

## DELUSIONS ARE BELIEFS (JUST NOT THE KIND YOU THOUGHT)

### Los delirios son creencias (solo que no del tipo que pensabas)

RAJEEV R. DUTTA, B.S. <sup>a</sup>

<https://orcid.org/0000-0002-6534-8812>

duttarr@hs.uci.edu

<sup>a</sup> Department of Philosophy and School of Medicine, University of California, Irvine, California, United States.

#### Abstract

The idea that delusions are beliefs is supported by the observation that delusions, similarly to beliefs, are used in reasoning. However, delusions also exhibit other features that are difficult to explain under this doxastic view—they strongly resist evidence and sometimes conflict with an agent's actions (in ways in which beliefs seemingly do not), giving rise to what is known as the *double bookkeeping phenomenon*. These features have motivated non-doxastic views, arguing that delusions are other types of mental phenomena (e.g., imaginings or empty speech acts). While these non-doxastic views account for the features of delusions like evidence resistance and double bookkeeping, they struggle to explain the belief-like features of delusions. We are thus at an impasse, where neither the doxastic nor the non-doxastic account of delusions can explain all the features of delusions. In this paper, I aim to make progress in this debate by showing that the doxastic view can, after all, account for both the belief-like and the other features of delusions. Drawing on recent literature in epistemology, I argue that (rational) beliefs can be motivated by non-evidential (e.g., pragmatic and emotional) factors, and that these factors help explain delusions' otherwise puzzling features. I also propose that categorizing beliefs across two functional compartments, the implicit and the explicit, is helpful for understanding the dynamic interplay of delusions and non-delusional beliefs.

**Key words:** Delusions; Doxastic; Non-Evidential; Beliefs.

#### Resumen

La idea de que los delirios son creencias es apoyada por la observación de que los delirios, al igual que las creencias, son usados en el razonamiento. Sin embargo, los delirios tienen otras características que son difíciles de explicar por medio de una teoría doxástica —por ejemplo, los delirios persisten a pesar de la evidencia en su contra y algunas veces entran en conflicto con las acciones del agente (de maneras en que las creencias parecerían no hacerlo), dando lugar a lo que se conoce como el *fenómeno de la doble cuenta*—. Estas características han motivado teorías no doxásticas, que argumentan que los delirios son otro tipo de fenómeno mental (e.g., imaginaciones o actos de habla vacíos). Mientras que las teorías no-doxásticas pueden explicar dichas características de los delirios—como su persistencia a pesar de la evidencia en su contra y el fenómeno de la doble cuenta— estas

teorías tienen dificultades para explicar las características de los delirios que los hacen similares a las creencias. Nos encontramos entonces en un punto muerto en el que ni los enfoques doxásticos ni los no doxásticos pueden explicar todas las características de los delirios. En este artículo, propongo avanzar en este debate mostrando que la propuesta doxástica puede, después de todo, dar cuenta tanto de las características de los delirios que los hacen similares a las creencias como de sus demás características. Basándome en literatura reciente en epistemología, argumento que las creencias (rationales) puede estar motivadas por factores no evidenciales (e.g. pragmáticos o emocionales) y que estos factores sirven para explicar las características aparentemente desconcertantes de los delirios. También propongo que categorizar las creencias a través de dos compartimentos funcionales, el implícito y el explícito, nos ayuda a entender la interacción dinámica entre creencias delirantes y no delirantes.

**Palabras clave:** Delirios; Doxástico; No evidencial; Creencias.

## 1. Delusions in Context

Consider the following excerpts from case studies about delusions:

**CIA-Hospital:** “One of our patients from an open ward claimed that the **hospital was surrounded by CIA agents** only waiting to kill him. Nonetheless, **he went to buy an ice cream** apparently undisturbed in a kiosk **outside the hospital**” (Parnas et al., 2020; emphasis mine in all case studies).

**False Confession:** “The patient... stated that ‘I feel everything is unreal. I feel suicidal and guilt’ and endorsed a plan to either walk into traffic or shoot himself in the head due to increasingly distressing thoughts and memories. According to the patient, **he had reported to the police that he raped his ex-girlfriend** a year previously, although **she denied the claim to the police**” (Chhaya, 2017).<sup>1</sup>

**Pope-Farmer:** “...the conduct of the patients is inadequate. They really do nothing to attain their goal; **the [self-proclaimed] emperor and the [self-proclaimed] pope help to manure the fields; the [self-proclaimed] queen of heaven irons the patients’ shirts or besmears herself and the table with saliva**” (Bleuler, 1924, pp. 391-392).

<sup>1</sup> We might worry that this case is one in which the ex-girlfriend lied to the police to protect the man (since this dynamic is known to occur in abusive relationships). Chhaya makes clear in the case study that this man is having delusions (that is, multiple delusions, only one of which I have mentioned here).

Though delusions are heterogenous in many respects (including, but not limited to their etiology and content), I take it that these cases help us understand some of the common features of delusions, listed below.<sup>2</sup>

**Irrationality:** Delusions are irrational (i.e., they do not accord with evidence available to the agent).<sup>3</sup>

**Double Bookkeeping:** Delusions need not lead to action concordant with their content.

**Evidence-Resistance:** Delusions are resistant to counter-evidence.<sup>4</sup>

**Irrationality** and **Evidence-Resistance** are well-known features of delusions, often considered necessary conditions for classifying a mental state as a delusion. **Double Bookkeeping**, by contrast, is not as commonly known. Psychiatrist Josef Parnas writes “Even though the phenomenon is probably very well known to most experienced clinicians (though not in an explicit or conceptual way), it is completely neglected in contemporary mainstream psychiatry. However, in the last ten years we have witnessed emerging interest in the phenomenon of double bookkeeping” (Parnas et al., 2020). According to Parnas, the phenomenon is gaining traction because its complex clinical presentation seems to suggest that “the patient’s experience of the world must not simply be mistaken, but somehow altered or transformed in a global way” (Parnas et al., 2020). **Double Bookkeeping** is thus another factor of both epistemic and clinical interest in the effort to characterize delusions.

Gaining a better understanding of the nature of delusions may have clinical implications. For example, if delusions are most akin to beliefs, cognitive-behavioral models of delusion remission might be more effective. If, by contrast, delusions are more akin to imaginings, approaches such as interactive guided imagery therapy might be more effective to promote a patient’s ability to distinguish imaginings from reality.

The doxastic view of delusions (i.e., the view that delusions are beliefs) has emerged as the standard account (*Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, 2013) and

<sup>2</sup> Two of them are taken from clinical studies, and the third is from Eugen Bleuler’s seminal *Textbook of Psychiatry* (Bleuler himself coined the term “schizophrenia”).

<sup>3</sup> This is a provisional characterization. In section 3, we will develop more precise vocabulary with which we can update **Irrationality**.

<sup>4</sup> Flores (2021) argues convincingly that delusions are *resistant* to evidence, rather than completely irresponsive to evidence. Those unmoved by her argument may choose to formulate **Evidence-Resistance** as “Delusions are irresponsive to evidence” instead (i.e., *Evidence-Irresponsiveness*).

consequently, has been often defended in the literature (see, e.g., Bortolotti, 2010, Bayne & Pacherie, 2005).

Traditional doxastic views compellingly describe the belief-like behaviors of delusions. Consider, for instance, **False Confession**, the case in which a man has a delusion that he assaulted his girlfriend. In accordance with the doxastic view, the man reasons with and acts on his delusion: he turns himself into the authorities because he is convinced that he has committed a crime. The doxastic view of delusions thus seems to yield the intuitive conclusion that the man genuinely believes the content of his delusion and is disposed to act in accordance with it. Further, since delusions are a subset of beliefs, it follows that, since not all beliefs are rational, delusions can be irrational (and in my view are constitutively so). In the **Pope-Farmer** case, a farmer professes the belief that he is the Pope, which is irrational given his evidence. The doxastic view has a straightforward way of explaining **Irrationality**: the agent is holding an irrational belief.

