliar would be singularly preoccupied with understanding what makes her unfamiliar. If so, there may be a good reason why subjects do not believe bizarrely false propositions outside the specific domain of the delusion: they simply do not entertain them at the time when cognitive resources are unavailable for rejection.

Of special interest here is the fact that in the majority of monothematic delusions anomalous experiences are not only object specific, which explains why they may trigger some thoughts and not others, but also context sensitive. In Capgras delusion, for instance, the anomalous experience is thought to arise exclusively in contexts where the loved one is visually present

(Ramachandran and Blakeslee, 1998). This means that while the mechanisms of belief revision may be dismantled at t_1 , they may resume correct functioning at t_2 (the time at which the loved one is not in sight), allowing subjects to reject beliefs as normal.

So far, so good. But there remains a loose end that needs tying up. Object specificity and context sensitivity are common to most, though not all, monothematic delusions. In Cotard delusion, for example, it is hypothesised that subjects undergo a generalised experience of unfamiliarity with respect to the environment, to which they respond by denying their own existence (Young *et al.*, 1992). As applied here, the aforementioned proposal does not help explain the fact of monothematicity. For one thing, the experience of unfamiliarity may be directed at a multiplicity of intentional objects (i.e. anything around the subject), resulting in a wider range of propositions being entertained. For another, it is not contingent on particular circumstances (e.g. the presence or absence of a loved one), rendering increased cognitive load invariant across contexts.

We are back then to the question of what prevents subjects from adopting implausible beliefs in domains other than that of the delusion (in this case, Cotard). I think 'nothing' is perhaps the best answer here. Let me explain. As hinted earlier, to say that whenever we entertain that p we automatically come to believe that p is to say that we increase our credence in p to such an extent that it would, in the right circumstances, govern inference and behaviour (see footnote 4). Part of what this means is that there are circumstances in which our beliefs will fail to control inference and behaviour, one being when their content is irrelevant to our current motivational-affective concerns. So, for instance, my belief that John Coltrane started playing saxophone on the alto will hardly lead to any content-congruent behaviour or reasoning in the context where I am rushing to catch a flight. Now suppose I have Cotard delusion and undergo a variety of visual illusions while my appropriate processes for belief rejection are disabled. Suppose, for instance, I look into an Ames distorted room and as a result entertain the implausible hypothesis that the person moving across the room is actually changing size, and not just appearing to change size.¹⁶ Spinozan doxasticism has it that I will eventually acquire and fail to refute the corresponding belief, but does not imply that I will necessarily do something with it. As we just saw, beliefs can lie largely quiescent outside of settings in which their contents are relevant, and the belief contents relevant in the Cotard setting are normally confined to narrow domains, mostly death related.

What should make of all this? The moral to be drawn, I think, is that people with Cotard delusion may adopt and maintain more bizarrely false beliefs than it is apparent from their verbal and non-verbal behaviours. Perhaps, however, while the delusion is prominent in their mental lives and manifest in at least verbal utterances, such other beliefs remain inactive and thus unavailable for guiding inference and behaviour. The reason may

be that, unlike the former, the latter are ill suited for expressing how subjects feel in the current situation or for explaining why their anomalous experiences have the contents they do.

To sum up, there are at least two ways that Spinozan doxasticism might account for the monothematicity of delusions. One is to say that delusion-relevant propositions are the only bizarre propositions being entertained at the moment when the rejection process is suspended, and that is why subjects do not end up with additional bizarre beliefs. The second response concedes that subjects may end up with other bizarre beliefs besides the delusion but argues that those beliefs will tend to go undetected as a result of being of no use, and hence inactivated, in the context in question.

7.3. FRAGMENTATION, CIRCUMSCRIPTION AND DOUBLE-BOOKKEEPING

Before concluding this section, let us see how the Spinozan theory allows for circumscription and double-bookkeeping while retaining the doxastic status of delusions.

One way that doxasticists can attempt to accommodate circumscription is by appeal to fragmentation.¹⁷ As long as beliefs are fragmented, they can be stored separately without having to guide all of our behaviour in all contexts. Instead, because they are stored in separate compartments of the mind, they can be accessed at different moments to drive different behaviours in different contexts for different purposes (Elga and Rayo, forthcoming). To say that fragmented beliefs would guide behaviour in some, but not all, contexts is equivalent to saying that they are *behaviourally* circumscribed. And, of course, the same thing can be said, *mutatis mutandis*, concerning *affective* circumscription and the emotions driven by beliefs. Finally, if we posit a fragmented architecture of belief storage, that makes it possible that beliefs housed in disparate fragments are mutually inconsistent without an agent having to draw out the logical entailments among them or believing their conjunctions. This lack of closure would explain how we can have beliefs that are *inferentially* circumscribed and poorly integrated with the rest. As we saw earlier, it may make perfect sense to say of an agent that she believes (i) that *p* and (ii) that *p* implies *q* even if she does not believe that *q*, provided that we treat her as fragmented and that (i) and (ii) are stored in different fragments.

For doxasticists, the importance of this is that it bolsters their argument by way of the following reasoning. There is a clear correlation between fragmentation and circumscription. So, if our belief system is in fact fragmented, then circumscription supports rather than undermines the continuity between delusions and beliefs (Bortolotti, 2010, p. 89).

One possible objection is that the idea of beliefs being fragmented is an ad hoc stipulation to save doxasticism: (P1) belief fragmentation is typically defended on the grounds that beliefs are not always consistent, deductively

closed and effective in guiding an agent's behaviour (e.g. Egan, 2008b); (P2) the main limitation of doxasticism, as standardly devised, is that it merely asserts that, and sheds no light on why, belief has those features; (P3) the only thing to which doxasticism can appeal in order to explain why belief has these features is that belief is fragmented; (C) therefore, the appeal to fragmentation is not revealing but rather begging the question.

