



# Superstitious Confabulations

Anna Ichino<sup>1</sup>

© The Author(s) 2018

## Abstract

Superstition and confabulation are extremely pervasive in our cognitive lives. Whilst both these phenomena are widely discussed in the recent psychological literature, however, the relationship between them has not been the object of much explicit attention. In this paper, I argue that this relationship is actually very close, and deserves indepth consideration. I argue that superstitious and confabulatory attitudes share several key features and are rooted in the same psychological mechanisms. Moreover, some of the key features that superstitious and confabulatory attitudes share reveal such attitudes to be non-doxastic in nature, with important implications for our assessment of their epistemic rationality. Many instances of what we call superstitious and confabulatory ‘beliefs’ are not, in fact, beliefs; hence, entertaining them may be less irrational than it *prima facie* seems to be.

**Keywords** Confabulation · Superstition · Belief · Imagination · Irrationality · Dual-process theories

Confabulation is a heterogenous phenomenon, which varies across a number of dimensions—including content, mode of elicitation, aetiology, and more. While acknowledging this heterogeneity, recent philosophical discussions have focussed mostly on some particular kinds of confabulation: notably, confabulations that are about the self, and externally elicited—classic examples being cases of memory distortions and of choice blindness. With a few exceptions, such discussions highlighted the epistemic faults of these confabulations, especially in relation to self-knowledge.

In this paper, I draw the attention to a different sort of confabulations, which are typically about the world (as opposed to the self), and spontaneous (as opposed to externally elicited). I call them ‘superstitious confabulations’, since they originate in the domain of superstitious thinking and practices. I argue that these sorts of confabulations are peculiar in that they are often entertained in a non-doxastic way—with positive consequences for the assessment of their epistemic rationality; and that they can teach us important lessons on the epistemic and normative status of confabulations more in general.

In Sect. 1, I introduce paradigmatic instances of superstitions and of confabulations. In Sect. 2, I argue that they share some central features—including ill-groundedness, lack of deceitful intentions, motivational origins, and gap-filling function—hence, that we can identify a class of ‘superstitious confabulations’. In Sect. 3, I defend a non-doxastic approach to superstitious confabulations, and I explore its normative implications—suggesting that superstitious attitudes are less irrational than they are generally taken to be. I then conclude, in Sect. 4, by asking to what extents my non-doxastic approach and its implications may also apply to more paradigmatic, non-superstitious, forms of confabulation.

## 1 Superstition and Confabulation

### 1.1 Superstition

When I was at High School, I always used the same pen for written tests as I took notes with in my classes—after all, it already knew the right answers. When I cycle to work in the morning, I always make sure to get over the same ‘lucky crack’ in the road—it’s my secret trick to keep misfortune at bay. At the supermarket, I always pick the second item in the row on a shelf—obviously, that’s the only non-flawed one. I

---

✉ Anna Ichino  
anna.ichino@unimi.it

<sup>1</sup> Philosophy Department, Università degli Studi di Milano,  
Via Festa del Perdono 7, 20122 Milan, Italy

do also read my horoscope every Sunday—so as to find out what the rest of the week will be like.

These are just a few of the many small rituals and superstitious practices that fill my everyday life. And, apparently, I am in good company. Students, athletes, politicians, musicians, fishermen, policemen and gamblers are all categories of people well-known for the propitiatory rituals and lucky charms they engage with (see Vyse 2014, Ch. 2). You may have heard, for instance, of David Beckham’s pre-game rituals: like stepping on the pitch with the right foot first—to ensure right shots; or wearing a brand-new football outfit at each match—to bring out brand-new energies. And Beckham’s fans are ready to pay quite some money to possess his ‘old’ outfits, as indeed people do for such things as Lady Diana’s wedding dress, or John Lennon’s hand-written lyrics. We are also all familiar with a variety of traditional superstitious actions, like touching wood, crossing fingers, tossing spilled salt over one’s shoulder, and so on.

Psychological research provides systematic evidence—both from field surveys and from the lab—of how widespread among educated, intelligent adults is the tendency to behave in superstitious ways more or less similar to those just described (see Vyse 2014, Ch. 1–2 for an overview). In a series of seminal experiments, for instance, Rozin et al. (1986) found subjects reluctant to throw darts at the photograph of a loved one, unwilling to drink from a bottle labelled ‘Poison’, whilst knowing that it actually contained sugared water, or unwilling to wear an acknowledgedly fully sterilized sweater that they were told previously belonged to Adolf Hitler. Risen and Gilovich (2007, 2008) found subjects reluctant to perform actions that they take to ‘tempt fate’—like exchanging lottery tickets, commenting boastfully on a streak of success, or making ‘incautious’ optimistic remarks (such as ‘No one I know will ever get cancer’).

What all these otherwise disparate actions (and action omissions) have in common, as it turns out, is the peculiar kind of thinking that they seem to presuppose: a kind of thinking which departs in various ways from our natural/scientific view of the world—taking, for instance, action at distance to be possible (‘The positions of the stars at my birth will determine the rest of my life’) and objects to carry non-physical essences (‘Hitler’s sweater carries Hitler’s wickedness’); treating inanimate things as sentient creatures (‘My wise pen, who attended all the lectures, will help me in the exam’), surface similarities as essential ones (‘The image of my loved incorporates her essence’), and random coincidences as teleologically significant (‘The second item on the shelf was placed there on purpose’). More generally: seeing meaning, purpose, and agency where in fact there is none, or more of them where there is less.<sup>1</sup> This, indeed,

<sup>1</sup> Of course, there may be different explanations for the actions I mentioned that do not involve superstitious thoughts—and perhaps even explanations that do not involve any thoughts whatsoever, but

has been identified as the most distinctive feature of superstitious thoughts, which sets them apart from other sorts of scientifically ungrounded thoughts: a systematic tendency to exaggerate the extents to which properties like mindedness, rationality, and agential control are manifested in the world, thereby resulting in *category mistakes* where the core attributes of mental and physical processes are conflated (Lindeman and Aarnio 2007). A tendency to overestimate the world’s coherent and intelligent design—as Currie and Jureidini (2004) put it, dubbing this as ‘over-coherent thinking’.

Lots of interesting work has been devoted to characterising this kind of thinking more precisely, investigating the principles that underly it and the specific kinds of ontological confusions it involves (see Nemeroff and Rozin 2000; Hood 2012). But the general characterisation just outlined will suffice for the sake of the present discussion, and notably for the sake of appreciating some key similarities between superstitious and confabulatory thinking.

## 1.2 Confabulation

Henry has been married to Martha for over 30 years, and has four children with her aged between 22 and 32. Following neurosurgery, he has an impairment in autobiographic memory—which is why when the doctor asks him how long he has been with Martha, he incorrectly replies: ‘4 months’. To the further question about how they managed to have four children in 4 months, he replies that they are adopted.

Aitana, who suffers from Capgras delusion, claims that her son has been replaced by an impostor. When the doctor asks her how she can tell the difference between her son and the impostor, she replies that she can because ‘the impostor

Footnote 1 (continued)

only low-level associative routines. It might well be the case, for instance, that the idea that *the pen will tell me the right answers* never occurred to me, and I use that pen *just* out of habit; or that I never heard that *spilling salt is an evil omen*, and I toss spilt salt over my shoulder *just* to imitate what other people do. Insofar as this were true, the actions in question would not count as ‘superstitious’, since what makes an action superstitious is arguably the fact of involving some sort of superstitious thinking. But what matters to us here is that, even if alternative superstition-free explanations can always be conjectured, they do not seem to capture what goes on in the sort of cases I mentioned—which we are, indeed, typically classified as cases of superstition. This is something on which psychologists (as well folks’ ordinary talk) largely agree. Thoughts about good and bad luck *do* typically play some role in explaining why we cross our fingers or toss split salt—as our own verbal reports clearly indicate (Vyse 2014). Thoughts about objects’ invisible, contagious essences do typically play some role in explaining our willingness to own objects that belonged to someone we like—or unwillingness to own objects that belonged to someone we dislike (Rozin et al. 1986). And, as Risen (2016, p. 182) observes: “Why would people be afraid to comment on a streak of success, if they rejected the notion that the universe punishes such modest acts of hubris?”

has different-coloured eyes and is not as brawny as her son'; moreover, 'her real son wouldn't kiss her'.

Roger is an implicitly biased employer, who—in the course of a hiring decision—has just selected the CV of Giulio over the identical (if not even better) CV of Giulia. Without realising his bias, let alone its influence on his decision, when asked to explain the said decision he replies that Giulio's CV reveals his research to be of a higher quality than Giulia's.