However, traditional doxastic views have weaknesses, particularly in their response to **Double Bookkeeping** and **Evidence-Resistance**. For example, if the patient in **CIA-Hospital** believes that the CIA is waiting for him outside the hospital, and we accept that beliefs constitutively guide action, then we cannot explain the man's decision to leave the hospital to buy ice cream. If the farmer in **Pope-Farmer** has overwhelming evidence that he is not the Pope, and we accept that beliefs respond appropriately to evidence, we cannot explain why this belief persists in the face of conclusive evidence against it.

Non-doxastic views aim to solve these problems. These accounts encompass a range of views, some viewing that delusions are attitudes toward representations or imaginings (e.g., Currie & Ravenscroft, 2011), that an agent's having delusions indicates that they are occupying multiple different realities (e.g., Sass, 1994, Gallagher, 2009), or that advocate for different perspectives entirely (e.g., Berrios, 1991 suggests that delusions are empty speech acts). Since these views hold that delusions are not genuine beliefs, they can account for **Double Bookkeeping**. Under a non-doxastic account, the patient in **CIA-Hospital** leaves the hospital to get ice cream shows that he does not genuinely believe that the CIA is outside the hospital. The agent might be alternating between two or more experiential realities, one in which the CIA is outside the hospital and another in which the CIA is not outside the hospital.

The multiple reality view could explain **Evidence-Resistance** as well. Counter-evidence may be ineffective when the agent is in one reality but effective when that agent is in another reality. Or, alternatively,

the delusion is an imagining that the subject mistakes as being real (absolutely), making counter-evidence ineffective. A notable strength of non-doxastic views is that they can incorporate the impact that emotions may have in the formation of delusions, which traditional models of (at least knowledge-apt) belief formation generally do not accommodate. The multiple reality view, for example, allows for the possibility that an agent's emotional distress can contribute to the transition between one reality to a delusional one.

However, these non-doxastic views also have drawbacks. While the view that delusions are imaginings rather than genuine beliefs explains why **Double Bookkeeping** can occur, it cannot so easily explain when and why an agent acts in accordance with their delusion. In **False Confession**, that the man genuinely believes he has committed the crime helps explain why he turns himself in to authorities. But if delusions are imaginings, we need to explain why he holds steady to this confession as he would for a belief. Alternatively, if delusions result from an agent oscillating between multiple realities, we need to explain, in a systematic way, what governs the transition between these realities. Put differently, if we abandon doxasticism, we need an alternative account of the belief-like features of delusions.

The view I introduce in this paper aims to revise the doxastic view so that it can readily incorporate the intuitions that non-doxastic views preserve. That is, I think that traditional doxastic views are ill-equipped to handle challenges like **Double Bookkeeping** and **Evidence-Resistance**. Thus, I will develop a modified doxastic view which aims to reassess how beliefs are formed and how they function, illuminating properties that are intuitive for both ordinary beliefs and delusions. Specifically, I propose a view of delusions that can account for more than just errors in evidence-based reasoning in the formation and maintenance of delusions. In addition, I propose a scheme in which beliefs are functionally compartmented (into implicit and explicit compartments), which I argue gives rise to many of the features of delusions that traditional doxasticism fails to address.

I begin by clarifying the challenge that **Double Bookkeeping** poses to the doxastic account of delusions (section 2). I then propose a novel doxastic view to meet this challenge. I draw on Susanna Rinard's work in epistemology to defend the idea that non-evidential considerations can motivate belief formation (section 3), which is central to my doxastic view. I also propose that beliefs (and thus delusions) operate between two functional compartments, the implicit and the explicit (section 4). My view has the advantage of rescuing the standard doxastic view from **Double**

**Bookkeeping** and **Evidence-Resistance**, offering a better foundation from which to explain how delusions are formed and interact with other beliefs.

## 2. Double Bookkeeping's Challenge to Doxasticism

Let us consider how **Double Bookkeeping** can be used to challenge doxastic accounts of delusions.<sup>5</sup> Some people with delusions fail to act concordantly with their delusions. The case studies above display this feature: the patient who claims the CIA is waiting outside to kill him leaves the hospital to get ice cream (in **CIA-Hospital**), the man who claims he raped his girlfriend finds no support for this claim from his girlfriend (in **False Confession**), and the self-proclaimed Pope still toils in the field like a farmer (in **Pope-Farmer**). If subjects constitutively act in accordance with their beliefs, delusions cannot be beliefs. Bortolotti (2010) has offered the following helpful argument against the doxastic view based on **Double Bookkeeping**:

P1. It is constitutive of beliefs that they guide action in their relevant circumstances.

P2. Some delusions fail to guide action in their relevant circumstances (**Double Bookkeeping**).

C. Some delusions are not beliefs.

The non-doxastic argument from **Double Bookkeeping** against the traditional doxastic view seems, *prima facie*, to hold. To make matters worse, the doxasticist has no compelling explanation for **Evidence-Resistance**. In the next section, we will consider a view of belief formation that can help us both to (immediately) respond to **Evidence-Resistance** and set us up to respond to the argument from **Double Bookkeeping** (section 4). Various authors have addressed these objections to the doxastic view in a variety of ways. Bortolotti (2010) attempts to demonstrate that **Double Bookkeeping** is not clearly a problem for *all* delusions (since many people with delusions do, in fact, act in accordance with them). Bortolotti further argues that ordinary beliefs are not always action-guiding, and thus an action-guiding constraint on delusions fails as a disanalogy between delusions and beliefs. While much more could be said about Bortolotti's

<sup>5</sup> It is important to note that this objection is not the *only* potent concern with the doxastic view of delusions, but some (see Frankish, 2009) have noted it as one of the greatest challenges to the view (**Evidence-Resistance** is another challenge to the view which will be addressed in this paper).

response to **Double Bookkeeping**, I do not approach her work in-depth here. For the moment, it is enough to say that I think Bortolotti's response to **Double Bookkeeping** is plausible, but not decisively convincing in addressing the anti-doxastic challenge. Bortolotti's argument appeals to two kinds of deficits in agents with delusions: those of rational agency (i.e., failing to act in accordance with beliefs) and negative symptoms of psychotic disorders (e.g., avolition, global affective disturbances). However, once **Double Bookkeeping** is understood as not merely a failure to act but the presence of quite odd actions given the delusion (e.g., tending to the fields even though I claim that I am the Pope), Bortolotti's appeal to rational agency becomes less potent: failing to act on a delusion is one thing, while acting *inappropriately* given a delusion is quite another. Similarly, the negative symptoms that Bortolotti appeals to will fail to account for this sharper formulation of **Double Bookkeeping**.

### 3. Introducing Equal Treatment for Belief

Rinard (2019) provides a compelling account of non-evidential belief formation. Central to Rinard's view is the idea that pragmatic considerations (which I call, more broadly, non-evidential considerations) can rationally motivate belief formation. A version of this idea will support our new doxastic model of delusions. On this view, the doxasticist has a response to **Evidence-Resistance**: beliefs motivated by evidence can be expected to respond to evidence, while beliefs not motivated by evidence (like delusions) may not be expected to readily respond to evidence.<sup>6</sup> But first, let us discuss what non-evidential considerations are.

#### 3.1. Non-evidential considerations, adaptivity, and maladaptivity

Non-evidential considerations motivating belief formation is not uncommon. For example, consider someone who finds no solid evidence supporting religious belief but still believes in God because of the emotional comfort that religious belief provides. Rinard offers another example. A patient has terrible chronic pain but believes that acupuncture treatments will encourage the flow of raw energy through her body and relieve her pain. Incredibly, she finds that these acupuncture treatments *do* work to relieve her pain. However, the acupuncture treatments would cease to work

<sup>6</sup> As will become clear from my proposal in section 4, I do not think that this response to **Evidence-Resistance** is the most satisfying one available to doxasticists. However, if the reader is convinced only by section 3 but not section 4, this response to **Evidence-Resistance** (or some closely related variant of it) will be the best available.

if she stopped believing that they promoted the flow of energy through her body (like a placebo effect).