While this objection has some force against the standard defence, it has no, or at least considerably less force, against Spinozan doxasticism. The reason is that the Spinozan architecture is particularly well placed to accommodate a fragmented picture of belief. For the Spinozan, we automatically believe propositions before being able to reject them. In order to reject p , we must already believe p . Rejecting p can only take place if a conflict is detected between p and other beliefs we hold. Because we already believe p , the very existence of a conflict hinges on the possibility of inconsistent beliefs. As such, the second stage of the Spinozan theory, the evaluation of newly acquired beliefs, makes no sense unless our beliefs are fragmented. Moreover, because the rejection process is often disrupted by cognitive load, many inconsistencies are left unresolved.

What has this got to do with doxasticism? Assume the Spinozan theory is true. Then clearly our belief system ought to be fragmented, because, necessarily, many of our beliefs are inconsistent with the rest, and often remain so, even when conflicts are detected. Given this, an appeal to fragmentation here is not susceptible to the charge of being ad hoc, because it is based on a principled theory of belief fixation. Consequently, there is a more robust case to be made that delusions are beliefs.¹⁸

With these observations in mind, let us turn to double-bookkeeping. How do we explain someone's being aware of the delusional nature of their delusions, and yet failing to make up their minds about them? Note that the awareness that is involved by double-bookkeeping can be implicit or explicit. Some subjects might act or react contrary to their delusions, without actually consciously recognising the tension. This is the case, say, with a Capgras subject who states that his wife has been replaced by an imposter and yet continues to live in friendly terms with her. Other subjects, by contrast, might report conflicting beliefs in the course of a single conversation. Even though they explicitly recognise such conflicts, they do nothing to resolve them. For example, McKay and Ciolotti (2007) describe the case of a young woman with Cotard delusion, LU, who claimed to be dead. When asked how she would know when someone is dead, LU replied that dead people lay motionless with their eyes closed. Later in the interview, she recognised the inconsistency between her being dead and yet being able to move and speak, but she continued to maintain that she was dead.

Again, I stress that the question we need to ask is not (at least not only) whether there are quotidian states that exhibit double-bookkeeping and which we are prone to attribute to others as beliefs. The question, rather,

concerns how double-bookkeeping is tied to those states being beliefs. In answering this question, the doxasticist can appeal once more to fragmentation: the phenomenon of double-bookkeeping is difficult to square with a unified web of belief, but naturally explained by a fragmented mind that allows for coexistence of dissonant beliefs. Still, we might feel that this leaves out something important about why subjects are indifferent to dissonance among their beliefs. One can grant that ordinary believers might be inconsistent, and yet expect them to restructure their beliefs once the inconsistency is brought into awareness (Bendana and Mandelbaum, forthcoming). Consider a case where an agent believes p and believes $\neg p$, but experiences no conflict because only the fragment containing p is activated at the time. Because no inconsistency is detected, there is no incentive for the agent to restore coherence in her belief system. Now, imagine another scenario with the same individual, but where p and $\neg p$ are both in activated fragments. Here, we would expect the two fragments to be rendered consistent, because the agent is co-attending to mutually conflicting beliefs. The worry, then, is that appeal to fragmentation alone may not be sufficient to account for double-bookkeeping. To see how Spinozan doxasticism might address this worry, recall that for the Spinozan, rejecting beliefs is a breakdown prone process, one which stalls under cognitive load. What happens when the rejection process goes awry? There are at least two possibilities, each arguably corresponding to one of the two senses of double-bookkeeping outlined earlier.

One possibility is that the agent *thinks* she has discarded some belief in virtue of its inconsistency with the rest, when in fact she has not. Consequently, she refrains from using the belief for conscious planning and verbal behaviour, but continues to access it in low-level behaviour. For instance, one may explicitly disavow any form of racial prejudice, yet nevertheless continue to act in racist ways. This is similar to the case of the Capgras subject who denies his own wife being his wife, but never ceases to treat her as such. The idea is simple. Because the agent thinks she has already re-established coherence in her belief system, she no longer experiences dissonance. However, due to a failure of the revision process, the apparently 'discarded' belief persists and exerts a continuing impact on her behaviour.

A second possibility is that the agent realises that something has gone amiss with the revision process, and remains aware of having inconsistent beliefs, but ultimately tolerates such inconsistencies. This might be what happens with LU in the example earlier, whereby the belief that dead people are speechless coexists with (the verbal expression of) the belief that one is dead. It also may explain the continuity between cases like LU's and non-delusional cases of superstition. People readily combine superstitious beliefs about magic and supernatural causation (e.g. divine intervention) with a scientific stance towards the world. Although they know that such beliefs are irreconcilable, they often fail to decisively resolve the conflict in one way or another, even when called upon to do so (Vyse, 2014).

If this suggestion is on the right track, then a Spinozan architecture could be used to clarify not only why beliefs are fragmented and variously circumscribed but also why people retain inconsistencies between coactive beliefs, resulting in double-bookkeeping.

8. *Remaining concerns*

I have argued that Spinozan doxasticism can accommodate features of delusions that many have felt are telling against their status as beliefs. More importantly, I have argued that Spinozan doxasticism is more explanatory than standard doxasticism, because it explains why delusions, in virtue of being beliefs, have the features they do. As already indicated implicitly, but as I will now make more explicit, to say that delusional beliefs have these features *in virtue of being beliefs* is not to say that *all* of our beliefs have these features. Of course, as Bortolotti (2010, 2018) has convincingly argued, some non-delusional beliefs, such as superstitious beliefs, share many of the key features of delusional beliefs: among other things, they are very resistant to change, even when counterevidence becomes available. However, not all of our beliefs are of that kind. Your belief that it is raining outside is swiftly dislodged when you look outside and see that it is raining. So the question arises, if beliefs really are Spinozan, what explains why delusional beliefs have features that most non-delusional beliefs do not have? The answer lies with the fact, pointed out in Section 7.1, that the former typically arise and are maintained under conditions of high cognitive load (i.e. attentional demands and time pressure), combined with unusual, persistent and distressing experiences (which too are load inducing). The upshot, then, is that Spinozan beliefs need not always display the kinds of features that obtain in cases of delusion, but they do if a particular set of circumstances come together.