Maria is a participant in the now very famous experiment that Wilson and Nisbett (1978) carried out in a shopping mall, where four *identical* pairs of nylon stockings were put on display and then, under the cover story of a customer survey, passers-by were asked to judge which of those items was of the best quality. Without realising that the stockings were in fact identical, most subjects ended up choosing the one furthest to their right, arguably due to position effects such as the 'right-hand bias'. Yet, being unaware of such effects, when asked to explain their choice the vast majority of them were willing to offer seemingly rational explanations, mentioning allegedly different features of the items. Maria, for instance, promptly said that it was the knit and elasticity that she felt superior.

Miri is a participant in another famous experiment, by Haidt (2001), where subjects were presented with a story about two siblings who make love. According to the story, they make love only once, just to see how it feels, taking all sorts of birth-control precautions and making sure they do not harm each other in any way. Both of them enjoy the experience, which they then keep as a special secret that makes them feel even closer to each other. On hearing this story, most subjects automatically judged the siblings' action to be morally wrong—arguably due to deeply entrenched moral intuitions driven by low-level emotional reactions (notably, disgust). Lacking introspective access to such reactions, however, when asked to explain their moral judgements, many of them came up with reasons that were clearly unsubstantiated by the story itself. For instance, "they pointed out the dangers of inbreeding, only to remember that [the siblings] used two forms of birth control", or "argued that [the siblings] will be hurt, perhaps emotionally, even though the story makes it clear that no harm befell them" (Haidt 2001, p. 814). Eventually, they found themselves dumbfounded, but still refused to change their initial judgements.

Subjects' responses in all these cases—which I reported, with only minimal adaptations, from the recent psychological literature—have been described as instances of *confabulation*. See for instance Bortolotti (2017) on Henry-like cases of *mnemonic confabulation* and Miri-like cases of *confabulation about moral attitudes*; Coltheart (2017) on Aitana-like cases of *confabulation in delusion*; Hirstein (2005) on Maria-like cases of *confabulation about consumer choices*;

Sullivan-Bissett (2015) on Roger-like cases of *confabulation about implicitly biased choices*.<sup>2</sup>

The term 'confabulation' comes from the Latin 'fabula', which means story. It refers to the fact that subjects in these cases seem to lack access to a piece of evidence relevant to the explanation of a given fact/event; but, rather than admitting ignorance, they go on telling something—some sort of story, indeed—about it: a story which is therefore ill-grounded, and yet sincerely endorsed and communicated with no intention to deceive.

Far from being limited to the clinical domain, so-called 'everyday confabulations' like those of Maria, Roger, and Miri are shown to be pervasive in the non-clinical population, constituting a common feature of normal cognition. As Hirstein (2005, p. 4) observes: "Once one forms a concept of confabulation from seeing it in the clinic or in the neuropsychological literature, one starts to see mild versions of it in normal people. We are all familiar with people who seem unable to say the words *I don't know* and will quickly produce some sort of plausible-sounding response to whatever they are asked".

Whilst this variety of manifestations across different domains makes it hard to provide a definition in terms of necessary and sufficient conditions (Bortolotti and Cox 2009), there is general agreement on some key features shared by paradigmatic confabulations, as well as on some dimensions along which they differ from each other. Paradigmatic confabulations are typically: (i) ill-grounded, (ii) reported without deceitful intentions, (iii) involving a motivational component, and (iv) playing a gap-filling function. On the other hand, they can vary with respect to: mode of elicitation (spontaneous vs. provoked), aetiology (pathological vs. non-pathological), format (verbal vs. non-verbal), content (what subjects confabulate about), subjective degree of conviction, and possibly in other ways.

In what follows I shall discuss the key features of confabulation (i)–(iv), and argue that they are shared by paradigmatic superstitions; whilst differences between superstitions and confabulation can be traced back to the dimensions of variation along which instances of confabulations themselves differ from each other. Superstitions, I shall then conclude, can be seen as a sub-kind of confabulation.

<sup>2</sup> Confabulation has also been observed in such conditions as anosognosia, split-brain, post-hypnotic amnesia, and indeed various more (see e.g. Hirstein 2005; Coltheart 2017; Schnider 2018).

## 2 Key Shared Features

### 2.1 Ill-Groundedness And Lack of Deceitful Intentions

Ill-Groundedness is a key feature of confabulation highlighted by virtually all accounts of this phenomenon (see e.g. Hirstein 2005; Sullivan-Bissett 2015; Bortolotti 2017). This is due to the fact that confabulations occur when subjects have impaired access to a piece of evidence relevant to the explanation of a given event—whether it is a piece of memory about one’s past, as in Henry’s case, a piece of information about the causes of one’s attitudes/choices, as in Maria, Aitana, Roger and Miri’s cases, or something else. Whether this lack of access is due to pathological or non-pathological factors, subjects don’t recognize it, but seek to explain the event in question by making up a story. In so doing, they ‘tell more than they can know’—as Nisbett and Wilson (1977) famously put it—providing ill-grounded, and indeed very often false, claims.<sup>3</sup>

<sup>3</sup> An important terminological note is in order here, since my use of the term ‘ill-groundedness’ departs in one important respect from its ordinary meaning. The ordinary meaning of ‘ill-groundedness’ is a *thick* concept, which includes both: (1) a *description* of the relation between a given entity (typically, a representational state) and evidence, and (2) a (negative) *evaluation* of the relation described in (1). Generally speaking, indeed, when we say that a given representational state R is ‘ill-grounded’ we are not just describing R as scarcely supported by evidence, but we are also criticising R, saying that, due to this scarce evidential support, R is somehow faulty. For the sake of the present discussion, however, I am assuming a normatively *thin* notion of ‘ill-groundedness’, that includes only its descriptive component: saying that a state R is ‘ill-grounded’ in my sense here only amounts to saying that R isn’t (fully) supported by evidence, without providing any normative evaluation about this state of affairs. The reason why I am assuming this normatively thin notion of ‘ill-groundedness’ is that the ordinary, thick notion can arguably only apply to *doxastic* representations (like beliefs, judgements, or assertions), that are in some relevant sense *supposed* to get evidential support, and can therefore be criticized insofar as they lack it; whilst it does not seem to be applicable to *non-doxastic* representations (like imaginings, or fictional speech-acts), which are free from evidence norms, hence also immune from epistemic criticism. Given that one of the aims of this paper is precisely to question the doxastic status of confabulations (as, indeed, of superstitions) arguing that they are more imagination-like than belief-like, I need a normatively neutral way to characterise their relation to evidence to begin with. From now on, then, my talk of ‘ill-groundedness’ shall be *thinly* understood as simply *describing* the property of having little (or no) evidential support, without any negative (nor indeed positive) *normative* implications. The fact that confabulations and superstitions are ‘ill-grounded’ in this normatively thin sense is uncontroversial—this is the point of agreement between myself and most authors in the literature. Whether or not confabulations and superstitions can be taken to be ‘ill-grounded’ also in the thick, ordinary sense of the term, on the other hand, will depend on whether or not they are doxastic states—which will be the subject of my discussion in Sect. 3–4 below (I will argue they many of them are not).

This should not suggest confabulators to be typically insincere, however. On the contrary, a striking feature highlighted by most accounts is that confabulations are *not* produced with an intention to deceive (see e.g. Hirstein 2005; Bortolotti 2017). This is usually explained by appealing to the fact that confabulators are unaware of the ill-groundedness/falsity of their claims, whereas at least some degree of such awareness is necessary for ascribing deceitful intentions (Sullivan-Bissett 2015, pp. 552–553). And even those who question the extents to which confabulators are unaware of the ill-groundedness of their claims—highlighting their expressions of doubt and low-confidence (Fiala and Nichols 2009)—do not thereby take them to be insincere, but simply to be tentatively communicating something for which they acknowledgedly lack conclusive evidence. This is then something on which virtually everyone would agree: whether or not Henry, Aitana, Roger, Maria, and Miri are aware of the ill-groundedness of their claims, their intention in making such claims is not a deceitful one. If it were, they would not count as confabulators: confabulations are not lies (unless there could be such thing as ‘honest lies’).

Ill-groundedness and lack of deceitful intention, however, are not unique to cases of confabulations. There are many ‘honest mistakes’ that do not thereby count as confabulatory—such as mistakes due to momentary confusion, absent-mindedness, or other such factors.<sup>4</sup> What sets confabulation aside from these other sorts of honest mistakes are arguably the latter two features I mentioned, having to do with their motivational origins and, notably, gap-filling function.

### 2.2 Motivational Origins and Gap-Filling

Most accounts take confabulation to crucially involve a motivational element—that is, to be a purposeful construction with ‘a *raison d’être* in the teleological sense’ (McKay and Kinsbourne 2010, p. 291), due to the role played in it by goal-directed states like desires, intentions, or other pro-attitudes. Motivational elements can be operative at two different, though related, levels.