Consider a further case of Rinard's, in which a man has a mysterious illness. His chance of surviving the mysterious illness can only be increased by believing he will survive the illness. Suppose his chance of surviving the illness without believing he'll survive is 25% (let's say the mortality statistics point to this figure). But, if he believes that he will survive, his chance of survival is 45%. In this case, Rinard argues that the patient *should* develop the belief that he will survive, even though the evidence suggests that he will die. Though he will be arguably mistaken about the reality of his condition, he will increase his chances of surviving.<sup>7</sup>

Importantly, however, for the purpose of understanding delusions, we are not interested in what one *should* believe in the guidance-giving sense that Rinard describes. Though it may be a requirement of developing delusional states that delusions are at some point consciously adopted (see Frankish, 2009), whether a person *should* adopt a delusion or not is beyond the scope of this discussion. It is enough for our purposes to understand that non-evidential considerations can and do motivate belief. Further, I will suggest that instances of non-evidentially motivated belief formation can be categorized as adaptive or maladaptive.

While Rinard illustrates cases in which an agent evaluates non-evidential considerations and develops beliefs accordingly (e.g., the case of the acupuncture patient), I also think the opposite is possible: an agent may adopt a belief by *inappropriately* evaluating non-evidential considerations. Here's what an example of an inappropriate evaluation of non-evidential considerations might look like. Recall the case of the patient whose chance of surviving the mysterious illness will increase if he believes he will survive. Suppose another study shows that people with the illness who believe their disease is contagious have only a 1% chance of surviving. Suppose further that there is no evidence suggesting that the disease is contagious. If the patient still adopts the belief that his disease is contagious, he will have adopted a belief that is ultimately harmful for him.

Since a belief like the belief in the mysterious disease's contagion is not formed by any (mis)interpretation of relevant evidence, I will not call such a belief *irrational*, but rather *maladaptive* (though it is possible

<sup>7</sup> Crucially, Rinard does not claim that non-evidential considerations always override evidential ones. In fact, she notes that it is common for evidential and non-evidential considerations to have a consensus on what one should believe. After all, believing what evidence suggests is *usually* beneficial and useful in leading an ethical and fulfilling life. The relevant takeaway for our discussion is this: instead of focusing solely on evidence as motivation to believe, we should also tender non-evidential considerations.



that some belief can be *both* irrational *and* maladaptive. This would be the case *both* if the belief was contrary to available evidence *and* if adopting the belief led to harm for the agent or others). Maladaptive beliefs are beliefs which yield a net harm to the agent or work contrary to the agent's practical interests. Analogously to how irrational beliefs are (often) formed from a poor interpretation of evidence, I think maladaptive beliefs are (often) formed from a poor interpretation of non-evidential considerations.

Though I have chosen the word "maladaptive," I do not mean that these beliefs are maladaptive strictly in an evolutionary sense. Maladaptivity, as I am using it here, is context-dependent; an adaptive belief in one context may be a maladaptive belief in another. In this sense, the maladaptivity of a belief might be best attributed by a clinician who is aware of the believer's social and cultural context. That is, maladaptivity may be thought of as a clinical determination, which emphasizes how the belief by itself can be innocent or nondetrimental, but the surrounding social circumstances can problematize or vindicate the belief. A detrimental belief, thus, might be one that (within a sociocultural context) leads to difficulty completing activities of daily life, prohibits the formation of healthy relationships, motivates harm to oneself or to others, etc. In other words, the sorts of features that would make a belief maladaptive are those presented in common psychiatric practice as distressing (see the diagnostic codes for delusions in American Psychological Association, 2013).

Yet, our account also accommodates varied attributions of adaptivity and maladaptivity for the same belief given shifts in external circumstances. Here's an example. Suppose I am a doomsday preparer who believes that a zombie apocalypse is occurring. I pour nearly all my income into buying weapons, building a bunker, and acquiring supplies for surviving the apocalypse. I am driven to isolation as a result, and my psychiatrist evaluates my obsession as unhealthy. In this case, my belief that there is a zombie apocalypse is maladaptive. Now, let's say that a few days later, zombie apocalypse actually *does* occur. Suddenly, I am more prepared than anyone else and my stream of weapons and supplies provide me a survival advantage. Further, getting used to social isolation prepared me emotionally for the tolls of the zombie apocalypse. Now, the same belief that was maladaptive before has become adaptive.<sup>8</sup>

Non-evidential considerations are, as I have laid them out so far, a broad category of factors. While they exclude evidential considerations (e.g., chances of an event obtaining based on experience), I take them to include pragmatic motivators (i.e., considerations that recommend taking

<sup>8</sup> I thank Ram Neta for this example.

the action that leads to the best practical outcome) and *emotional* factors that go into belief formation. Emotions have been shown to play a role (sometimes accuracy-increasing and sometimes accuracy-decreasing) in conclusions we reach through logical reasoning (Jung et al., 2014). More relevantly, *emotional reasoning*, or using emotions rather than evidence to form beliefs, is a trait associated with certain mental disorders (like General Anxiety Disorder) under cognitive-behavioral models (Berle & Moulds, 2013). (Whether emotions ought to be included in accounts of *rational* belief is a different discussion that I do not approach here.)

Thus, in my view, a delusion is constitutively both irrational and maladaptive (i.e., it is harmful to the agent's functioning in a social and cultural context). Further, a delusion is constitutively motivated by non-evidential considerations (which is a property of some ordinary beliefs as well). My view, therefore, suggests an *evaluative* component to delusion attribution (i.e., the delusion must be evaluated as irrational and maladaptive) and suggests an *etiological* component (i.e., the delusion is formed for non-evidential motivations).

### 3.2. *Evidential considerations, rationality, and irrationality*

I have been intentionally noncommittal on exactly what makes a belief rational since this problem is a subject of controversy. For the purposes of this discussion, I will say that an agent has a rational belief at time  $t$  if the agent's belief is at least more supported than unsupported by the evidence available to the agent at time  $t$ . For example, if I (who knows that fair six-sided dice have six equiprobable outcomes) believe the chance of a fair six-sided die landing on the number "4" is 100%, my belief is irrational.<sup>9</sup>

<sup>9</sup> When I talk about considerations for rational, irrational, adaptive, and maladaptive beliefs, I am talking specifically about the considerations that *motivate* the agent's adopting the belief. In other words, it is not enough that some consideration merely causes a belief to be adopted. Consider an example from Rinard (2018). Imagine you're speeding down the highway and then see a cop, so you slow down your car. Pressing the brake pedal is the cause of your slowing down, but it isn't *why* you slowed down. Slowing down because you don't want to get a ticket is the kind of motivation I am interested in when talking about the rationality and/or adaptivity of a given belief. Adapted for belief, Pascal's wager cases serve as an analog for this idea. Suppose an atheist is moved by some immense practical benefit in favor of believing in God, deciding she wants to revise her disbelief and believe in God. This person might then choose to expose herself to certain books, communities of people, or events that will help her selectively acquire evidence to believe in God. The evidence she gains is a mere cause of her belief in God but it is not *why* she now believes in God. Her belief in God is caused by certain kinds

Unlike Rinard, I distinguish between evidential considerations for beliefs (which are evaluated by their rationality or irrationality) and non-evidential considerations for beliefs (which are evaluated by their adaptivity or maladaptivity). She would subsume what I call adaptive beliefs under the category of rational beliefs. In other words, a belief I call *either adaptive or rational* (in the next sub-section, I will propose a special term for these kinds of beliefs: *orthodox*), Rinard would call them *rational*.<sup>10</sup> I do not at all object to her naming convention but find it useful to bifurcate evidential and non-evidential considerations to provide more granularity for the discussion that follows.

One last deviation from Rinard is worth mentioning. Whereas Rinard discusses what beliefs an agent *should* have (i.e., what beliefs an agent is rationally permitted to have), my argument is not concerned with normative epistemology. Rather, I am interested in accessing a more psychological account of what beliefs it is possible for an agent to have and how those beliefs might be categorized as rational, irrational, adaptive, and/or maladaptive based on the considerations by which they are motivated.