Before closing, I have two further issues to address. The first is whether delusions actually fit the distinctive profile of Spinozan beliefs. The second is why doxasticists in general should want to accept such an unorthodox conception of belief.

To the first task, Spinozan beliefs are unconscious propositional attitudes that are not available to introspection; delusions, on the contrary, are conscious, at least in the sense that they are manifested in consciousness by an occurrent thought. How does one reconcile these viewpoints into a single conception of delusions as beliefs? My answer to this is that the conflict in question is not a genuine one. Spinozans deny that we are introspectively aware of beliefs *qua* beliefs. But they do not deny that we can introspect the content of beliefs. So, it may well be that delusions are unconscious beliefs, in the sense that they are content-carrying mental states whose content one can introspect, but whose functional role one cannot.

Qua beliefs, delusions are contents with a certain functional role. This includes their characteristic function in one's mental economy (e.g. their relation to evidence, other kinds of intentional states and behaviour). These relations are not themselves conscious. Thus, one could conclude that delusions are not conscious tout court. All one is aware of is their belief content, yet a content is, in itself, not a belief.

Having said that, it is very much up for grabs whether *all* forms of delusion are unconscious beliefs whose content is present to mind and available for verbal report. Some delusional statements, while sincere, might not be genuine belief reports, that is, reports expressing actual belief contents.

Recall (Section 3.3) that there are two ways in which the Spinozan thinks we find out what we believe. One way is by a direct introspecting act that makes us aware of the content of our beliefs. When this happens, what we sincerely report believing is largely coincident with what we actually believe. Another way proceeds via self-interpretation. We infer what we believe by observing our own behaviour, by considering what seems more reasonable or through other sorts of interpretive strategies. In this case, the 'beliefs' that we report having may well end up being nothing over and above what we actually believe. But most often they are the products of confabulations made up on the spot, which have little if anything in common with our actual beliefs. This is because self-interpretation is influenced by a multiplicity of factors, especially social (e.g. group identity) and motivational (e.g. reducing anxiety). As such, belief reports are often calibrated to fit the beliefs and reactions of those around us.

In essence, there is a distinction to be drawn between beliefs and belief reports. The former are brute architectural matters (Mandelbaum, 2014); our minds are designed to automatically believe any propositions to which they are exposed. The latter express judgments or claims about the contents of beliefs, and as such, they can be genuine (e.g. direct introspective reports) or spurious (e.g. mere endorsements). It is not my aim here to consider which delusions are genuine belief reports and which are not. That will depend, *inter alia*, on how delusional hypotheses arise in the first place. To illustrate, consider the following two scenarios.

- i The sight of your wife does not evoke the characteristic feeling of familiarity that normally accompanies the recognition of her face. You find passing thoughts popping into your mind, among which the implausible hypothesis (i.e. 'that woman who looks like my wife is not really my wife'). Tokening the proposition causes you to believe it, but only weakly at a nonconscious level. After days or weeks, you raise your credence in that proposition, until it becomes available in the form of a judgment (reporting your actual belief) that your wife is not really her.

- ii Other types of delusion might be analysed as the combined result of personal and social concerns, plus the self-evaluation of being relatively immune to such concerns. Concerns might arise with regard to your autonomy and power in relation to others, for instance, about the trustworthiness of friends and associates (paranoia), your class status (grandiosity) or your appealingness or love-worthiness (erotomania). Your struggle to establish an emotionally acceptable self-definition makes it so that you are caught in increasingly arbitrary and idiosyncratic interpretations of yourself and the world, which slowly transform into prolific delusional systems. Some have suggested that these interpretations appear idiosyncratic to others because they are developed in isolation from social relations, perhaps due to impaired Theory of Mind skills (Bora *et al.*, 2009; Bentall, 2018).

Obviously, these are oversimplifications but nonetheless useful ones for our current purposes. Thinking in Spinozan terms, the former scenario (i) corresponds to a case where an agent rightly reports what she believes, and the latter (ii) to cases where an agent merely infers what she takes herself to believe. This distinction brings about an additional advantage of Spinozan doxasticism. It is flexible enough to account for the heterogeneity present in the class of delusion. While some delusional reports may reflect the content of one's actual beliefs, others may be the products of an online elaboration, in which on-the-spot hypotheses are constructed and defended with arguments. It is important to realise, however, that when an agent takes herself to believe that *p*, she automatically believes that *p*. This means that even delusions that arise as *mere* belief reports (i.e. ones that are not reflective of actual belief contents) are liable to become beliefs.

To the second task, standard doxasticists like Bortolotti think delusions are beliefs where what they mean by belief is a state whose fixation can be explained on a psychological level, a level where talk of rational and irrational inferences makes sense. Why should they care if on some other view delusions are 'beliefs', but belief fixation is a brute causal mechanism operating below the psychological level? I make two points about this.

First, for Bortolotti no less than for the Spinozan, beliefs come cheap.¹⁹ In the place of rationality constraints, Bortolotti (2010) offers more realistic features of beliefs that she thinks are to guide everyday interpretation:

- i Beliefs have *some* inferential relations with other beliefs, wishes, desires and so forth.
- ii Beliefs display *some* sensitivity to evidence and argument.
- iii Beliefs can be, but need not be, manifested in behaviour.

- iv Beliefs can be self-ascribed (i.e. acknowledged as one's own), and their content can be endorsed (i.e. defended with reasons).