First, they have been pointed out to play a role in determining the specific contents that confabulations come to have (McKay and Kinsbourne 2010; Sullivan-Bissett 2015; Schnider 2018, Ch. 7). This is how, for instance, Maria’s desire to show expertise about weaving techniques might be what motivates her to mention weave quality (rather than, say colour) as the reason for her choice; or Roger’s desire to be competent and professional might be what motivates him to mention the quality of Giulio’s research (rather than, say, Giulio’s good look) as the decisive reason for hiring

<sup>4</sup> With ‘mistakes’ here I refer to ‘ill-grounded claims’ – in the normatively thin sense specified above.

him. These motives are typically positively biased, in that they seek to embellish the subject's situation, producing self-enhancement and "wishful reality distortions" (Schneider 2018, p. 188; see also; McKay and Kinsbourne 2010, pp. 291–293).

But motivational elements have also been taken to play a more fundamental role—contributing to determine not only why we form confabulations *with a specific content*, as opposed to a different one, but also why we form confabulations *at all*, as opposed to not forming them (Sullivan-Bissett 2015:552). It is at this latter motivational level that lies what is arguably the most distinctive feature of confabulation, which I called 'gap-filling'. At a basic level, indeed, paradigmatic confabulations seem to arise from our need to fill some gap: to make sense of bits of information that do not clearly fit together, due to a missing piece of evidence which cannot be accessed.

This idea was especially prominent in the first 'narrow' accounts that focussed on clinical cases of mnemonic confabulation—characterised as the result of patients' unconscious desires to cover-up gaps in their impaired autobiographic memory (see Schneider 2018, Ch. 1). But recent accounts suggested that gap-filling should be intended in a more comprehensive way. As Sullivan-Bissett (2015, p. 552)—building on Hirstein (2005)—puts it: "The gap-filling claim needs to be construed more broadly and not couched only with respect to gaps *in memory*. We might instead think of confabulatory explanations as filling gaps *at a certain level in the cognitive system*, insofar as they help produce 'complete, coherent representations of the world'."

In a similar vein, Wheatley (2009) characterises gap-filling as a basic *modus operandi* by which our minds "create meaning from confusing and often contradictory inputs" *at all cognitive levels* (Wheatley 2009, p. 204)—from memory, that is always a largely constructive process, to the lower levels of sensory and motor cognition, where our perceptual system is constantly engaged in the task of completing underdetermined and ambiguous stimuli in meaningful ways. Confabulations are nothing but *hypertrophic* manifestations of this gap-filling in a variety of domains—including, notably, the domain of self-representations—where the natural search for meaning and coherence amounts more to the fabrication of new material than to the interpretation of underdetermined inputs.

### 2.3 A Drive for Causal Understanding

This broader characterisation of the relevant gaps not just as memory gaps, but as explanatory gaps of a variety of kinds, fits well with an account recently put forward by Coltheart (2017), according to which we should understand confabulations as a product of the basic property of human cognition that Gopnik (2000) called "the drive for causal

understanding". This is a drive that, at an unconscious level, motivates us to provide causal explanations for the facts and events that are relevant to us, since tolerating gaps in our causal maps of the world is deeply distressing. As Gopnik observes, indeed, being in the state of wondering why something happened is an unsettling and arousing experience, whilst "finding an explanation for something is accompanied by a satisfaction that goes beyond the merely cognitive" (Gopnik 2000, p. 311). This is why we happen to produce explanations even when we lack the relevant information and we should better recognize ignorance: what happens in confabulation is that "the sense of release achieved by arriving at an answer outweighs any concerns about the plausibility of that answer (...); since it is better for us to have an explanation—any explanation, no matter how bizarre—than to have none" (Coltheart 2017, p. 68).

I actually take Coltheart's view to require one important refinement. Whilst his idea of a drive for causal understanding at the basis of confabulations seems right, this is not a generic drive towards *any* explanations whatsoever, as the above quote suggests, but rather a drive towards *one particular kind* of causal explanations: causal explanations which feature *reasons*—typically, psychological or mentalistic ones. This is especially evident in confabulatory explanations of our own attitudes and choices—which are typically rationalising explanations in terms of high-level motivating reasons (e.g. 'I choose this item because I thought it was of a high quality'), rather than merely causal explanations in terms of low-level factors like unconscious heuristics and biases (e.g. 'I choose this item because of right-hand position effect'). Of course, confabulators do not know that what *actually* explains their attitudes and choices are indeed causal factors of that latter sort. But the point is that they do not even seem consider such causal factors as possible explanatory options. On the contrary, they may even actively resist acknowledging them—as in Wilson and Nisbett's study—where, being "directly probed as to whether the position of the stockings could have influenced their decisions, all but one denied that such a thing could have had any influence" (1978, p. 124). What they were after, apparently, were explanations in which they could recognize themselves as rational agents driven by coherent motives, goals and values (cf. Keeling 2018 for a related point).

This sort of reasons-based gap-filling is central to the analogy with superstitions, to which I now turn.

### 2.4 Superstitious Confabulations

The features that I identified as most distinctive of confabulatory *verbal reports* apply also, *mutatis mutandis*, to the relevant mental states that underly them—that is, to confabulatory *thinking*. In what follows I shall argue that such features are shared by superstitious thinking.

This is quite obvious with respect to ill-groundedness and lack of deceitful intentions. Superstitions' ill-groundedness is a direct consequence of their tendency to go beyond the available evidence, exaggerating the world's mindedness and rationality in the ways I described in Sect. 1.1. As such, this feature figures in many definitions of superstitions, like the popular definition of the *Encyclopaedia Britannica*, in terms of 'beliefs or practices *groundless in themselves* and inconsistent with the degree of enlightenment reached by the community to which one belongs' (quoted in Vyse 2014, p. 23; emphasis mine). This also means that—like confabulations—superstitions are typically false. And yet—again like confabulations—they are not typically accompanied by deceitful intentions. Although of course there may be exceptions to this (as with cheating astrologists), common expressions—verbal or non-verbal—of superstition are not associated with insincerity.<sup>5</sup>

This, admittedly, is not yet enough to establish my view about superstitions' confabulatory nature—given that, as we have seen, the conjunction of ill-groundedness and lack of deceitful intentions is not an exclusive prerogative of confabulations. But whilst it is true that not all 'honest mistakes' amount to confabulating, there is reason to think that those involved in superstitions do, given that superstitions share also the other two, most distinctive features of confabulation: motivational origins and, relatedly, gap-filling functions.

Those familiar with the psychological literature on the origins of superstitious thinking will have already noted various striking similarities with what I said about confabulations. Superstitions' formation has indeed been associated to the same gap-filling tendency that we have seen at play in confabulations: that is, our tendency to actively manipulate incoming inputs by filling their gaps in meaningful ways—the most meaningful ones for us generally being ways related to human-like agency, purposes, reasons. This tendency is operative from the lowest sensory levels—with, for instance, our perceptual system naturally responding to stimuli that are even remotely face-like (whether they are clouds, landscapes, or even just some lines and spots on

a piece of paper), leading us to 'see faces' into them (cf. Fig. 1). And superstitions, like confabulations, have been taken to be high-level, *hypertrophic* manifestations of it (Guthrie 1993; Currie and Jureidini 2004; Hood 2012). As we tend to (perceptually) see human-like features into inanimate objects, completing our perceptions in anthropomorphic ways, so we tend to (cognitively) see hidden meaning, agency and reasons in the events of our lives, completing our propositional representations of them in the 'over-coherent' ways I described (see Sect. 1.1 above). This is why we are naturally keen to ascribe meaning and purpose to 'apparently irrelevant' facts such as a cycling over a certain crack, spilling some salt, or being born on a given date—inscribing them into a larger teleological design.

As with confabulation, this hypertrophic search for teleological coherence has been traced back to a strive for causal explanations. Risen (2016, p. 287), for instance, argued that "an appealing feature of magic and superstition is its ability to offer an explanation for any and all phenomena", contending that "it is people's natural tendency to infer causal relationships that makes this so". And, indeed, Colheart himself (2017, p. 67) suggests that the same "drive for causal understanding" that underlies confabulation is likely to be at play also in superstitious and magical thinking.

