Topoi (2020) 39:189–201
https://doi.org/10.1007/s11245-018-09629-y

Confabulating Reasons

Marianna Bergamaschi Ganapini[1]

Published online: 8 January 2019

Abstract
In this paper, I will focus on a type of confabulation that emerges in relation to questions about mental attitudes (e.g. belief, emotion, decision) whose causes we cannot introspectively access. I argue against two popular views that see confabulations as mainly offering a psychological story about ourselves. On these views, confabulations are the result of either a cause-tracking mechanism or a self-directed mindreading mechanism. In contrast, I propose the view that confabulations are mostly telling a normative story: they are arguments primarily offered to justify one's attitudes, and they are produced by our *argumentative* reasoning mechanism driven by the biological goal of presenting ourselves as good reasoners and as reliable sources of information.

Keywords Confabulation · Reasoning · Reasons · Causes · Basis · Self-knowledge

1 Confabulations

In this paper, I will explore the nature of confabulations that are linked to our common mental attitudes (e.g. belief, emotion, decision) by looking at the mechanism primarily responsible for those confabulations.[1] Roughly, a confabulation (of this kind) is a false and/or ill grounded claim (Hirsten 2005, pp. 33–4; Bortolotti 2017), often prompted by a why-question concerning some mental attitude ϕ, or provoked by some challenge moved against ϕ. These confabulations can be found in both healthy subjects and subjects with various pathologies (e.g. delusions, brain injuries). It is also a defining feature of confabulation that those who confabulate are not being dishonest but they are genuinely confident of the truth of what they say (Hirsten 2005; Coltheart and Turner 2009). Another fairly widely held point is that these confabulations appear in conjunction with the agent's inability to introspectively access the causes that produced ϕ in the first place (Turner and Coltheart 2010).

It is still unclear what mechanism(s) is responsible for confabulations. For both the 'normal' and the pathological cases, the received wisdom is that these confabulations infer—or at least attempt to do so—the *causes* of mental attitude ϕ, when the confabulator lacks introspective access to those causes. The related hypothesis is that these confabulations are either the result of a general cognitive mechanism that pushes us to understand the world in terms of causal relations (Coltheart 2017), or the result of a self-directed mindreading mechanism (Carruthers 2011).[2]

However, I will argue that this approach fails to acknowledge that in many cases confabulations are there not (only) to produce self-reports but mostly to tell a story about how good we are at following rational norms. The narrative confabulations are meant to present is primarily normative (and psychological only in a derivative sense): confabulations are mostly presented to *justify* decisions, beliefs, and actions. More specifically, my main claim is that confabulations are primarily offered as premises to an argument, while they are psychological explanations for attitudes *as a result of* their normative function. On the view defended here, what drives subjects to confabulate is *typically* a mechanism specialized

✉ Marianna Bergamaschi Ganapini
bergamam@union.edu

[1] Union College, Schenectady, NY, USA

[1] Finding an accurate account of the phenomenon of confabulation—which encompasses a large variety of different cases—has proven particularly hard (Bortolotti and Cox 2009). In this paper, I only focus on one type of confabulation, leaving open the question whether it is possible to find one overarching account for *all* types of confabulations.

[2] The literature on confabulation presents a pluralistic account of the mechanisms causally responsible for confabulation (Bortolotti and Cox 2009, pp. 954–5). One thing that has been widely noted, though, is that confabulation is found with or without a pathology (Nisbett and Wilson 1977; Dutton and Aron 1974; Haidt 2001), so there is reason to think that some of the mechanisms involved in the pathological cases may also be at work in the non-pathological ones (Coltheart 2017).

in producing arguments (rather than tracking causal relations per se), namely our argumentative reasoning mechanism (Mercier and Sperber 2011). That strongly suggests that we confabulate—instead of admitting we have or had no reason for believing or doing something—mostly as a result of our reasoning mechanism's biological and social function of showing others that we are reliable agents and informants. Hence, whatever failure of self-knowledge confabulation may ultimately reveal, in my view it is the result of a more fundamental attempt to project an image of ourselves as normative, rational agents.