Each of these features fits squarely with the Spinozan conception of belief. From the foregoing discussion, it should be obvious that Spinozan beliefs are (i) inferentially promiscuous and able to (iii) cause behaviour. It should also be clear that at least some Spinozan beliefs are (ii) sensitive to evidence. For one thing, they can be revised in the wake of new information, although the process of revision is often short-circuited by cognitive load. For another, they can be formed, albeit contingently, based on evidence. Think, for example, of perceptual beliefs, whereby one believes that p based on a perceptual experience with the content that p . Finally, we have seen that, (iv) while beliefs are acquired in ways not available to introspection, agents can ascribe beliefs to themselves by introspecting their contents. In doing so, they also endorse such contents, which is to say that they are disposed to defend them with reasons or using rhetorical strategies.

Second, Spinozan doxasticism yields two results that all doxasticists should like. First, it entails that delusions are the same kind of attitudes as beliefs concerning garden-variety facts, such as the belief that there is some leftover pizza in the fridge. The entailment is based on the following reasoning. To have any attitude towards a propositional content p is equivalent to believing p . In having a delusion that p , one is psychologically related to p . Therefore, being deluded that p is the same kind of attitude as believing that there is some leftover pizza in the fridge. Secondly, it offers a simple and intuitively satisfying explanation of why the key features of delusion should be taken as features of belief: delusions are the way they are as a result of the way belief fixation actually works.

9. Conclusion

The Spinozan theory of belief fixation holds that the mere act of representing a proposition leads to immediately believing it. Minds like ours are such that they cannot *merely* represent a proposition. Rather, propositions are believed as quickly as they are represented. I have argued here that this view has important consequences for the debate over the doxastic status of delusions. Specifically, if we accept this view, it gives a more perspicuous defence of doxasticism than the dominant one in the literature, offered by Bortolotti. Doxasticism has been criticised on the grounds that delusions fail to conform to certain rationality standards that we expect beliefs to conform to. Against this objection, Bortolotti points to typical cases of belief ascription that also fall foul of rationality standards. However, some may say this is putting the cart before the horse. Doxasticism is supposed to make us

understand why delusions qua beliefs behave as they do. It does not do that if it only focuses on belief ascription. I have shown that Spinozan doxasticism, unlike the standard doxastic defence, is capable of providing such an understanding.²⁰

Department of Philosophy
University of Birmingham

NOTES

¹ For current purposes, to ‘entertain’ a proposition is to have access to it, though not necessarily consciously so. Because one has access to p whenever p is embedded in any propositional attitude, entertaining p is integral to any propositional attitude one bears to p . Instances of ‘entertaining that p ’ are ‘perceiving that p ’, ‘considering that p ’, ‘supposing that p ’, ‘imagining that p ’, ‘accepting that p for the sake of argument’ and so forth.

² If you are to take a doxastic attitude towards p , three alternatives are possible: believing that p is true (believing p), believing that p is false (disbelieving p) or refraining from either believing or disbelieving p (suspending judgment about p). To keep things simple, I will henceforth speak only in terms of belief and disbelief. For more on suspended judgment, see Friedman (2013).

³ Examples are imaginings that one misinterprets as beliefs (Currie and Jureidini, 2001; Currie and Ravenscroft, 2002), perceptual illusions (Hohwy and Rajan, 2012), empty speech acts (Berrios, 1991), in-between states part belief-like, part not belief-like (Schwitzgebel, 2012), ‘bimaginings’ part belief-like, part imagining-like (Egan, 2008a), acceptances (Frankish, 2013; Dub, 2017) or thoughts reporting the contents of default processes (Gerrans, 2014).

⁴ A cautionary note about terminology is due before setting off. The Spinozan theory works with a graded notion of belief, which allows that one can believe things with varying degrees of confidence or credence (Mandelbaum, 2014). One’s belief that p is assigned a value in the interval $[0,1]$ that represents one’s degree of confidence in that proposition – the higher the number, the greater the degree of confidence. This contrasts with the standard notion of belief as a binary, all-or-none state, where one either believes or does not believe that p . For the Spinozan, entertaining a proposition p causes one’s credence in p to become higher. How high? Mandelbaum (2014) contends that the credence is raised to a level that, under the right circumstances, would guide action and that would allow p to serve as a premise in inferences – presumably a level much higher than 0.0001 (as that would be behaviourally inert) and yet it need not be extremely high (one that equals or exceeds 0.9).

⁵ ‘Effortful’ here is used in the narrow sense of ‘cognitively demanding’, covering any mental task that consumes substantial cognitive resources. In this narrow sense, it is possible for a mental task to be effortful and yet occur outside of conscious awareness and without conscious intent (Mandelbaum, 2010, 2014).

⁶ More studies gave results purportedly inconsistent with the Spinozan theory. Hasson *et al.* (2005) presented their participants with statements that differed in their informativeness. It was found that the cognitive load manipulations selectively impaired their ability to remember statements that were uninformative when false (i.e. this person walks barefoot to work), but had no effect on statements that were informative when false (i.e. this person owns a television). See Mandelbaum (2014, pp. 71–75) for a proposal on how to accommodate these findings within the

Spinozan theory. Other authors (Richter *et al.*, 2009) reported that cognitive load during feedback processing reduces memory for false statements for which participants had no or weak prior knowledge ('Krypton is a noble gas'), but not for false statements for which they had strong prior knowledge ('Soft soap is edible'). In contrast with this evidence, recently published findings suggest that belief in false statements is increased by repetition independent of prior knowledge, whether it be none, weak or strong (Fazio *et al.*, 2015; Unkelbach and Greifeneder, 2018).

⁷ In a technical sense, 'inconsistent' does not mean the same as 'contradictory'. If beliefs are inconsistent, then they cannot both be true but can both be false (e.g. 'Vicky is my sister' and 'Vicky is my daughter'); whereas if beliefs are contradictory, then both cannot be true and cannot be false (e.g. 'Vicky is my sister' and 'Vicky is not my sister'). For the sake of simplicity, however, I will here use 'inconsistent' in a broad sense to also include contradictory beliefs.