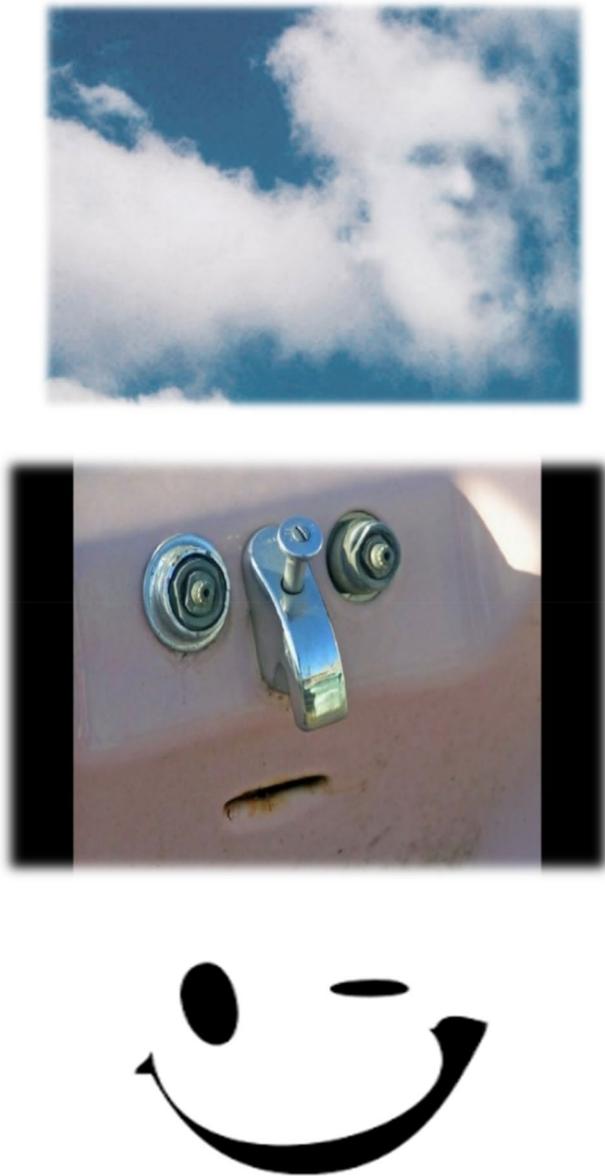
Although here again I would argue that what superstitions strive for are not just *causes*, but causes that are also *reasons*. In this case, reasons of a finalistic sort: superstitions seek to make rational sense of what happens to (and around) us, presenting it as teleologically meaningful—hence also more manageable and controllable.

The pervasive human desire/need for control, indeed, has been identified as one of the most basic motivations that superstitions seek to satisfy, by creating an 'illusion of control' upon our otherwise worryingly unpredictable lives (see e.g. Vyse 2014, pp. 159–161; Risen 2016, p. 188).

Besides, a variety of more specific motivational factors have been taken to influence superstitions' particular contents. Typically, again as in cases of paradigmatic confabulations, positively biased factors—such as desires "to belong, to find the world benevolent, and to maintain self-esteem" (Lindeman 1998, p. 258; cf. also Vyse 2014; Hood 2012). "Even a short glance at these motives—Lindeman notes—shows that pseudoscience [i.e. superstition] fulfils them more easily and better than does scientific information (...) replacing the scientific picture of a highly abstruse, random and unjust world (...), with a more coherent, controllable and positive view". Like paradigmatic confabulations, superstitions, too, seem then to involve 'wishful reality distortions'.

Overall, then, although not every honest mistake counts as confabulatory, the ways in which superstitions are mistaken look so similar to the ways in which paradigmatic confabulations are, that I think we should recognize, within

<sup>5</sup> Why not? As with confabulation, there may be different routes to this conclusion. Sometimes deceitful intentions are easily ruled out because subjects do not even publicly express their superstitions, so a fortiori they do not express them *with an intention to deceive* (see Vyse 2014, pp. 16–18 on subjects' reluctance towards public superstitious avowals). In other cases, deceitful intentions can be ruled out because subjects introduce their superstitious reports with explicit acknowledgements about their falsity/ill-groundedness (see discussion of 'superstitious acquiescence' in Sect. 3 below). Yet in other cases, lack of deceitful intentions may depend on the fact that subjects are unaware of the falsity/ill-groundedness of their superstitions—whereas, as I noted above, at least some such awareness seems necessary to charge someone with an intention to deceive (note however that, as I shall argue in Sect. 3, cases of this latter sort are not the norm in superstition).



**Fig. 1** Some examples of perceptual gap-filling: seeing human faces into clouds, sinks, and random marks on paper

the broad confabulation category, a sub-class of ‘superstitious confabulations’.

Of course, this does not amount to denying differences between superstitions and other more paradigmatic confabulations. One significant difference, as we have seen, concerns their (kinds of) contents. Whilst paradigmatic confabulations are in some relevant sense about *the self* (about our own attitudes/choices/motivations), superstitions typically concern *the external world* (or, at least, what happens in it that is relevant to us). This explains why in paradigmatic forms of confabulation the hypertrophic search for reasons and intentions does not involve the category mistakes that it involves in paradigmatic superstitions: over-ascribing rationality

and purpose to self-conscious agents is mistaken, but not *categorically* mistaken as it is ascribing them to inanimate objects. Moreover, and relatedly, whilst superstitions are spontaneously and autonomously produced, confabulations typically occur in conversational contexts where they are elicited by external questions—indeed, pretty challenging questions that threaten subjects’ rationality and agential ownership (Bortolotti 2017). This is also why paradigmatic confabulations, but not always superstitions, typically get verbal expression.

Whilst recognizing the significance of these—and possibly other—differences, however, we can think of them as variations within the same broad psychological family. After all, as we have seen, there is wide agreement on the fact that confabulations can vary in content, mode of elicitation, format, and many more ways. Insofar as this is true, and insofar as we envisage a characterisation of confabulation encompassing the full variety of phenomena that have been labelled as such in the literature, it’s not clear on what grounds superstitions should be excluded from it.

That said, the issue here is at least partly terminological, depending on how fine-grained we want our mental categories to be; and I don’t need to insist at any cost on my label of ‘superstitious confabulations’. What matters to me is that we recognize that these two phenomena—commonly called, respectively, ‘confabulation’ and ‘superstition’—share some key features and can be seen as close relatives.

Why does that matter? It matters for our general understanding of how the mind works—making us realise how pervasive certain features of our cognition are. And it matters for a better understanding of the two phenomena in particular—fostering closer interactions between research programs in these two areas, which so far have gone on in almost total disregard from each other.

In the remaining part of the paper I will provide an example of how fruitful such interactions can be. I shall argue that an important theoretical model recently developed to explain superstitious attitudes reveals such attitudes to be non-doxastic in nature (Sect. 3), and that this model may be fruitfully applied also to other more standard forms of confabulation (Sect. 4)—in both cases with important implications for the sorts of normative questions that philosophers have been asking about these two phenomena.

### 3 A Non-doxastic Approach to Superstitious Confabulations

So far, I talked in generic terms of superstitious *thoughts*, remaining deliberately neutral on what sort of mental *attitudes* such thoughts exactly are. Yet, much of what I said is relevant to that question, and my aim in this section is

to address it more explicitly and explore its normative implications.

‘Attitude’ here refers primarily to functional role: what kinds of functional states superstitious thoughts exactly are? The standard answer is that they are beliefs: we commonly ascribe to each other superstitious ‘beliefs’, or ‘beliefs’ in magic. However, I shall argue that psychological work in this area provides grounds to question such ascriptions, suggesting that superstitious attitudes differ from beliefs in some key functional respects—including, notably, sensitivity to evidence and reasons.

Arguing along these lines, note, does not presuppose an idealised picture of belief as perfectly sensitive to evidence. Of course, ‘hot’ factors like desires or emotions can influence belief formation in various ways (see Bayne and Fernández 2009). But granting this is compatible with assuming that at least some degree of evidence-sensitivity is what crucially characterises belief’s functional profile at the inputs level—while, conversely, lacking relevant degrees of such sensitivity compromises doxastic status. It is this latter assumption that lies at the basis of my argument.

Importantly, among the kinds of evidence to which beliefs respond, there is what we can call ‘inferential evidence’: the evidence constituted by a subject’s pre-existing beliefs. This explains why our beliefs are inferentially integrated with each other into a (more or less) holistically consistent system; and it is the reason why, insofar as we take some degree of evidence-sensitivity to be necessary for belief, we should take some corresponding degree of inferential integration to be necessary as well.

This is, then, my point: lacking relevant degrees of evidence-sensitivity and inferential integration, superstitious attitudes can hardly classify as beliefs. Vice versa, the ways in which they are formed and maintained speak in favour of classifications in non-doxastic terms—such as in terms of imaginings, or of other non-doxastic categories in the vicinity.