### 3.3. *Orthodox and delusional beliefs*

With these clarifications in place, we can diagram our belief categories (see **Figure 1**). The rows of the grid represent evidential considerations, where “rational” means that the adopted belief accords with available evidence (and “irrational” means that the belief does not). The columns represent non-evidential considerations, where “adaptive” means there is a net practical benefit in adopting the belief (and “maladaptive” means that there is a net practical loss).

The figure represents cases in which *both* non-evidential *and* evidential considerations are at play. For example, imagine a person has a headache for most of the day, so they decide to take some pain medication. The belief at work here is that taking the medication will alleviate their pain. A belief like this one would occupy the top right cell (and would thus be orthodox) because it is motivated by both rational evaluation of evidence

---

of evidence but motivated by her perceived pragmatic advantage in believing in God. It is this second kind of motivation, the *why*, that engenders what I mean by “based on” or “non-evidential considerations for belief.”

<sup>10</sup> Rinard writes that she is “doubtful that there is a special epistemic sense of rationality—I suspect this is a philosophers’ invention” (Rinard, 2019). It seems to me that she is broadening the definition of the word *rational* to include non-evidential considerations.

		Non-Evidential Considerations	
		Maladaptive	Adaptive
Evidential Considerations	Rational	Orthodox	Orthodox
	Irrational	Delusional	Orthodox

**Figure 1: Evidential and Non-Evidential Considerations for Belief Formation**

(taking pain medication has helped with pain in the past) and adaptivity (having this belief facilitates the alleviation of pain).

*Orthodox* beliefs occupy three out of the four cells of **Figure 1**. I take these beliefs to be *either* rational (i.e., according with the agent's current evidence) *or* adaptive (i.e., they afford a net practical benefit to the agent). Thus, an irrational belief is orthodox if and only if it affords a net practical benefit (i.e., is adaptive). The last cell consists of (candidate)<sup>11</sup> *delusional* beliefs, which are *both* irrational *and* maladaptive. I sometimes use "delusion" to refer to these beliefs, but I will often prefer "delusional belief" when comparing the belief in question to an "orthodox belief."

A consequence of my view is that there is no such thing as an adaptive delusion. All beliefs that are truly adaptive (despite what the evidence might say about them), are orthodox beliefs on this view. I don't think that this commitment is nearly as problematic as it may appear at

<sup>11</sup> Of course, not *all* irrational and maladaptive beliefs are clinically designated as delusions (e.g., sexist beliefs). But this combination of the irrational and the maladaptive is what is encapsulated by the necessary condition **Irrationality**. Thus, for the purpose of this discussion about delusions, I will refer to beliefs that are both irrational and maladaptive as delusional beliefs.

first glance. As Rinard herself mentions, many adaptive beliefs are adaptive in virtue of their being true (or, at least, supported by the evidence) since it is, often, adaptive to believe what is supported by your evidence. The exceptions to this principle are cases such as the mysterious illness (where believing, against evidence, that you will survive the disease, is adaptive). On my view, beliefs like those are not delusions, precisely because they are adaptive.<sup>12</sup>

We now have new tools with which to update the **Irrationality** condition of delusions from section 1. Below, I include the updated **Irrationality** condition, which reflects the foregoing about non-evidential considerations for belief and reproduces the other two features. Recall that none of these conditions are sufficient conditions, nor do they altogether constitute sufficient conditions for classifying a belief as a delusion. Psychiatric qualifications, including distress to the agent, the level of dysfunction the delusion causes, persistence of the delusion over time, and more, must be made. This paper is concerned with epistemological components and conditions for delusion classification, not diagnostic criteria for a delusion.

**Irrationality\*:** Delusions are *both* irrational (conflict with available evidence) *and* maladaptive

**Double Bookkeeping:** Delusions need not lead to action concordant with their content

**Evidence-Resistance:** Delusions are resistant to counter-evidence

### 3.4. Sensitivity of beliefs to counter-evidence

To round out this notion of non-evidentially motivated beliefs and evidentially motivated beliefs, we may want to say more about **Evidence Resistance** in terms of a belief's susceptibility to counter-evidence. Suppose a young married couple discovers that 60% of couples similar to them in demographic and educational factors have marriages that end in divorce (and that this couple is averse to divorce). Since they are more likely to divorce than not, they may believe that they will divorce. Alternatively, they may believe, against evidence, that they will not divorce,

<sup>12</sup> Even if a patient thinks that his delusion is adaptive (or if it is the case that the delusion is the most adaptive option that he can identify), according to my view it must be maladaptive because it is a delusion. A clinician could find a more adaptive alternative for the patient (rather than believing the delusion). So, perhaps the agent is not blameworthy for believing the delusion (since it is the most adaptive option he can identify), but that does not mean the delusion itself is, all things considered, adaptive.

which may be an adaptive belief (if this belief decreases their likelihood of divorcing). Imagine this couple believes, against evidence, that they will remain married.

Suppose further research emerges, suggesting that 65% of couples like them end up divorced. Would this couple, already believing against evidence, begin to believe in a manner motivated by evidence? Perhaps. However, the burden of proof, it seems to me, rests on the one claiming that an agent presently believing for non-evidential motivations would cease in this motivation and accept evidence as the grounds for a revised belief instead, inconsistent with their previous belief. Much more likely, I think, is that the couple continues to believe, against all evidence, that they will remain married.

Perhaps more extreme situations reverse the intuition. Say that even newer research suggests that 99% of couples like this couple end up divorced. *This* counter-evidence may shake the couple's belief in their continued marriage. But what would separate this case from the previous one in terms of the likelihood of belief revision?

I think the answer to this question rests in the extent to which the agent is convinced of this belief. If the members of the couple strongly believe that their marriage will persist, this belief is more likely to resist counter-evidence. This relationship seems intuitive: The more strongly an agent believes  $p$  for non-evidential reasons, the less likely they are to begin believing  $\sim p$  for evidential reasons. In the example of this married couple, we would expect that they would not revise their belief that they will remain married in the face of counter-evidence suggesting a 65% chance of divorce but may revise their belief in the fact of counter-evidence suggesting a 99% chance of divorce. However, if they were strongly convinced in their staying married, even the stronger counter-evidence may not be enough to lead to belief revision.

A constitutive feature of delusions is the conviction with which they are held. We might think that delusions are held with (at least close to) the maximal conviction with which a belief can be held. Thus, for the purpose of this discussion on delusions (since these are, by my account, beliefs held with strong conviction), delusions motivated by non-evidential considerations are resistant to counter-evidence.

#### 4. Implicit and Explicit Compartments

Let's take stock of the discussion so far. In my view, delusions are beliefs. I have further argued that beliefs (and thus delusions) can be motivated by non-evidential considerations and that the specific type of

beliefs that are delusions both conflict with available evidence and are maladaptive to the agent. This distinction allows us to appreciate the deeper level of error that agents with delusions have. If I am a delusional agent who comes to believe that I am the Pope, my error is not simply that I have made a poor assessment of my available evidence—I have also mistaken a pragmatic or emotional factor (e.g., I feel that I am more important or powerful than people acknowledge) as motivation for the belief that I am the Pope, ultimately to my detriment.

As described earlier, this distinction helps us explain **Irrationality** in more complete terms by appealing to evidential and non-evidential dimensions of delusion formation. Additionally, the use of non-evidential considerations helps the doxasticist respond to the challenge from **Evidence-Resistance**. Since delusions are motivated, at least in part, by non-evidential considerations, counter-evidence will be ineffective in reversing a delusion.<sup>13</sup> However, to explain **Double Bookkeeping**, we also need to understand how an agent can simultaneously hold two contradictory beliefs. The goal of this section is to outline a compartmentalization scheme that refines our new doxastic model, explaining **Double Bookkeeping** along the way.

I propose that there are two functional compartments of belief: the explicit and the implicit. **Double Bookkeeping** arises from these two different compartments of belief possessing inconsistent beliefs. I suggest that which compartment (either the implicit and explicit) possesses which belief (the orthodox or the delusional) at one time gives rise to many of the features of delusions. Additionally, as the examples in section 4.2 will suggest, the ability to hold contradictory beliefs across compartments is not exclusive to agents with delusions. Rather, it is a feature of ordinary cognition.