⁸ It goes without saying that this implies a doxastic model of implicit bias, according to which the implicit attitudes driving racist behaviours are unconscious beliefs. This is a minority view among implicit bias theorists (see, e.g. Mandelbaum and Quilty-Dunn, 2015; Frankish, 2016; Mandelbaum, 2016; cf. Levy, 2015; Holroyd, 2016; Madva, 2016).

⁹ For instance, perhaps you went hiking in the Swiss Alps a month ago, so you will infer that you are not prone to altitude sickness or you would have had troubles breathing in.

¹⁰ A natural worry here is that this strategy turns on a surreptitious slide between the folk concept of belief, comprising the properties that we stereotypically associate with belief, and the concept of belief operative in the Spinozan theory, comprising the psychological laws that allegedly govern belief fixation. That is, one might suspect that with such a strategy the 'belief' status of the latter concept would depend on whatever properties the folk ascribe to the former. The worry is misplaced for the reason that Mandelbaum is doing something altogether different: showing that there are several important respects in which these two concepts are more alike than different. The take-home message, as we will see shortly, is that the states under discussion are belief-like enough to count as 'beliefs' by commonsense standards. Of course, some people will disagree as to how much is enough, but that need not concern us here, as we are not interested in whether the Spinozan theory is true, but only in whether, if true, it supports doxasticism about delusions.

¹¹ This brings us naturally to the following question: why do people who interact with the deluded person not also adopt the delusion? Were all else equal, they too should entertain the delusional proposition (e.g. 'that is not his wife') and thus believe it. But all else is not equal. As we will see (Section 7.1), explaining why they do not believe it is going to be part of the story for why the deluded person does. A further, related, question is why the deluded person cannot drop his delusion by just hearing someone deny its content (e.g. by hearing the misidentified person say 'I am your wife'). This too is addressed in Section 7.1.

¹² The 'evidence' for some delusions comes in the form of irregular perceptual experiences (e.g. Stone and Young, 1997; Maher, 1999; Davies *et al.*, 2001). Also, some delusions are sensitive to arguments and evidence to the contrary (e.g. Bortolotti, 2010; Flores, forthcoming).

¹³ It should be noted that the degree to which delusions are behaviourally and affectively circumscribed can vary considerably from case to case. It is not as if delusions *never* motivate behaviour and affect in the way we would expect them to. Clinical studies have reported cases in which Capgras subjects grow extremely distressed and act against the alleged imposters, at times with deadly consequences (e.g. Christodoulou, 1977; De Pauw and Szulecka, 1988). This, however, does not make circumscription any less problematic from the point of view of doxasticists. Merely citing the existence of conflicting cases does not really explain what needs explaining, which is why delusions are *so often* weakly behaviour guiding and not accompanied by emotional responses appropriate to their content (Dub, 2017).

¹⁴ If 'evidence' were understood loosely to mean 'information' or 'data', one could claim that Spinozan belief fixation is hyperresponsive to evidence, because every proposition represented is thus believed. Here, however, 'evidence' should be understood in the epistemological sense, namely, as something that makes a difference to what is reasonable for one to believe

(e.g. Kelly, 2014). So, when I say that Spinozan belief fixation is ‘unresponsive to evidence’, I mean that believing occurs irrespective of what one is justified in believing.

¹⁵ I am grateful to an anonymous referee for raising this problem.

¹⁶ For the reader who has never heard of Ames Room before, it is a room specially designed to create an illusion in which items appear to change size as they are moved from one corner to the other (e.g. Gregory, 1970, pp. 26–28).

¹⁷ Bayne and Pacherie (2005) argue that behavioural circumscription is excusable by reference to non-standard features of the circumstance one finds oneself in. For example, it is difficult to say which action or attitude would be consistent with a person’s belief that someone is inserting thoughts into her mind. Clearly, however, not every case of behavioural circumscription can be excused in this way: it is plausible to expect of someone who believes her mother has been replaced by an imposter to report her missing. But there is more. Bayne and Pacherie’s point is not just that it is hard to know what to expect in some cases of delusion. Rather, the point is that even in cases where the behaviour we would expect is not manifested, that can be excused: it might be that a person does not turn to the police because she knows that the imposter is a perfect lookalike and she fears she will not be taken seriously. This is plausible, but cannot be generalised across all the behavioural dispositions that delusional individuals fail to manifest. What should excuse a person who believes she is being poisoned from eating her food? It is hard to imagine anyone choosing the possibility of severe illness or death over being sectioned or whatever else.

¹⁸ In order to avoid confusion, let me emphasise that I am not trying to argue that the Spinozan theory provides the best explanation (tout court) for belief fragmentation. My point is just that Spinozan doxasticism fares better than standard doxasticism in this respect. As I have been arguing, standard doxasticism does not tell us why key features of delusions, such as circumscription and inconsistency, or equivalently, double-bookkeeping, are features of beliefs. What the standard doxasticist can do is appeal to belief fragmentation: beliefs are circumscribed and mutually inconsistent because they are fragmented. The problem with this is that it may strike one as circular, because circumscription and inconsistency are themselves among the best evidence supporting fragmentation. Spinozan doxasticism avoids the charge of circularity by giving an account of belief fixation that ‘ensures’ fragmentation.

¹⁹ Note that this is not unique to Bortolotti. Doxasticists in general are likely to be liberal about belief as compared with non-doxasticists, who tend to be stricter (e.g. Archer, 2013; Ichino, 2020). Here, I focus on Bortolotti as the most prominent representative of doxasticism.

²⁰ Thanks to Lisa Bortolotti, Ema Sullivan-Bissett, Tim Bayne, Henry Taylor and Eric Mandelbaum for extremely useful comments and feedback. Thanks especially to two anonymous referees for this journal whose constructive comments helped improve the paper. Thanks also to the Arts and Humanities Research Council for their generous support.