### 3.1 Superstitions’ Formation and Maintenance

The motivational accounts of superstitions’ formation that I discussed above have already suggested that superstitions are not really a product of evidence-tracking mechanisms. Their formation is rather characterised by a tendency to ‘jump beyond’ the available evidence, fabricating representations the main function of which is to satisfy our basic desires for purpose and control, as well as other contingent motivations and affective inclinations/aversions we may have. Even though, as I granted, this sort of motivational and emotional factors may also play a role in belief formation, that is not the systematic and central role that they play in the case of superstitions—where, on the other hand, the role of evidence is merely a causal, ‘triggering’ one. Superstitions are

*prompted* by evidence in largely associative ways, which remind closely of imaginative prompting. My memory of some good outcome that followed a given event (e.g. cycling over a certain crack), together with my desire for control, prompts me to imagine a causal connection between the said event and outcome; the observation that I did well in the exam after I used a certain pen, together with a desire to keep doing well in future exams, prompts me to imagine that the pen itself can help me; seeing the label ‘Poison’ on a bottle, prompts me to imagine the bottle to actually contain poison;<sup>6</sup> and so on.

This is the picture emerging from recent psychological theorising—which, if not typically talking in terms of ‘imaginative’ prompting, traces superstitious thinking back to the interaction of motivational and affective factors, on the one hand, and low-level associative heuristics, on the other. For instance, Lindeman (1998, pp. 257–258) argues that superstitions arise from “the interplay between basic motives and experiential thinking”—where the latter is defined as “the realm of cognitive heuristics (...) the goal of which is not to find the correct answer, but rather to find a quick decision to guide behaviour”. In a similar vein, Hood (2012) traces superstitions’ formation back to what he calls our ‘supersense’, which is all but an evidence-tracking mechanism—its operations being variously characterised in terms of ‘wishful thinking’ (p. 8), ‘gut feelings’ (p. 40), ‘emotional urges and feelings that run counter to reason’ (p. 85), ‘System-1 heuristics’ (p. 259) (see also Lindeman and Aarnio 2007; Vyse 2014, Ch. 2–3).

But the functional difference between superstitious and doxastic attitudes emerges even more strikingly when we look at how superstitions are maintained: without being revised in the face of contradictory evidence and reasons, even when—this is the key point—subjects are fully aware of such contradictions.

To appreciate this point, we should emphasise once again that, being a natural product of how our minds work, superstitious thinking is widespread in cognitively normal population. Against the traditional view that relegates superstitious thinking to young/primitive/ignorant creatures, recent research highlights its pervasiveness among ‘intelligent, emotionally stable adults’ (Risen 2016, p. 182)—without even finding consistent evidence of inverse correlation with

<sup>6</sup> In this latter case—referring to Rozin and colleagues 1986’s experiment—the claim that my representation of the bottle as *poisoned* is not a response to evidence but is merely associatively triggered by it, is further supported by the observation that, in a variant of the same experiment, subjects seemed to form the very same representation when presented with a label which read ‘NOT Poison’. For a more extensive discussion of this variant of the experiment, defending an account of it in terms of imaginings, see Currie and Ichino (2012), pp. 791–792.

levels of scientific education (Vyse 2014, pp. 49–51). This means that, for most superstitious subjects, superstitions are not only ill-grounded, but also in tension with many other beliefs to which they are committed. And a striking feature of superstitious cognition is that the subjects in question are typically aware of such ill-groundedness and tensions, which they explicitly recognize, without apparently feeling much pressure to resolve them—as in the popular tale reporting the following exchange between Niels Bohr and an incredulous friend who noted a horseshoe hanging over his door: “Niels, it can’t possibly be that you, a brilliant scientist, believe that foolish horseshoe superstition?!”—“Oh, of course not... But I’m told it works even if you don’t!”.

Far from being just anecdotal, similar openly recognised ‘mistakes’ are something with which psychologists have long been familiar (see remarks about them in Rozin et al 1986; Campbell 1996), and which have been object of increasing attention. In a series of recent studies, Risen and her colleagues empirically demonstrated their widespread occurrence, arguing that in order to account for the paradoxical feature of superstitious thinking that such mistakes reveal we need to revise standard cognitive models which conflate error-detection with error-correction, and to recognise the psychological reality of what Risen (2016) dubbed ‘acquiescence’: a process by which ideas that are explicitly understood to be false/ill-grounded are nonetheless willingly endorsed and allowed to play action-guiding roles.

### 3.2 The ‘Decoupled Dual-Process Model’ of Superstition

The standard models to which Risen refers are the well-known ‘dual-process models’, that identify two different systems by which our minds process information: the quick and dirty ‘System 1’, which operates in associative, heuristic-based, effortless ways; and the slower ‘System 2’, which operates in deliberate, analytic, effortful ways. According to the influential ‘Corrective’ (or ‘Default Interventionist’) version of such models, the two systems are both active when we are awake, though System-1 does most of the work, providing the default intuitions upon which System-2 only plays a monitoring role (Kahneman 2011). This means that normally System-2 endorses System-1 suggestions as they are, whilst it only actively intervenes when it detects an error in such suggestions, correcting them consequently. When System 1 intuitions are in line with System-2 rational analysis, instead—or when for some reason System 2 fails to detect an error—System 1 intuitions are let stand.

This model, which successfully accounts for much everyday cognition, has also proven helpful to understand various aspects of superstitious thinking (see Fig. 2)—including notably superstitions’ formation, where, as we’ve seen, System 1 heuristics have been taken to be at work.

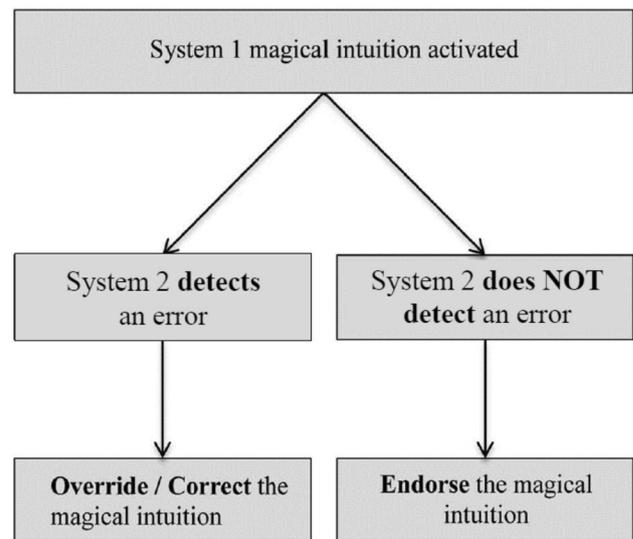


Fig. 2 The ‘corrective dual-process model’ of superstitious thinking. Adapted from Risen (2016), p. 184

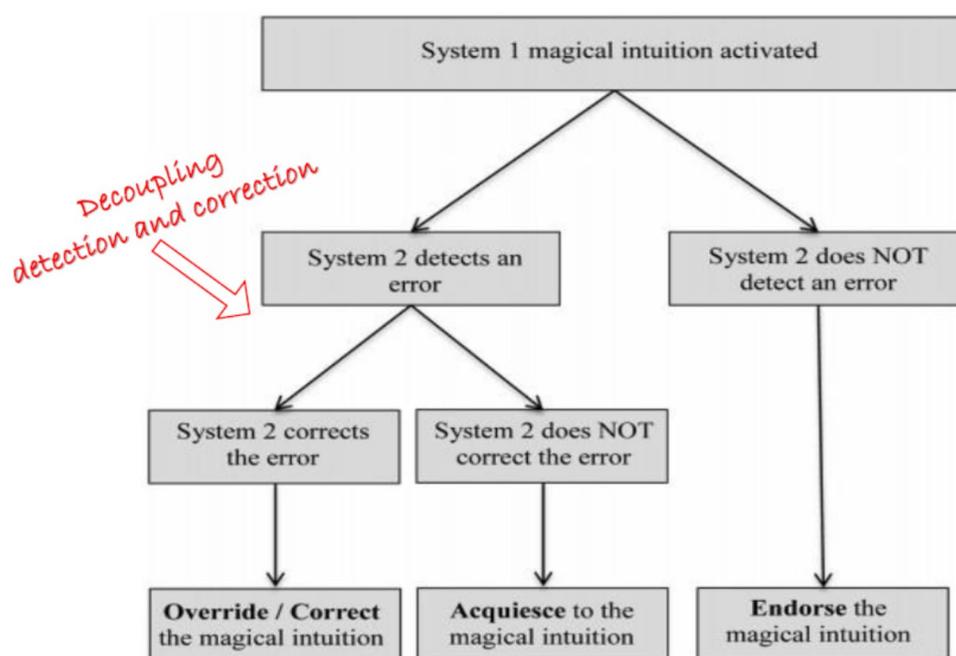
Yet, the model *as it is* seems unable to explain the characteristic ‘acquiescence’ by which superstitions are maintained even when subjects are aware of their incorrectness. This is because the model rests on the unstated assumption that, in the interactions between the two systems, *detection implies correction*: when people, through the rational analysis of System 2, detect a mistaken superstitious intuition generated by System 1, they will *ipso facto* correct it, abandoning/amending the relevant superstition. Although this might indeed happen in some cases, we have seen that in many other cases it doesn’t: we often *acquiesce* to acknowledgedly ill-grounded superstitions.