#### 4.1. *Inconsistent beliefs in agents with delusions*

Ordinary agents adopt inconsistent beliefs. One possible explanation for how this happens is that there are *fragments* in the mind. Lewis (1982) offers an example of fragmentation. Suppose Lewis believes (1) Nassau Street runs roughly from east to west. He also believes (2) that a nearby

<sup>13</sup> Refinements to this claim (which will be described further in subsection 4.1.2) will depend on the reader accepting the compartmentalization scheme I propose in this section. In other words, I think that a doxasticist who accepts only non-evidential belief formation (section 3) can, at best, provide a basic response to **Evidence-Resistance**. By contrast, a doxasticist who accepts both non-evidential belief formation and my compartmentalization scheme (section 4) can provide a satisfying response to **Evidence-Resistance** and **Double Bookkeeping**.

railroad runs roughly from north to south. He also believes (3) that Nassau Street and the railroad are roughly parallel. The first two beliefs are inconsistent with the third, but the inconsistent triad is never in view to him all at once. Perhaps when he is on Nassau Street, he brings (1) and (2) to mind. Maybe when he is discussing the roads and railroads with a friend, he endorses (3) and (1).

Lewis' scheme isn't the only way to conceptualize fragmentation. To mention a few alternatives, we might posit, as Gendler (2008) does, that there is some distinction between belief and "alief" which explains features of our mental states. Or we might distinguish between "implicit" and "explicit" fragments, as Greco (2015) appeals to. My scheme calls for a distinction between implicit and explicit *compartments* (rather than fragments<sup>14</sup>) in which an agent holds beliefs.

Here's a bird's-eye view of how my compartmentalization scheme works. Every belief  $p$  that an agent  $S$  has is in the implicit compartment, in the explicit compartment, or in both compartments. If  $p$  is in the implicit compartment,  $S$  is disposed to treat  $p$  as true in unreflective action. If  $p$  is in the explicit compartment,  $S$  is disposed to endorse  $p$  and defend  $p$  in argument.<sup>15</sup> It is possible that  $p$  is in both compartments, such that  $S$  is both disposed to treat  $p$  as true in unreflective action and in argument. It is also possible that  $p$  is believed in one compartment while  $\sim p$  is believed in the other.

Before saying more about what these compartments are, I will emphasize what these compartments are *not*. I do not assume that implicit and explicit compartments necessarily reflect a *structural* neurobiological reality. Rather, I think of these compartments as a way to motivate intuitive *functional* properties of beliefs. These properties include our ability to have implicit action based on unconsciously formed beliefs about the world, the occasional discordance between our professed beliefs and our actions, and our tendency to quickly form beliefs without the careful consideration of evidence. As such, implicit and explicit beliefs are not meant to suggest any particular structures of the mind, imply that beliefs are stored in any specific kind of network, etc. Rather, implicit and explicit beliefs point to two seemingly different functional profiles of beliefs: Those on which we unreflectively act (implicit beliefs) and those which we genuinely endorse and are disposed to defend in argument (explicit beliefs).

<sup>14</sup> I use the terms "compartmentalization" and "compartments" rather than "fragments" (as Egan, 2008, Lewis, 1982, and others use) to emphasize that any given belief an agent holds will fall into one of the two compartments (implicit or explicit) or both.

<sup>15</sup> My description of the implicit/explicit compartments leans heavily on Greco's (2015) description of fragments.



Also, I do not claim that the division between implicit and explicit beliefs is novel. Notable examples of comparable projects include distinctions between “System 1” and “System 2” (see Kahneman, 2011) and the distinction between “aliefs” and “beliefs” (see Gendler, 2008). I propose my own distinction between implicit and explicit compartments, not because either of these two distinctions appear particularly problematic or flawed to me, but because they are both more ambitious (i.e., they work to explain more phenomena) than I need for the purpose of characterizing delusions. For example, System 1 describes a broad range of psychological phenomena, including visual processing, a wide range of heuristics, accompanying physiological sensations, etc. The System 1/System 2 division does much more explanatory work than I need for this discussion. Therefore, I will make a less-revisionary, independent characterization of implicit and explicit compartments. Likewise, implicit beliefs need not be Gendler’s aliefs. For while there can be exchange of beliefs between implicit and explicit compartments, Gendler may not be committed to there being an exchange of content between aliefs and beliefs (at least not in the same way I do for implicit and explicit beliefs, as I describe in subsection 4.1.4).<sup>16</sup>

The upshot of the foregoing is that while I think either System 1/System 2 or the alief/belief distinction could be adapted to do the work that I have in mind for implicit/explicit beliefs, I refer to the implicit/explicit distinction in this paper to provide my account of the functional behaviors of beliefs. Note further that the implicit and explicit distinction is not the same as a dispositional and occurrent belief distinction. One can intuitively have explicit beliefs that are either dispositional or occurrent, whereas one probably has only dispositional implicit beliefs. Though, if one can have occurrent implicit beliefs, this is no worse off for my claim that dispositional/occurrent beliefs are not the same as implicit/explicit beliefs.

#### *4.1.1. Implicit beliefs reduce cognitive load*

We demand a lot from our minds. A great deal of our daily activities and meaningful relationships depends on our ability to form (roughly) accurate beliefs. Moreover, we must be able to form these beliefs *quickly* if

<sup>16</sup> In addition, there are challenges concerning whether the alief category varies meaningfully from belief or aliefs it can do the explanatory work Gendler has in mind for them (see Mandelbaum, 2013) for which I do not have full responses. Accordingly, I opt to avoid using aliefs to describe orthodox and delusional mental states. Since I hold that implicit and explicit beliefs are both the same kind of state (i.e., they are both beliefs, just sensitive to different contexts and considerations), I take it that my distinction is free from the concerns that Mandelbaum raises for aliefs.

we are to adapt appropriately to rapidly changing conditions. These beliefs seem to come from instincts that need not be validated by experience. A man who has never encountered or heard of a tiger may immediately believe, upon encountering a tiger, that it is dangerous (though he may not be able to tell why he believes it to be dangerous). We should think that belief in  $p$ , “This tiger is dangerous” is a genuine belief (rather than *only* an instinctual push toward safety or some other kind of non-belief state) because the subject is disposed to use it in reasoning, and it provides an answer to the question “Is this tiger dangerous?” If the man is forced to choose between running into a forest where the tiger is or take a different route to get home that avoids the tiger, his belief  $p$  will be used to decide that chancing an encounter with the tiger is more dangerous.

I think the implicit compartment handles most of these types of beliefs, usually (but not constitutively) without our awareness. If we had to actively engage in empirical observation, induction, or deduction to form each and every one of our beliefs, we would not have a cognitive arrangement compatible with survival, much less the demands of lively society. We are constantly forming beliefs about whom we trust, distrust, like, think likes us, and more.

#### 4.1.2. *Implicit beliefs are more sensitive to non-evidential considerations*

In exchange for this efficiency, we trade accuracy in several cases. Our belief in the danger of tigers may be formed automatically so that we can quickly avoid one in the wild. But what about when we see a tiger in an enclosure? Most people would probably be (at least a little) afraid if they came face-to-face with a tiger, even if separated by glass. If you asked an agent in this situation, “Are you in danger?” they may respond, honestly, “I am not in danger because of the glass: but somehow I still feel afraid.” The left and right sides of the colon demonstrate the explicit and implicit beliefs, respectively. The explicit belief, formed on consideration of the evidence, holds that the person is safe, protected from the tiger by the sturdy glass. The implicit compartment, on the other hand, holds  $p$ , “This tiger is dangerous” in such a way that is *resistant* to the evidence that there is sturdy glass between himself and the tiger. This belief is not evidentially-motivated (insofar as emotions do not count as strong epistemic reasons for beliefs).