REFERENCES

- Alexander, M. P., Stuss, D. T. and Benson, D. F. (1979). ‘Capgras Syndrome: A Reduplicative Phenomenon,’ *Neurology* 29(3), pp. 334–339.
- Appel, M. and Richter, T. (2007). ‘Persuasive Effects of Fictional Narratives Increase Over Time,’ *Media Psychology* 10(1), pp. 113–134.
- Archer, S. (2013). ‘Nondoxasticism About Self-deception,’ *Dialectica* 67(3), pp. 265–282.
- Aronson, E. (1999). ‘Dissonance, Hypocrisy, and the Self Concept,’ in E. Harmon-Jones and J. Mills (eds) *Cognitive Dissonance: Progress on a Pivotal Theory in Social Psychology*. Washington: American Psychological Association, pp. 103–126.
- Asp, E. W., Khan, L. B., Jonason, A. B., Adkins-Hempel, M. M., Warner, K. A., Pardilla-Delgado, E., Payne, J. D., Mullineaux, P., Denburg, N. L., and Tranel, D. (n.d.).

- Second-guessing of Spinoza: Psychophysiological and Behavioural Evidence that Believing Is Default During Proposition Comprehension. Unpublished Manuscript.
- Bayne, T. and Pacherie, E. (2005). 'In Defence of the Doxastic Conception of Delusions,' *Mind & Language* 20(2), pp. 163–188.
- Bendana, J. and Mandelbaum, E. (forthcoming). 'The Fragmentation of Belief,' in D. Kinderman, A. Onofri and C. Borgoni (eds) *The Fragmented Mind*. Oxford: Oxford University Press.
- Bentall, R. P. (2018). 'Delusions and Other Beliefs,' in L. Bortolotti (ed.) *Delusions in Context*. Cham: Springer International Publishing, pp. 67–95.
- Bentall, R. P. and Kaney, S. (1989). 'Content Specific Information Processing and Persecutory Delusions: An Investigation Using the Emotional Stroop Task,' *British Journal of Medical Psychology* 62, pp. 355–364.
- Berrios, G. E. (1991). 'Delusions as 'Wrong Beliefs': A Conceptual History,' *British Journal of Psychiatry* 159(S14), pp. 6–13.
- Berti, A., Ladavas, E., Stracciari, A., Giannarelli, C. and Ossola, A. (1998). 'Anosognosia for Motor Impairment and Dissociations with Patients Evaluation of the Disorder: Theoretical Considerations,' *Cognitive Neuropsychiatry* 3(1), pp. 21–43.
- Bora, E., Yucel, M. and Pantelis, C. (2009). 'Theory of Mind Impairment in Schizophrenia: Meta-analysis,' *Schizophrenia Research* 109(1–3), pp. 1–9.
- Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.
- Bortolotti, L. (2012). 'In Defence of Modest Doxasticism About Delusions,' *Neuroethics* 5(1), pp. 39–53.
- Bortolotti, L. (2018). 'Delusions and Three Myths of Irrational Belief,' in L. Bortolotti (ed.) *Delusions in Context*. Cham: Palgrave Macmillan.
- Brighetti, G., Bonifacci, P., Borlimi, R. and Ottaviani, C. (2007). 'Far from the Heart Far from the Eye': Evidence from the Capgras Delusion,' *Cognitive Neuropsychiatry* 12, pp. 189–197.
- Campbell, J. (2001). 'Rationality, Meaning, and the Analysis of Delusion,' *Philosophy, Psychiatry, and Psychology* 8(2), pp. 89–100.
- Christodoulou, G. N. (1977). 'The Syndrome of Capgras,' *British Journal of Psychiatry* 130(6), pp. 556–564.
- Coltheart, M. (2013). 'On the Distinction Between Monothematic and Polythematic Delusions,' *Mind & Language* 28(1), pp. 103–112.
- Coltheart, M., Langdon, R. and McKay, R. (2011). 'Delusional Belief,' *Annual Review of Psychology* 62, pp. 271–298.
- Currie, G. (2000). 'Imagination, Delusion and Hallucinations,' *Mind & Language* 15(1), pp. 168–183.
- Currie, G. and Jureidini, J. (2001). 'Delusion, Rationality, Empathy: Commentary on Martin Davies et al,' *Philosophy, Psychiatry, and Psychology* 8(2), pp. 159–162.
- Currie, G. and Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.
- Davidson, D. (1982). 'Paradoxes of Irrationality,' in R. Wollheim and J. Hopkins (eds) *Philosophical Essays on Freud*. Cambridge: Cambridge University Press, pp. 289–305.
- Davidson, D. (1985). 'Incoherence and Irrationality,' *Dialectica* 39(4), pp. 345–354.
- Davies, M., Coltheart, M., Langdon, R. and Breen, N. (2001). 'Monothematic Delusions: Towards a Two-factor Account,' *Philosophy, Psychiatry, and Psychology* 8(2/3), pp. 133–158.
- De Pauw, K. W. and Szulecka, T. K. (1988). 'Dangerous Delusions. Violence and the Misidentification Syndromes,' *British Journal of Psychiatry* 152(Jan.), pp. 91–96.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