Hence Risen’s conclusion that superstitious thinking, and superstitious acquiescence in particular, require us to revise dual-process accounts, by ‘decoupling’ detection and correction (see Fig. 3): “A modified dual process account that decouples detection and correction can help explain why superstitious beliefs are maintained even when people know they are not true. System 1 produces a magical intuition. System 2 recognizes the error, but *acquiesces* to the initial intuition. And, when this happens, people may find themselves simultaneously believing and not believing: System 2 knows that the intuition is an error, but lets it stand. Most simply, decoupling detection and correction recognizes the possibility that people can detect an error, but *choose* not to correct it.” (Risen 2016, pp. 194–195).

### 3.3 The Non-doxastic Nature of Superstitious Attitudes

This decoupled dual-process account speaks against the doxastic nature of superstitious attitudes; or, at least, of all

**Fig. 3** Jane Risen’s ‘decoupled dual-process model’ of superstitious thinking. Taken from Risen (2016), p. 194



the instances of such attitudes that involve *acquiescence*, which is a clearly non-doxastic phenomenon.<sup>7</sup> Indeed, it is a widely acknowledged feature of beliefs that, precisely because they are in some relevant sense dependent on evidence and truth, they are not the sort of states that one can consciously choose to hold (see Williams 1973 for a classic statement of this view). And a corollary of this seems to be that they are not the sort of states that one can consciously ‘choose to let stand’ in the way described by Risen with respect to acquiescence.

This talk in terms of a ‘choice to let superstitions stand’, note, is not meant to deny that it might be difficult for superstitious people to do otherwise. Indeed, Risen herself emphasises that superstitions are ‘powerful intuitions’ strongly supported by our motives and emotions; hence that acquiescing to them is a natural and often compelling inclination. The power of such inclination will vary—ranging from cases where it is relatively easy to override it, to cases of proper compulsion, for which one might even argue that acquiescence would be better described as a ‘surrender’ rather than as a ‘choice’.<sup>8</sup> But even in these latter cases, what characterises acquiescence is the fact of depending entirely on the said motives and emotions, in *open* conflict with consciously

recognised evidence and reasons. It is the self-conscious status of this conflict with evidence and reasons that makes it hard to see acquiescence as a doxastic endorsement, which is a form of endorsement crucially characterised by a commitment to truth.

This is why, on the whole, Risen’s talk of “superstitious beliefs maintained even when people know they are not true” sounds very much like a contradiction in terms, suggesting that the attitude she is referring to would be better described in non-doxastic terms.

What non-doxastic terms, precisely? Once again, imagination looks as an appealing candidate—given that, differently from beliefs, imaginings are precisely the sort of things that one can (decide to) entertain against all her acknowledged evidence and reasons. Moreover, and relatedly, superstitious attitudes resemble imaginings in their circumscribed inferential and motivational outputs. Indeed, people fail to draw many obvious consequences of their superstitions, which remain somewhat inferentially isolated—thereby also failing to motivate many of the actions that we would expect if one really held corresponding beliefs (which, vice versa, would be inferentially connected with other background beliefs).

That said, highlighting functional *similarities* with imaginings is admittedly not enough to conclude that superstitious attitudes simply *are* imaginings. This conclusion would require more argument, considering also other candidate non-doxastic categories—like Gendler’s (2008) ‘aliefs’; or Van Leeuwen’s (2014) ‘credences’ (introduced to characterise religious attitudes—which can indeed be seen as

<sup>7</sup> But note that also instances of superstition which are not a result of System-2’s deliberate acquiescence, but rather of its failure to detect System-1’s ill-grounded intuitions, are arguably non-doxastic, due to their motivational origins and inferential isolation.

<sup>8</sup> In relation to cases of this latter sort, see Einstein and Menzies (2004) and Vyse (2014), Ch. 6 on superstitions occurring in obsessive-compulsive disorder.

close relatives of superstitions).<sup>9</sup> For the sake of the present discussion, however, it's enough for me to have shown that superstitions do *not* respond to evidence and reasons as genuine beliefs do, hence are better understood in *non-doxastic* terms. Leaving a positive account of what precisely such non-doxastic terms are for future research, here I shall rather address a challenge to my anti-doxastic view—a challenge which, in turn, will lead me to discuss its important normative implications.

### 3.4 Normative Implications

Although the idea that we can't believe something whilst explicitly recognizing it to be false is widely accepted, there have been some recent attempts to question it. Here I shall focus on one of them, by Huddleston (2012), which is especially relevant to my discussion—not only because it is built precisely around an example of superstitious attitude, but also because it introduces some important normative considerations that it is now time to address.

Against what he recognizes as “the orthodox position”, Huddleston argues for the possibility and existence of what he calls “naughty beliefs”: beliefs held “contrary to one’s *consciously-apparent* knowledge of what the facts are” (Huddleston 2012, p. 213). His chief example is his own superstitious ‘belief’ that raising wine glasses in a toast without looking into your partner’s eyes will bring seven years of bad sex—which he explicitly recognizes to be ill-grounded and at odds with much of what he believes.

Interestingly, Huddleston grants that a ‘belief’ of this sort is “of a particular and unusual kind”, different from most of his other beliefs in that it is not evidence-sensitive/regulated for truth, and it is inferentially isolated (Huddleston 2012, p. 210). Hence, he grants that, from a *functional* point of view, his superstition would be better described in non-doxastic terms—such as, he suggests, either imaginings or aliefs (Huddleston 2012, pp. 216–218). Yet—he argues—similar non-doxastic descriptions should be rejected on *normative* grounds. Indeed, if we classify superstitions non-doxastically, we cannot criticize them as irrational: “we get the *misleading* implication that they float free of the truth norms associated with ordinary belief” (Huddleston 2012, p. 217; emphasis mine).

But why take such implication to be misleading? Huddleston does not provide any reason for this conclusion, which is less obvious than it may seem. Although the view that superstitions are irrational is admittedly rather common, indeed, this does not mean that it should be uncritically

accepted. Vice versa, what we have seen so far provides good reasons to question it for a large class of superstitious attitudes.

Insofar as superstitions are not produced by evidence-tracking and reason-sensitive mechanisms, indeed, it is not clear in what sense they could even be *supposed* to track evidence and reasons, hence on what grounds we may want to criticise them for *failing* to do that. After all, something can arguably only fail at tasks that it is in some relevant sense supposed to perform. Far from being epistemically faulty, then, superstitions do not seem to be even epistemically assessable. Moreover, due to the characteristic inferential isolation that I noted above, superstitions do not typically lead us to adopt a cascade of new ill-grounded beliefs. Hence, they do not seem to involve serious risks of epistemic degeneration, either—thereby further undermining any ground for epistemic criticism.

One here might reply that what makes superstitions irrational—and, more generally, liable to epistemic evidence norms—is their motivational role in action-guidance: for the best way for a mental state to guide action successfully is to represent reality accurately, and representing reality accurately requires evidence-sensitivity (for an argument along these lines, see Bayne and Hattiangadi 2013, p. 130).

However, the link between successful action and evidence-sensitivity is less tight than this argument presupposes. Psychological research in a variety of domains provides examples in which acting upon factually inaccurate representations is (not by virtue of sheer coincidence) conducive to successful outcomes. And, importantly, superstition is one of such domains. As Vyse (2014, p. 11) observes: “One of the most exciting recent findings was a clear demonstration that belief in luck can improve your performance at a skilled activity. In a 2010 study that became an instant classic, Damisch and colleagues showed that participants were better at putting a golf-ball into a cup if they were told their ball had been lucky that day than if they were merely handed a ball”. In line with this, and in open contrast with the purely negative assessment of superstitions characteristic of much traditional research, the idea that superstitious actions bring about a number of pragmatic benefits (social, as well as psychological ones), and that those benefits might translate into better performances, is a recurrent theme in the recent literature (see e.g. Lindeman and Aarnio 2007; Hood 2012).

Admittedly, this in principle would be compatible with what Bortolotti (2017) calls a ‘Trade-Off View’, according to which superstitions’ pragmatic benefits obtain *in spite* of their epistemic faults. But insisting on the existence of such faults at this point would be question-begging: what I am pressing Huddleston and his allies into providing, is precisely a reason to accept the claim that superstitions ‘ought’

<sup>9</sup> I have actually argued against alief-based explanations of superstitious attitudes—and against the psychological reality of aliefs more in general—in Currie and Ichino (2012).

to be evidence-sensitive—and we have just seen that the fact of guiding action is not, in itself, sufficient for that sake.

Importantly, this does not mean that there is nothing whatsoever wrong with superstitions, and superstitious agents are always immune from criticism.