This result should appear intuitive, given the general features of the implicit compartment: While explicit beliefs are genuinely endorsed and defended arguments (suggesting that there is conscious evaluation of evidence to arrive upon or revise explicit beliefs), implicit beliefs are not

as sensitive to evidence given that they are formed from more immediate mechanisms related to navigating environments (or so I suggest). Implicit sexist (who are explicit anti-sexists) have excellent reasons to be anti-sexist (which they genuinely endorse), so the persistence of implicit sexism in explicit anti-sexists is puzzling unless the consideration that implicit beliefs are less sensitive to evidence is accepted. The same holds for the tiger behind the glass: The zoo-goer has strong evidence for believing that the tiger cannot harm them, and their implicit attitudes toward the tiger are difficult to explain without appeal to implicit beliefs' decreased sensitivity of evidence.

The benefit of the implicit compartment is that, in the wild, the person facing the tiger would decide immediately to attempt an escape from the tiger, not wasting valuable time -considering anecdotes and statistics about tiger attacks to construct an evidentially motivated judgment about the danger of the tiger. The drawback is that, since the belief was not formed from evidence, it is resistant to competing evidence (recall the discussion in subsection 3.4., concerning resistance to evidence of non-evidentially motivated beliefs). This reasoning helps motivate **Evidence-Resistance** in delusions.

#### 4.1.3. *Explicit beliefs are more sensitive to evidential considerations*

As you are reading this paper, considering the arguments and examples within it and connecting it to your existing beliefs about beliefs and delusions, you are utilizing explicit beliefs. These are the beliefs that you would be disposed to defend in argument (e.g., if you disagree with the argument of this paper and I asked you about your own view, you would be articulating explicit beliefs).

Further, I think there are some ordinary cognitive states that are explainable through the explicit compartment's ability to respond to evidence and the implicit compartment's resistance to evidence. Imagine I have no evidence to suggest that my partner is cheating on me. Maybe I have evidence that suggests her *loyalty* (there are no unexplained disappearances, she communicates regularly and honestly when I voice concerns, etc.). Still, despite the evidence (in casual conversation, sometimes people say "the rational side of me says that everything is okay"), I might feel unconvinced that my partner is loyal. It seems possible that this state occurs because I have an explicit belief (based on my evidence) that my partner is loyal and a competing implicit belief (based on my emotions) that my partner is cheating.

#### 4.1.4. Both compartments are “leaky”

While I have claimed that implicit beliefs are more sensitive to non-evidential considerations and explicit beliefs are more sensitive to reason and evidence, I do not mean to say that this bifurcation is absolute. I see no reason why non-evidentially motivated beliefs cannot be recognized by the agent and defended in argument like other explicit beliefs.

I also think that beliefs can be exchanged from one compartment to the other, though they must be “translated” in some way. For example, I may develop an implicit belief that tigers are dangerous based on only one encounter with a tiger (not even having heard of a tiger until that encounter). If asked about the experience by a friend, I might attempt to rationalize the belief in several ways (e.g., “Its teeth look like they were designed to kill me, its posture signaled aggression”), effectively reproducing the implicit belief in the explicit compartment. Note that I formed the explicit belief *post hoc* because I consciously thought of reasons as to why the tiger was dangerous only *after* the encounter. The reasons I devise need not be perfectly accurate—they just have to be the kinds of things I can use to argue for my belief.

Conversely, I think the explicit compartment can be efficacious on the implicit compartment, though I think there must be a similar “translation” or “training” involved. Imagine that I have arachnophobia and believe  $p$ , “Spiders are deathly dangerous.” I am disposed to defend this belief in argument and to demonstrate my belief through unreflective action (screaming, cursing, and running away from spiders when they are presented to me). It appears that belief  $p$  is in both my explicit and implicit compartments.

Imagine I see a therapist who eventually convinces me that I am overestimating the dangerousness of spiders. I come to accept the therapist’s view and revise my explicit belief in  $p$  to  $\sim p$ . Now, when asked by a friend, I argue that spiders are not deathly dangerous and that I don’t believe I would die if I touched one. Still, most of the symptoms of my phobia have persisted. I still scream, curse, and run away from spiders when they are presented to me. My implicit belief is still in  $p$ . To help me, my therapist suggests we undergo systematic desensitization. I start by thinking about spiders, graduate to imagining touching a spider, then I progress to seeing a spider from a safe distance, and then end my journey by finally touching a spider under the supervision of my therapist. I have not gained any new evidence,<sup>17</sup> but my implicit fear of spiders has

<sup>17</sup> While, through the experience of touching the spider, I have seen that touching

subsided. My belief in  $\sim p$  is now represented in my implicit compartment as well.

#### 4.1.5. *Compartments can have duplicate or contradictory beliefs*

As suggested in the previous example (and as will be presented in future examples), the explicit and implicit compartments may both have some belief  $p$ , or one might have  $p$  while the other has  $\sim p$ . The status of each compartment toward a delusional belief  $p$  (and what each of these states may correspond to for a person occupying them) will be discussed in subsection 4.3.

### 4.2. *Compartments in action*

In the following subsections, I discuss two examples that further illustrate the explicit and implicit compartments, their interactions with evidence, and their interactions with each other.

#### 4.2.1. *Implicit and explicit compartments: Alex the implicit sexist*

Consider a case of Alex the implicit sexist, the details of which I draw heavily from Schwitzgebel (2010) and Greco (2015). In private conversations and public statements, Alex, a professor, is an avowed anti-sexist. He is prepared to argue, sincerely, for the equality of intelligence between genders and has done so in the past. Yet, at the same time, it is clear to others that he is biased against women in multiple contexts. In the classroom, he consistently grades women's papers more harshly than men's (but equally when he is unaware of the gender of the authors). When he serves on hiring committees, it never appears to him that the women are as bright as the male candidates. At the same time, Alex seems committed to fighting his implicit bias. Perhaps he reminds himself to be more open-minded and interpret women's comments in his class more generously when he would usually be uncharitable. When he witnesses sexist behavior from his colleagues, he is quick to correct them, even when he worries there may be social consequences for doing so.

The upshot of this characterization is that there is no one clear answer to the question "Is Alex sexist?" There are many reasons to think

---

a spider doesn't mean I will die on the spot, I was already willing to defend that in argument before. In other words, the belief that spiders are not deathly dangerous was already in my explicit compartment.

that he is, as well as some reasons to doubt that he is *entirely* sexist. We might say that in the explicit compartment, Alex holds  $p$ , “Women are equally capable as men.” In his implicit compartment, either as a function of socialization, out-group bias, etc., Alex holds  $\sim p$ .

As noted earlier, I think that the two compartments can be efficacious on each other. If Alex is rigorous enough in challenging himself to treat women charitably, he may eventually begin to unreflectively treat his students equitably. It would seem, then, that both circuits would possess the belief  $p$ . Such revision demonstrates the efficacy of the explicit compartment on the implicit compartment. Conversely, if he is confronted about his biased behavior and rationalizes his behavior, he transfers an implicit compartment belief to the explicit compartment, so both compartments contain  $\sim p$ . Alternatively, he may continue with his explicit compartment containing the one belief,  $p$  (that women are equally capable as men) while his implicit compartment contains the opposite belief,  $\sim p$ .

#### 4.2.2. *Implicit and explicit compartments: Driving to university*

Imagine I am driving to my university. I live about one hour away from campus, in a suburban town. As a result, I must get on and off a couple of freeways and make several specific turns. About ten minutes into my journey, I notice a black car with tinted windows behind me. Twenty minutes into my journey, the car is still following me. I take care not to change my route, but as I get onto new roads, I can't help but notice the same black car is still behind me. The thought crosses my mind that the vehicle belongs to an FBI agent, and I am being followed for some reason. After all, this car looks like the ones I've seen on television in crime drama shows. There is a little bit of (questionable) evidence in favor of believing the FBI is following me, and perhaps a few transient feelings of worry bring me to consider the question: “Am I being followed by the FBI?”