- Dub, R. (2017). 'Delusions, Acceptances, and Cognitive Feelings,' *Philosophy and Phenomenological Research* 94(1), pp. 27–60.
- Egan, A. (2008a). 'Imagination, Delusion, and Self-deception,' in T. Bayne and J. Fernandez (eds) *Delusion and Self-deception: Affective and Motivational Influences on Belief Formation*. Hove: Psychology Press.
- Egan, A. (2008b). 'Seeing and Believing: Perception, Belief Formation and the Divided Mind,' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 140 (1), pp. 47–63.
- Elga, A. and Rayo, A. (forthcoming). 'The Fragmentation of Belief,' in D. Kinderman, A. Onofri and C. Borgoni (eds) *The Fragmented Mind*. Oxford: Oxford University Press.
- Ellis, H. D., Young, A. W., Quayle, A. and de Pauw, K. W. (1997). 'Reduced Autonomic Responses to Faces in Capgras Delusion,' *Proceedings of the Royal Society: Biological Sciences* B264, pp. 1085–1092.
- Fazio, L. K., Brashier, N. M., Payne, B. K. and Marsh, E. J. (2015). 'Knowledge Does Not Protect Against Illusory Truth,' *Journal of Experimental Psychology: General* 144(5), pp. 993–1002.
- Fazio, L. K., Rand, D. G. and Pennycook, G. (2019). 'Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements,' *Psychonomic Bulletin and Review* 26, pp. 1705–1710.
- Fear, C., Sharp, H. and Healy, D. (1996). 'Cognitive Processes in Delusional Disorders,' *The British Journal of Psychiatry* 168(1), pp. 61–67.
- Flores, C. (forthcoming). 'Delusional Evidence-responsiveness,' *Synthese*.
- Fodor, J. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press.
- Frankish, K. (2013). 'Delusions: A Two-level Framework,' in M. Broome and L. Bortolotti (eds) *Psychiatry as Cognitive Neuroscience*. Oxford: Oxford University Press, pp. 269–285.
- Frankish, K. (2016). 'Playing Double: Implicit Bias, Dual Levels, and Self-control,' in M. Brownstein and J. Saul (eds) *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford: Oxford University Press, pp. 23–46.
- Friedman, J. (2013). 'Suspended Judgment,' *Philosophical Studies* 162(2), pp. 165–181.
- Gerrans, P. (2014). *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought*. Cambridge, MA: MIT Press.
- Gilbert, D. T. (1991). 'How Mental Systems Believe,' *American Psychologist* 46(2), pp. 107–119.
- Gilbert, D. T., Krull, D. S. and Malone, P. S. (1990). 'Unbelieving the Unbelievable: Some Problems in the Rejection of False Information,' *Journal of Personality and Social Psychology* 59(4), pp. 601–613.
- Gilbert, D. T., Tafarodi, R. W. and Malone, P. S. (1993). 'You Cannot Not Believe Everything You Read,' *Journal of Personality and Social Psychology* 65(2), pp. 221–233.
- Gotlib, I. H. and McCann, C. D. (1984). 'Construct Accessibility and Depression: An Examination of Cognitive and Affective Factors,' *Journal of Personality and Social Psychology* 47, pp. 427–439.
- Green, M. C. and Brock, T. C. (2002). 'In the Mind's Eye: Transportation-imagery Model of Narrative Persuasion,' in M. C. Green, J. J. Strange and T. C. Brock (eds) *Narrative Impact: Social and Cognitive Foundations*. Mahwah: Lawrence Erlbaum, pp. 315–341.
- Gregory, R. L. (1970). *The Intelligent Eye*. London: Weidenfeld and Nicolson.
- Hamilton, A. (2007). 'Against the Belief Model of Delusion,' in M. C. Chung, K. W. M. Fulford and G. Graham (eds) *Reconceiving Schizophrenia*. Oxford: Oxford University Press, pp. 217–234.
- Hasson, U., Simmons, J. P. and Todorov, A. (2005). 'Believe It or Not: On the Possibility of Suspending Belief,' *Psychological Science* 16, pp. 566–571.

- Hirstein, W. and Ramachandran, V. S. (1997). 'Capgras Syndrome: A Novel Probe for Understanding the Neural Representation of the Identity and Familiarity of Persons,' *Proceedings of the Royal Society of London B: Biological Sciences* 264, pp. 437–444.
- Hohwy, J. and Rajan, V. (2012). 'Delusions as Forensically Disturbing Perceptual Inferences,' *Neuroethics* 5, pp. 5–11.
- Holroyd, J. (2016). 'What Do We Want from a Model of Implicit Cognition?' *Proceedings of the Aristotelian Society* 116(2), pp. 153–179.
- Huebner, B. (2009). 'Troubles with Stereotypes for Spinozan Minds,' *Philosophy of the Social Sciences* 39(1), pp. 63–92.
- Ichino, A. (2020). 'Superstitious Confabulations,' *Topoi* 39, pp. 203–217.
- Kapur, S. (2003). 'Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia,' *American Journal of Psychiatry* 160 (1), pp. 13–23.
- Kapur, S. (2004). 'How Antipsychotics Become 'Anti-psychotic' – From Dopamine to Salience to Psychosis,' *Trends in Pharmacological Sciences* 25(8), pp. 402–406.
- Kelly, T. (2014). 'Evidence,' in E. N. Zalta (ed.) *Stanford Encyclopedia of Philosophy (Fall)*. <https://plato.stanford.edu/entries/evidence/>
- Langdon, R., Connaughton, E. and Coltheart, M. (2014). 'The Fregoli Delusion: A Disorder of Person Identification and Tracking,' *Topics in Cognitive Science* 6(4), pp. 615–631.
- Leafhead, K. M., Young, A. W. and Szulecka, T. K. (1996). 'Delusions Demand Attention,' *Cognitive Neuropsychiatry* 1(1), pp. 5–16.
- Levy, N. (2015). 'Neither Fish Nor Fowl: Implicit Attitudes as Patchy Endorsements,' *Nous* 49 (4), pp. 800–823.
- Levy, N. and Mandelbaum, E. (2014). 'The Powers that Bind: Doxastic Voluntarism and Epistemic Obligation,' in J. Matheson and R. Vitz (eds) *The Ethics of Belief*. Oxford: Oxford University Press, pp. 15–32.
- Lewis, D. (1982). 'Logic for Equivocators,' *Nous* 16(3), p. 431–441.
- Lucchelli, F. and Spinnler, H. (2007). 'The Case of Lost Wilma: A Clinical Report of Capgras Delusion,' *Neurological Sciences* 28(4), pp. 188–195.
- Madva, A. (2016). 'Why Implicit Attitudes Are (Probably) Not Beliefs,' *Synthese* 193(8), pp. 2659–2684.
- Maher, B. A. (1999). 'Anomalous Experience in Everyday Life: Its Significance for Psychopathology,' *The Monist* 82(4), pp. 547–570.
- Mandelbaum, E. (2010). *The Architecture of Belief: An Essay on the Unbearable Automaticity of Believing*. PhD Diss. Chapel Hill: University of North Carolina.
- Mandelbaum, E. (2014). 'Thinking Is Believing,' *Inquiry* 57(1), pp. 55–96.
- Mandelbaum, E. (2016). 'Attitude, Inference, Association: On the Propositional Structure of Implicit Bias,' *Nous* 50(3), pp. 629–658.
- Mandelbaum, E. and Quilty-Dunn, J. (2015). 'Believing Without Reason: or: Why Liberals Shouldn't Watch Fox News,' *The Harvard Review of Philosophy* 22, pp. 42–52.
- Marsh, E. J. and Fazio, L. K. (2006). 'Learning Errors from Fiction: Difficulties in Reducing Reliance on Fictional Stories,' *Memory and Cognition* 34, pp. 1140–1149.
- Matthews, R. J. (2013). 'Belief and Belief's Penumbra,' in N. Nottelmann (ed.) *New Essays on Belief: Constitution, Content and Structure*. London: Palgrave Macmillan, pp. 100–123.
- McGhie, A. and Chapman, J. (1961). 'Disorders of Attention and Perception in Early Schizophrenia,' *British Journal of Medical Psychology* 34(2), pp. 103–116.
- McKay, R. and Cipelotti, L. (2007). 'Attributional Style in a Case of Cotard Delusion,' *Consciousness and Cognition* 16(2), pp. 349–359.
- Nadarevic, L. and Erdfelder, E. (2019). 'More Evidence Against the Spinozan Model: Cognitive Load Diminished Memory for 'true' Feedback', *Memory and Cognition* 47, pp. 1386–1400.