For one thing, the non-doxastic nature of superstitious cognitions may not always be recognized as such by superstitious agents, who may therefore commit meta-cognitive mistakes, forming false meta-beliefs about their own superstitious attitudes. Such mistakes will arguably be the norm with superstitions resulting from System 2's failure to detect the ill-groundedness of System 1 intuitions. But they may also occur in cases of superstitions resulting from System 2's acquiescence—given that, strictly speaking, acquiescence only involves an awareness of the ill-groundedness of one's superstitions, which is not necessarily accompanied by the further awareness of their non-doxastic status. This is then a first way in which superstitious agents might be criticised even if their superstitious attitudes as such are not faulty: what may be faulty is their self-knowledge of such attitudes (with significant practical consequences: a failure to recognise superstitious attitudes' non-doxastic nature may be what leads subjects to take such attitudes 'too seriously', allowing them to guide action *in the place* of their beliefs, rather than *alongside* them<sup>10</sup>).

Moreover, and importantly, to say that superstitions are not beliefs and therefore do not undergo *epistemic* normative constraints, does not mean that they do not undergo normative constraints *of any sorts*. Quite the opposite: it means that the question of superstitions' rationality should be posed afresh—investigating their normative status on new, non-epistemic grounds.

Among the relevant non-epistemic norms that superstitions will undergo there are arguably *prudential* norms—norms having to do with promoting our well-being and inclusive fitness. This is how, for instance, stress-inducing superstitions like those characteristic of some obsessive-compulsive disorders (e.g. 'If I don't turn the light on and off seven times someone I love will die'), or idiosyncratic superstitions leading to social alienation (like some bad-luck-related superstitions), can be criticised:<sup>11</sup> even when their non-doxastic status is properly recognized by the agents who hold them,<sup>12</sup> such superstitions could be

charged with prudential irrationality. Besides, superstitions might also undergo *moral*—and possibly also *political* or *social*—norms: especially insofar as entertaining superstitious attitudes is a matter of deliberate acquiescence, indeed, subjects seem to be morally and socially responsible for holding them—and criticisable accordingly.

What precisely these non-epistemic norms that apply to superstitions look like and how they are related to each other (as well as, possibly, to other norms that I haven't so far identified) is a complex question, which here I was only aiming to raise, and not to exhaustively address. My point here was to suggest that, rather than grounding the *reductio* of non-doxastic accounts of superstitions envisaged by Huddleston, the normative implications of such accounts should be the springboard for further research in the (arguably yet underexplored) area of non-doxastic normativity.

#### 4 Non-doxasticism Beyond Superstitions

In the previous two sections I have argued, respectively, that superstitions can be understood as forms of confabulation, and that they are often entertained in a non-doxastic way. The question now naturally arises of whether and to what extents my anti-doxastic arguments may also apply to more paradigmatic, non-superstitious, forms of confabulation.

This question has so far received less explicit attention than the parallel question concerning neighbouring territories such as delusions or self-deception, where the debate between doxasticists and non-doxasticists has been live. In the confabulation literature, neutral talk in terms of confabulatory 'claims' or 'explanations' is fairly common—the underlying, and often unargued, assumption being that such claims or explanations are genuinely believed by the subjects who offer them.<sup>13</sup>

What I said so far, though, urges us to reconsider such an assumption more critically. Given the large variety of states that have come to be labelled as confabulations, a comprehensive investigation of their doxastic status would require much more space than I've left. But I shall conclude by outlining some reasons to think that what I argued about the formation and maintenance of superstitious attitudes may apply also to the formation and maintenance of (at

<sup>10</sup> Unwarranted superstitious actions will include verbal reports of superstitious attitudes *as beliefs*—which may be especially dangerous from a social-epistemological point of view, misleading other people into taking such reports as reliable testimonies.

<sup>11</sup> On superstitions in obsessive-compulsive disorder, see Einstein and Menzies (2004); on socially alienating superstitions, see Vyse 2014, p. 65.

<sup>12</sup> As it may happen with patients with a 'good of fair insight into their OCD condition'—see DSM V (American Psychiatric Association 2013).

<sup>13</sup> In fact, this assumption has been questioned more often for confabulations than for delusions—and some have even argued that this is where the two phenomena come apart. As Langdon and Bayne (2010, p. 325) observe: "a common objection to a conception of confabulation as delusion-like is that confabulations [but not delusions] are fanciful narratives that are not truly believed nor acted upon". That said, however, talk in terms of confabulatory 'beliefs' remains fairly common.

least some instances of) more paradigmatic confabulatory attitudes—thereby casting doubts on their doxastic nature.

This is easily shown with respect to *formation*—given that, as we have seen, paradigmatic confabulations are traced back to the same sorts of motivational, non-evidence-tracking mechanisms that are at play in superstitions' formation. Given what we have also seen about the essential connection between belief and evidence, this is already enough to put pressure on doxastic accounts of confabulation.

Further pressure in this sense comes from considerations about confabulation's *maintenance*—although the parallel with superstitions in this respect is less straightforward, and requires more careful discussion. The relevant question here is whether there is anything like 'confabulatory acquiescence', corresponding to the 'superstitious acquiescence' described by Risen. Admittedly, if there is, it is less glaring than in cases of superstition. However, as Risen (2016, p. 200) observes, whilst acquiescence is brought into sharpest relief in cases of superstition, "we may be able to export the lessons we learn from superstition (...), to study-cases in which acquiescence is more difficult to detect". Some instances of confabulation, I shall argue, are cases of this sort.

Interestingly, among the domains that Risen identifies as potentially involving acquiescence outside the domain of superstition, there is moral reasoning—which has also been identified as a characteristic domain of confabulation (see Sect. 1.2 above). In moral dumbfounding studies like the one by Haidt (2001), for instance, participants stick to their moral intuitions (e.g. *Sex between siblings is always wrong*) even when they explicitly recognize them to lack rational grounds (as Haidt (2001, p. 814) observes: "many people said something like: *I don't know, I can't explain it, I just know it's wrong*")—which is precisely what acquiescence amounts to. Hence Risen's conclusion that "Haidt's Social Intuitionist Model of moral reasoning (...) could be considered a model of moral acquiescence" (Risen 2016, p. 200).

To be sure, what subjects are acquiescing to in these cases is their original moral intuition, and not the confabulatory explanations by which they subsequently seek to defend it—explanations which, vice versa, they seem pretty ready to abandon when challenged with contrary reasons (i.e., when they are 'reminded' that the story they read explicitly contradicts their explanations). So, I am not presenting these cases as examples of *confabulations* maintained via acquiescence.<sup>14</sup> However, I take what happens in these

<sup>14</sup> To my knowledge, indeed, the original moral intuition in itself is not typically classified as confabulatory, arguably because it does not obviously instantiate the *ill-groundedness* characteristic of confabulation. Indeed, the question of whether a deeply entrenched feeling of moral disgust may be a *legitimate ground* for our moral intuition about the wrongness of incest is controversial (as it is controversial, more generally, the question of what ultimately grounds our moral judgements).

cases to suggest a more general hypothesis, according to which acquiescence might be involved in so-called 'primary confabulations'—the confabulations originally produced in response to a challenging question (or also in the absence of external prompting), which are typically stubbornly maintained in the face of counterevidence and counterarguments; differently from the 'secondary-confabulations' by which those primary ones are subsequently justified, which are less stubborn and more amenable to revisions.<sup>15</sup>

The idea that the stubbornness of primary confabulations may be determined by (more or less deliberate) acquiescence fits well with Fiala and Nichols 2009's suggestion according to which—contrary to the mainstream view that confabulation involves 'not knowing that one doesn't know'—confabulators are typically consciously aware of the poor evidential grounds of their initial confabulatory responses, which they indeed precede by open acknowledgements of ignorance and incertitude (see Sect. 2.1 above). A close look at experimental reports, Fiala and Nichols observe, reveals many confabulatory responses to self-knowledge questions (like 'why did you do X?' or 'why do you believe Y?') to begin with explicit admissions like 'Gee, I don't really know...' (Fiala and Nichols 2009, p. 145, quoting; Nisbett and Wilson 1977, p. 237). Insofar as this is true, it seems that confabulators' do actually 'detect' the ill-groundedness of the responses that they are about to offer, but they nonetheless decide to 'let such responses stand', offering them anyway; in other words, they *acquiesce* to their initial confabulatory responses.<sup>16</sup>

According to Fiala and Nichols, similar cases of confabulations accompanied by explicit avowals of ignorance are not exceptional, but are rather "typical (...) across many different research paradigms in which people confabulate" (Fiala and Nichols 2009, p. 245). Their examples include confabulatory reports in classic misattribution studies, in dissonance reduction studies, in post-hypnotic amnesia, and in split-brain conditions. Whilst granting that more systematic research is needed to determine the precise dimensions of the phenomenon, then, we seem to have good reasons to postulate the psychological reality of 'confabulatory acquiescence' in several cases of primary confabulations.<sup>17</sup> And

<sup>15</sup> Even if, possibly, revisions in favour of equally ill-grounded claims. On these and other differences between primary and secondary confabulations, see Turner and Coltheart (2010); Langdon and Bayne (2010).