At this point, I have considered  $p$ , “The FBI is following me.” As a rational person (at least rational enough for this purpose), I both consider the overwhelming evidence for  $\sim p$  and take that evidence seriously. While it's true I took a very specific route, it's also true that I am headed to a research university with tens of thousands of students, faculty, and staff. The likelihood of my happening to drive to a large university at the same time as someone else (who happens to own a dark car) is, intuitively, much higher than the odds that I'm being hunted by the FBI for some unclear reason. I adopt this orthodox belief in my explicit compartment. If someone were in the car with me and asked, “Do you believe the FBI is following you?” I would defend in argument that the FBI is not following me.

Now imagine that Robert, a young man just like me, is in the same situation. Imagine Robert satisfies the criteria for having prodromal schizophrenia, which means that he has not yet developed intense psychotic symptoms, but based on a combination of social isolation, impairment in role functioning, odd beliefs, magical thinking, unusual perceptual experiences, etc., can be said to have a high risk of developing schizophrenia (George et al., 2017). People who match these criteria may report feeling depressed, distant from others, and distant from reality. Robert, in driving to the university, might consider the same evidence that I do, and he might even reach the same rational conclusion that I do. But the difference for Robert is the increased weight he gives to the non-evidential considerations. Perhaps he implicitly feels some sort of concern for his safety and fear of being followed, and the FBI following him would justify these feelings. After all, if the FBI truly were following him, it might explain the longstanding tension generated by his prodromal symptoms (e.g., feeling chronically disconnected from others, having strange perceptual experiences that others do not have). The delusional belief that the FBI is following him is adopted into his implicit compartment because it resolves this emotional tension.

What I have described here is far from the full story about how agents may adopt delusional beliefs. For one, I have not yet described how Robert's implicit belief could become an explicit belief (it is, after all, when people express their delusions to others that we become aware of their psychopathology). I have also not said anything about how Robert might come to find remission from the delusion by my model. These aspects will be discussed in future work.

Next, I will outline and describe the possible states an agent with delusions can occupy based on the status of delusional belief acceptance or rejection between the two compartments.

### *4.3. Four belief states*

I note four orientations of the delusional and orthodox belief between the two compartments—I take it that each of these are possible states one can be in with respect to a proposition  $p$ . For example,  $p$  may be “The FBI is following me.” The implicit and explicit compartments may each hold a delusional (i.e.,  $p$ ) or orthodox (i.e.,  $\sim p$ ) belief, producing any of the four orientations that follow. An agent with a delusion need not spend equal amounts of time in each state. I aim to demonstrate a set of mutually exclusive and jointly exhaustive configurations a person can have with respect to  $p$  in her implicit and explicit compartments.

- DD.** Implicit, **D**elusional; Explicit, **D**elusional (non-double bookkeeping delusional state)  
**OD.** Implicit, **O**rthodox; Explicit, **D**elusional (double bookkeeping state)  
**DO.** Implicit, **D**elusional; Explicit, **O**rthodox (either prodromal or “normal”)  
**OO.** Implicit, **O**rthodox; Explicit, **O**rthodox (remission or non-delusional)

I now describe some features of the states in turn.

#### 4.3.1. *The (DD) state*

The (DD) state is characterized by delusional belief in both compartments. An agent in this state would be expected to express the delusional belief in declarations and arguments as well as unreflective action. That is why I consider an agent with a delusional belief in the (DD) state to be in a non-double bookkeeping state. An agent in this state would endorse a delusional belief (e.g., “the CIA is waiting outside the hospital to kill me”) while also acting in accordance with that delusion, even when acting without reflection (e.g., not stepping outside the hospital in fear of being killed).

#### 4.3.2. *The (OD) state*

The (OD) state is the **Double Bookkeeping** state, which we set out to characterize at the beginning of this discussion. (The implicit compartment holds the **O**rthodox belief while the explicit compartment holds the **D**elusional belief.) An agent in this state would defend the delusional belief in argument, but not act in accordance with the belief in unreflective action. For example, an agent in this state would endorse a delusional belief  $p$  (e.g., “the CIA is waiting outside the hospital to kill me”) *without* also acting in accordance with  $p$  when acting without reflection (e.g., going outside the hospital to buy ice cream).

Let me make explicit how this model provides us with the tools to explain **Double Bookkeeping**. The explicit compartment contains the delusional belief ( $p$ , that the CIA *is* waiting outside the hospital to kill him), so he defends this belief in argument.<sup>18</sup> However, the agent *also* has the orthodox belief ( $\sim p$ , that the CIA *is not* waiting outside the hospital to kill him) in his implicit compartment, so he unreflectively acts as though  $\sim p$ . This explains how his behavior is at odds with his belief that the CIA wants to kill him.

<sup>18</sup> Examples of such defenses of delusional beliefs can be found in Flores (2021) and Ramachandran & Blakeslee (1998).



### 4.3.3. The (DO) state

The (DO) state is characterized by **D**elusional belief in the implicit compartment and **O**rthodox belief in the explicit compartment. In the earlier example of Robert's drive to his university, he enters the (DO) state once the delusional belief ("The FBI is following me") is adopted to his implicit compartment. I think that the (DO) state is found both in people who would be considered prodromal/delusional by a clinician and those who are not. There are a few differences between these two groups of people, and I wish to say a little as to what may distinguish them.

Recall Alex, the implicit sexist, who holds inconsistent beliefs in the implicit and explicit compartments ("Women are inferior to men" in the former, and "Women are equally capable as men" in the latter). Alex could be said to be in the (DO) state, since he holds an irrational, maladaptive belief in his implicit compartment which is resistant to evidence (while holding an orthodox belief in the explicit compartment). Still, Alex would probably not be diagnosed as having delusions, though he is consistently in the (DO) state. Is this point of discrepancy between the model and clinical designations indicative of a failing of the model?

I do not think so, because the discrepancy between an implicit sexist and a typical psychotic agent of this day and age is due to the social circumstances and norms that demarcate which beliefs are simply troubling as opposed to pathological. In other words, whether (what I call) a delusional belief is clinically recognized as a delusion depends on social and cultural factors, both within a time and a place.

Research suggests that the content of delusions is influenced by social themes and that delusional beliefs may "spread" through groups, whether through mass hysteria (or more specifically, "mass delusion" and "social delusion") or through shared psychotic disorder, also known as *folie à deux*, even if the "recipients" of the delusions show no prior inclination toward adopting delusions (Bell, Raihani, & Wilkinson, 2020). Further, psychologists and medical anthropologists alike have observed that the presentation and course (i.e., length of morbidity, severity, etc.) of a delusion and psychotic illness vary widely from culture to culture (see Draguns & Tanaka-Matsumi, 2003 and Luhrmann & Marrow, 2016). The upshot of this research for present purposes is that the criteria that a clinician might use to diagnose someone as delusional has varied significantly across cultural contexts. What it means to have a clinical delusion versus a non-pathological, strange belief (or perhaps conviction about a conspiracy theory, a sexist belief, or a racist belief) is sensitive to time and place. For example, the belief that the government could send us a message if they

wanted to, causing the faces of our nation's leaders to appear on screens in our house might be considered a delusional belief for an agent living in ancient Mesopotamia. But in 21<sup>st</sup> century America, it is not a delusion. However, to be convinced that the FBI is following you with the intent of killing you when you've committed no crimes verges more on delusional. (Note that this delusion would not have been possible before 1908, when the FBI was formed. It seems possible to me that the prevalence of persecutory delusions has increased with globalization and advances in mobile technology. While such a claim is orthogonal to this discussion, my hope is that these possible differences in delusion presentation over time indicates that social and cultural factors are at play in determining what clinicians will designate as a delusion.)

In sum, occupying the (DO) state does not necessarily make one a delusional agent (in a clinical sense); sociocultural factors matter, too. People with contradictory beliefs of any kind could be considered as a part of the (DO) state with respect to that belief. Alex the implicit sexist and people who refuse to eat chocolate shaped like dog feces are examples of this.

#### 4.3.4. *The (OO) state*

The (OO) state is the state that agents without any delusions generally occupy. For example, if I hold the orthodox belief "The FBI is not trying to kill me" in both the explicit and implicit compartments, I occupy the (OO) state with respect to that belief. In argument, I will defend the claim that the FBI is not trying to kill me, and I also will reflect this belief in unreflective action (e.g., not acting particularly anxious near government buildings).