- Pilyshyn, Z. W. (1989). 'Computing in Cognitive Science,' in M. I. Posner (ed.) *Foundations of Cognitive Science*. Cambridge, MA: MIT Press, pp. 49–91.
- Quine, W. V. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Quine, W. V. and Ullian, J. S. (1970). *The Web of Belief*. New York: Random House.
- Ramachandran, V. S. and Blakeslee, S. (1998). *Phantoms in the Brain: Human Nature and the Architecture of the Mind*. London: Fourth Estate.
- Richter, T., Schroeder, S. and Wohrmann, B. (2009). 'You Do Not Have to Believe Everything You Read: Background Knowledge Permits Fast and Efficient Validation of Information,' *Journal of Personality and Social Psychology* 96, pp. 538–558.
- Rose, D., Buckwalter, W. and Turri, J. (2014). 'When Words Speak Louder than Actions: Delusion, Belief, and the Power of Assertion,' *Australasian Journal of Philosophy* 92(4), pp. 683–700.
- Sass, L. (1992). *Madness and Modernism: Insanity in the Light of Modern Art, Literature, and Thought*. New York: Basic Books.
- Sass, L. (1994). *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind*. Ithaca: Cornell University Press.
- Schwitzgebel, E. (2012). 'Mad Belief?' *Neuroethics* 5(1), pp. 13–17.
- Sims, A. (2003). *Symptoms in the Mind*, 3rd edn. London: Elsevier.
- Skurnik, I., Yoon, C., Park, D. C. and Schwarz, N. (2005). 'How Warnings About False Claims Become Recommendations,' *The Journal of Consumer Research* 31, pp. 713–724.
- Stalnaker, R. (1984). *Inquiry*. Cambridge: Cambridge University Press.
- Stich, S. P. (1978). 'Beliefs and Subdoxastic States,' *Philosophy of Science* 45(4), pp. 499–518.
- Stone, T. and Young, A. W. (1997). 'Delusions and Brain Injury: The Philosophy and Psychology of Belief,' *Mind and Language* 12, pp. 327–364.
- Stone, T. and Young, A. W. (2007). 'Delusions and Brain Injury: The Philosophy and Psychology of Belief,' *Mind & Language* 12(3–4), pp. 327–364.
- Street, C. N. H. and Kingstone, A. (2016). 'Aligning Spinoza with Descartes: An Informed Cartesian Account of the Truth Bias,' *British Journal of Psychology* 108, pp. 453–466.
- Unkelbach, C. and Greifeneder, R. (2018). 'Experiential Fluency and Declarative Advice Jointly Inform Judgments of Truth,' *Journal of Experimental Social Psychology* 79, pp. 78–86.
- Vyse, S. (2014). *Believing in Magic. The Psychology of Superstition*. New York: Oxford University Press.
- Wedgwood, R. (2017). *The Value of Rationality*. Oxford: Oxford University Press.
- Wegner, D. M., Erber, R. and Zanakos, S. (1993). On Trying Not to Be Sexist. Unpublished Manuscript.
- Young, A. W. (2000). "Wondrous Strange": The Neuropsychology of Abnormal Beliefs,' *Mind & Language* 15(1), pp. 47–73.
- Young, A. W. and Leafhead, K. M. (1996). 'Betwixt Life and Death: Case Studies of Cotard Delusion,' in P. W. Halligan and J. C. Marshall (eds) *Method in Madness: Case Studies in Cognitive Neuropsychiatry*. East Sussex: Taylor and Francis, pp. 147–171.
- Young, A. W., Robertson, I. H., Hellowell, D. J., de Pauw, K. W. and Pentland, B. (1992). 'Cotard Delusion after Brain Injury,' *Psychological Medicine* 22, pp. 799–804.