<sup>16</sup> Plausibly, they do that because they think it is better to provide an ill-grounded response than to provide none (cf. Coltheart 2017, pp. 65–66, discussed in Sect. 2.3 above).

<sup>17</sup> Fiala and Nichols acknowledge that their claims are mostly based on anecdotal evidence. On the other hand, they point out that the same is true of the mainstream claim about confabulators' unawareness of their ignorance, which has never been systematically investigated, either. As far as I know, Fiala and Nichols' plea for a more systematic empirical approach to this issue has not been taken up by anyone yet: I don't know of any study specifically investigating con-

then, given the non-doxastic nature of acquiescence itself, good reasons to recognize the non-doxastic nature of such confabulations.

As I suggested, moreover, a non-doxastic approach may also be defended for at least some secondary confabulations, due to their motivational, evidence-insensitive origin (as well as, relatedly, to their limited role in guiding inference and actions, which has been pointed out by various authors—cf. Langdon and Bayne 2010, p. 325).

Although still speculative, these non-doxastic hypotheses deserve serious empirical and philosophical consideration—not least because of the normative implications that, as my discussion in Sect. 3.4 indicated, would follow insofar as they were confirmed, putting into question common assumptions about confabulations' epistemic irrationality. These implications would be especially relevant for the burgeoning research program concerning confabulation's 'epistemic innocence', that seeks to investigate the extents to which confabulations might be not only pragmatically, but also epistemically, beneficial (see Bortolotti and Sullivan-Bissett 2018).

## 5 Conclusion

I have defended two main claims: (1) superstitions share key features with confabulations; (2) many instances of superstitious attitudes are non-doxastic in nature. I have argued that (1) is important also because it suggests that research programs in the psychology of superstition and of confabulation, which so far have gone on in almost total disregard from each other, can fruitfully interact. Then, in defending (2), I have provided an example of fruitful interaction: I have shown that recent psychological accounts of superstitious attitudes reveal their non-doxastic nature, and argued that similar accounts could apply to other forms of confabulations—in both cases, with significant normative implications. This latter part of my argument has been by and large programmatic, outlining avenues for future research.

**Acknowledgements** Early versions of this paper were presented in May 2018 at the Philosophy Work-In-Progress Seminar of the Van Leer Jerusalem Institute and at the Workshop on Confabulation and Epistemic Innocence organised by Elisabetta Lalumera at the University of Milano-Bicocca. I am very grateful to the audiences of both these events for their stimulating and encouraging feedbacks, which helped my ideas on the relationships between confabulation and superstition to get off the ground. I am also very grateful to Marianna

Bergamaschi Ganapini, Kengo Miyazono, and Ema Sullivan-Bissett, for their extremely helpful comments and suggestions on subsequent drafts of this paper. The latest versions of the paper have then been greatly improved by the feedback of two anonymous referees for this journal. I wish to thank them, as well as the editors Lisa Bortolotti and Sophie Stammers, for all their generous and precious work.

## Compliance with Ethical Standards

**Conflict of interest** The author declares that she has no conflict of interest.

**Research Involving Human and Animal Participants** This article does not contain any studies with human participants or animals performed by any of the authors.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- American Psychiatric Association (2013) Obsessive-compulsive and related disorders. In: Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Publishing, Arlington
- Bayne T, Fernández J (2009) Delusion and self-deception. Affective and motivational influences on belief formation. Macquarie Monographs in cognitive science. Psychology Press, Hove
- Bayne T, Hattiangadi A (2013) Belief and its bedfellows. In: Nottelmann N (ed) New essays on belief. Constitution, content and structure. Palgrave Macmillan, London
- Bortolotti L (2017) Stranger than fiction: costs and benefits of everyday confabulation. *Rev Philos Psychol* 3:1–23
- Bortolotti L, Cox RE (2009) Faultless' ignorance: strengths and limitations of epistemic definitions of confabulation. *Conscious Cogn* 18:952–965
- Bortolotti L, Sullivan-Bissett E (2018) The epistemic innocence of clinical memory distortions. *Mind Lang*. <https://doi.org/10.1111/mila.12175>
- Campbell C (1996) Half-belief and the paradox of ritual instrumentalism: a theory of modern superstition. *Br J Sociol* 47:151–166
- Coltheart M (2017) Confabulation and conversation. *Cortex* 87:62–68
- Currie G, Ichino A (2012) Aliefs don't exist, though some of their relatives do. *Analysis* 72(4):788–798
- Currie G, Jureidini J (2004) Narrative and coherence. *Mind Lang* 19(4):409–427
- Einstein D, Menzies R (2004) The presence of magical thinking in obsessive compulsive disorder. *Behav Res Ther* 42(5):539–549
- Fiala B, Nichols S (2009) Confabulation, confidence, and introspection. *Behav Brain Sci* 32:142–145
- Gendler T (2008) Alief and belief. *J Philos* 105:643–663
- Gopnik A (2000) Explanation as orgasm and the drive for causal knowledge. In Keil C, Wilson RA (eds) *Explanation and cognition*. MIT Press, Cambridge
- Guthrie S (1993) *Faces in the clouds: a new theory of religion*. Oxford University Press, New York
- Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev* 108:814–834

Footnote 17 (continued)

fabulators' awareness of the ill-groundedness of their claims. Given the important implications that it would have for our understanding of confabulations' epistemic and normative status, this seems an important topic for future research.

- Hirstein W (2005) *Brain fiction: self-deception and the riddle of confabulation*. MIT Press, Cambridge
- Hood B (2012) *Supersense. From superstition to religion. The brain science of belief*. Constable, London
- Huddleston A (2012) Naughty beliefs. *Philos Stud* 160:209–222
- Kahneman D (2011) *Thinking, fast and slow*. Penguin Books, London
- Keeling S (2018) Confabulations and rational obligations for self-knowledge. *Philos Psychol* 31(8):1215–1238
- Langdon R, Bayne T (2010) Delusion and confabulation: mistakes of perceiving, remembering and believing. *Cogn Neuropsychiatry* 15:319–345
- Lindeman M (1998) Motivation, cognition, and pseudoscience. *Scand J Psychol* 39:257–265
- Lindeman M, Aarnio K (2007) Superstitious, magical, and paranormal beliefs: an integrative model. *J Res Pers* 41:731–744
- McKay R, Kinsbourne M (2010) Confabulation, delusion, and anosognosia: motivational factors and false claims. *Cogn Neuropsychiatry* 15:288–318
- Nemeroff C, Rozin P (2000) The makings of the magical mind. In: Rosengren K, Harris PL (eds) *Imagining the impossible. Magical, scientific, and religious thinking in children*. Cambridge University Press, Cambridge
- Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. *Psychol Rev* 84(3):231–259
- Risen JL (2016) Believing what we don't believe: acquiescing to superstitions & other powerful intuitions. *Psychol Rev* 4:182–207
- Risen JL, Gilovich T (2007) Another look at why people are reluctant to exchange lottery tickets. *J Pers Soc Psychol* 93:12–22
- Risen JL, Gilovich T (2008) Why people are reluctant to tempt fate. *J Pers Soc Psychol* 95:293–307
- Rozin P, Millman M, Nemeroff C (1986) Operation of the laws of sympathetic magic in disgust and other domains. *J Pers Soc Psychol* 50(4):703–712
- Schnider A (2018) *The confabulating mind. How the brain creates reality*, 2nd edn. Oxford University Press, New York
- Sullivan-Bissett E (2015) Implicit bias, confabulation, and epistemic innocence. *Conscious Cogn* 33:548–560
- Turner M, Coltheart M (2010) Confabulation and delusion: a common monitoring framework. *Cogn Neuropsychiatry* 15:346–376
- Van Leeuwen N (2014) Religious credence is not factual belief. *Cognition* 133(3):698–715
- Vyse S (2014) *Believing in magic. The psychology of superstition* (updated edition). Oxford University Press, New York
- Wheatley T (2009) Everyday confabulation. In: Hirstein W (ed) *Confabulation: views from neuroscience, psychiatry, psychology, and philosophy*. Oxford University Press, New York, pp. 203–221
- Williams B (1973) Deciding to believe. In: Williams B (ed) *Problems of the self*. Cambridge University Press, Cambridge
- Wilson TD, Nisbett RE (1978) The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Soc Psychol* 41:118–131