## 5. Closing remarks

I have proposed a modified doxastic model that makes use of non-evidential motivators for belief formation and a compartmentalization strategy that appreciates the distinction between implicit and explicit beliefs. Using this model, I have suggested explanations for many clinical features of delusions. This includes a response to the challenge posed by **Double Bookkeeping**, in which I propose compartmentalization between implicit and explicit beliefs to explain how a person's actions can contradict their professed delusions. I also described the advantages of my view over traditional doxastic and non-doxastic views in explaining these behaviors of delusions.

The updated doxastic view seems to me a promising new voice in the conversation about delusions. This doxastic view, which incorporates the impact of emotional tension on delusion content and formation, enables to ask new questions about delusions that were previously unavailable. For example, we can start to have meaningful discussions about how and when emotions override evidence in reasoning. As mentioned before, emotional reasoning is already recognized as a part of even non-pathological human cognition, so a philosophical model of belief that carefully incorporates emotional reasoning seems promising. Specific to delusions, we might begin to track the emotional states associated with delusion etiology to see how strongly the connection between emotional tension and delusion formation holds.

If cultural factors have an impact on the emotional tension an individual experiences because of facilitated integration in society, my model provides a way to understand the varying severity and prevalence of delusional morbidity across cultures. The emotional dimension of delusions might also help explain trends observed in medical anthropology, including the tendency for minorities in less integrated communities to have a higher prevalence of psychotic disorders (see Halpern & Nazroo, 2000 and Boydell et al., 2001) and for cultures with tightly knit, intergenerational, family structures to have faster recovery from (or less severe presentation of) psychotic disorders (see Hopper et al., 2007).

If this model is accurate, the updated doxastic model can serve as another bridge between philosophy and clinical psychology. Or, at the very least, I hope that this discussion has shown that plausible ways to root seemingly complex mental pathology in ordinary psychological notions (e.g., beliefs and emotions) can be found.

## References

- American Psychiatric Association (2013). Schizophrenia and other psychotic disorders. In *Diagnostic and Statistical Manual of Mental Disorders* (5<sup>th</sup> ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Bayne, T., & Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind and Language*, 20(2), 163-188. <https://doi.org/10.1111/j.0268-1064.2005.00281.x>
- Bell, V., Raihani, N., & Wilkinson, S. (2020). Derationalizing delusions. *Clinical Psychological Science*, 9(1), 24-37. <https://doi.org/10.1177/2167702620951553>
- Berle, D., & Moulds, M. L. (2013). Emotional reasoning processes and dysphoric mood: cross-sectional and prospective relationships. *PLoS ONE*, 8(6). <https://doi.org/10.1371/journal.pone.0067359>

- Berrios, G. E. (1991). Delusions as “wrong beliefs”: A conceptual history. *British Journal of Psychiatry*, 159(S14), 6-13. <https://doi.org/10.1192/s0007125000296414>
- Bleuler, E. (1924). *Textbook of psychiatry*. Taylor & Francis.
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford University Press.
- Bortolotti, L. (2010). Double bookkeeping in delusions: Explaining the gap between saying and doing. In J. H. Aguilar, K. Frankish & A. A. Buckareff (Eds.), *New waves in philosophy of action* (pp. 237-255). Palgrave Macmillan.
- Boydell, J., Van Os, J., McKenzie, K., Allardyce, J., Goel, R., McCreadie, R. G., & Murray, R. M. (2001). Incidence of schizophrenia in ethnic minorities in London: Ecological study into interactions with environment. *British Medical Journal*, 323(7325), 1336-1338. <https://doi.org/10.1136/bmj.323.7325.1336>
- Chhaya, T. (2017). She’s not my mother: A 24-year-old man with Capgras delusion. *Federal Practitioner*, 34(12), 38-41.
- Currie, G., & Ravenscroft, I. (2011). *Recreative minds: Imagination in philosophy and psychology*. Clarendon Press.
- Draguns, J. G., & Tanaka-Matsumi, J. (2003). Assessment of psychopathology across and within cultures: Issues and findings. *Behaviour Research and Therapy*, 41(7), 755-776. [https://doi.org/10.1016/s0005-7967\(02\)00190-0](https://doi.org/10.1016/s0005-7967(02)00190-0)
- Egan, A. (2008). Seeing and believing: Perception, belief formation and the divided mind. *Philosophical Studies*, 140(1), 47-63. <https://doi.org/10.1007/s11098-008-9225-1>
- Flores, C. (2021). Delusional evidence-responsiveness. *Synthese*, 199, 6299-6330. <https://doi.org/10.1007/s11229-021-03070-2>
- Frankish, K. (2009). Delusions: A two-level framework. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience: Philosophical perspectives* (pp. 269-284). Oxford University Press.
- Gallagher, S. (2009). Delusional realities. In M. R. Broome & L. Bortolotti (Eds.), *Psychiatry as cognitive neuroscience: Philosophical perspectives* (pp. 245-268). Oxford University Press.
- Gendler, T. S. (2008). Alief and belief. *Journal of Philosophy*, 105(10), 634-663. <https://doi.org/10.5840/jphil20081051025>
- George, M., Maheshwari, S., Chandran, S., Manohar, J.S., & Sathyanarayana, T. S. R. (2017). Understanding the schizophrenia prodrome. *Indian Journal of Psychiatry*, 59(4), 505-509. [https://doi.org/10.4103/psychiatry.IndianJPsychiatry\\_464\\_17](https://doi.org/10.4103/psychiatry.IndianJPsychiatry_464_17)
- Greco, D. (2015). Iteration and fragmentation. *Philosophy and*

- Phenomenological Research*, 91(3), 656-673. <https://doi.org/10.1111/phpr.12086>
- Halpern, D., & Nazroo, J. (2000). The ethnic density effect: Results from a national community survey of England and Wales. *International Journal of Social Psychiatry*, 46(1), 34-46. <https://doi.org/10.1177/002076400004600105>
- Hopper, K., Glynn, H., Janca, A., & Sartorius, N. (2007). *Recovery from schizophrenia: An international perspective; A report from the WHO coordinated international study of schizophrenia*. Oxford University Press.
- Jung, N., Wranke, C., Hamburger, K., & Knauff, M. (2014). How emotions affect logical reasoning: Evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00570>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Lewis, D. (1982). Logic for equivocators. *Noûs*, 16(3), 431-441. <https://doi.org/10.2307/2216219>
- Luhrmann, T. M., & Marrow, J. (2016). *Our most troubling madness: Case studies in schizophrenia across cultures*. University of California Press.
- Mandelbaum, E. (2013). Against alief. *Philosophical Studies*, 165(1), 197-211. <https://doi.org/10.1007/s11098-012-9930-7>
- Parnas, J., Urfer-Parnas, A., & Stephensen, H. (2020). Double bookkeeping and schizophrenia spectrum: Divided unified phenomenal consciousness. *European Archives of Psychiatry and Clinical Neuroscience*. <https://doi.org/10.1007/s00406-020-01185-0>
- Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms of the brain: Probing the mysteries of the human mind*. William Morrow and Comp.
- Rinard, S. (2018). Believing for practical reasons. *Noûs*, 53(4), 763-784. <https://doi.org/10.1111/nous.12253>
- Rinard, S. (2019). Equal treatment for belief. *Philosophical Studies*, 176(7), 1923-1950. <https://doi.org/10.1007/s11098-018-1104-9>
- Sass, L. A. (1994). *The paradoxes of delusion: Wittgenstein, Schreber, and the schizophrenic mind*. Cornell University Press.
- Schwitzgebel, E. (2010). Acting Contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4), 531-553. <https://doi.org/10.1111/j.1468-0114.2010.01381.x>

Received 25<sup>th</sup> August 2023; revised 13<sup>rd</sup> November 2023; revised 2<sup>nd</sup> March 2024; accepted 5<sup>th</sup> March 2024.