2 The Standard View: Confabulations as (Mistaken) Causes and Motivations

Beliefs, decisions but also emotions and desires can—when expressed and made available to others—elicit (direct or indirect) questions such as 'why do you believe that?', 'what brought you to make that decision?', or 'why do you feel that way?'. At times, when answering these questions, we confabulate (Hirstein 2005, p. 20).[3] Although these questions (and their answers) are asking for (and providing) reasons for our attitudes, the word "reason" itself has three possible different meanings.[4] The *reasons why* agent A formed attitude ϕ are 'explanatory reasons', namely the causes that influence her ϕing: the reason *why* I desire a Coke is that I just saw a Coke-commercial on TV, and the reason why I am in a bad mood is that I did not sleep well last night. These causes explain my attitudes but are not the reasons in light of which I have them, and they did not motivate me in any way, but worked to form those attitudes in a way that is often outside my control and capacity to introspect. Second, a *normative reason* to ϕ is the consideration, fact or proposition that speaks in favor of ϕing (Scanlon 1998). These are either facts/true propositions supporting the content of one's ϕ or mental attitudes representing those facts/ true propositions. To illustrate, there is a reason for me to (intend to) study hard tonight, namely the fact that I have a difficult test tomorrow. Likewise, the recent raising temperatures and see-levels are normative reasons for us to believe in global warming. My daughter's outstanding piano performance is a reason for me to be proud of her. And so on. Finally, *reasons for which* are—in the most general sense—things that motivate us to ϕ. *Good* motivating reasons are motivating reasons that are also normative reasons. On the so-called "standard view" of motivating reasons for actions (Velleman 1992), motivating reasons are the mental states (desire-belief pairs) that *cause* actions *in the right way* (Davidson 1963).[5] The standard view extends beyond action: it has been argued that motivating reasons for beliefs are causing psychological states too (Turri 2009). My belief that the average surface temperature is rising causes my belief that global warming is real and provides the basis upon which I come to form that belief. Similarly, my belief that you broke my guitar motivates my anger by causing it, while my desire to see you is caused and based on the belief that you are my friend.[6]

Many accounts of confabulations share the view that confabulations are offered as reasons connected to attitudes, but what reasons? The standard view—which, I will argue, is vastly insufficient to explain confabulation—is that confabulations talk about the *psychological causes* of attitudes, in the sense that they are (failed) attempts to explain the origins of one's mental attitudes.

There are, however, two different versions of the standard view: one says that confabulations are offered as bare causal explanations for our attitudes, while the other points out that confabulations are offered as the *reasons for which* we have those attitudes. I will first take a look at these two different claims, and then I will offer objections against them to show that there is a much richer story to be told about confabulations and the mechanism behind it.

2.1 Motivations and Mindreading

On one possible reading of the standard view, confabulations are there to provide the *reasons for which* we made a choice

[3] Confabulations of this kind are very similar to rationalizations (Audi 1985; D'Cruz 2015; Schwitzgebel and Jonathan 2017), although later I will argue that confabulations are not merely forms of rationalization.

[4] This tripartite way of classifying reasons is widely used in the metaethical literature on reasons for action, but also transferable to mental attitudes more generally. See Smith (1994), Alvarez (2007), and Hieronymi (2011).

[5] For a causal account involving intentions rather than the belief-desire pair, see Bratman (1987).

[6] Here I mainly refer to the "standard view" of motivating reasons, but there are alternative views out there. For instance, an alternative view is that my motivating reasons are *not* mental states but facts I know (Hyman 1999). On this view, the impending exam is *my* reason for studying hard tonight. Here I am not talking about my psychology but about a fact that speaks in favor of and explains my decision to study tonight. That said, of course, the agent's reasons might figure in her psychology as the content of her mental states (e.g. I believe *I have an exam tomorrow*), and be used in her deliberating about what to do (e.g. Should I study or go to the movie tonight? Well, I have an impeding exam tomorrow so I'd better study). Alternatively, against the standard view, some have offered a dispositional account of basing in which epistemic grounds need *not* be the causes that produce our beliefs (e.g. Evans 2013. We'll come back to this view below). Finally, some argue that motivating reasons (sometime also called 'operative reasons') are the considerations I take to be or treat as normative reasons (Korsgaard 1996; Scanlon 1998; Schroeder 2007). The impending exam (or that I have an impeding exam) is *my* reason for studying hard tonight, in the sense that it is what I take to be a normative reason for studying hard.

or formed an attitude. For instance, in a recent paper Keeling (2018, p. 3) explains that "[c]onfabulators tend to mistakenly self-ascribe putative motivating reasons in particular, that is, they believe there is a reason for which they formed the attitude." Similarly, Johansson et al. (2006, p. 673) clarify that, "[i]n the choice blindness paradigm participants fail to notice mismatches between their intended choice and the outcome they are presented with, while nevertheless offering introspectively derived *reasons for why* they chose the way they did." Although there is a plurality of views out there offering conflicting definitions of 'motivating reasons', as mentioned above the predominant wisdom is that motivating reasons are causes. Indeed, on the so-called "standard view" of motivating reasons, these are intended as psychological states that are partly causally responsible for why we formed a certain attitude. On this view, then, when the confabulator answers the question "why did you ϕ?" by mentioning the (psychological) factors that motivated her to ϕ, she fails to see the actual origins of her attitude, and instead offers a mistaken picture of the basis of her ϕing. Thus, confabulation is a phenomenon "where participants construct plausible but inaccurate accounts of their own motivations" (Scaife 2014, p. 470).

For those who believe that confabulations are (putative) motivating reasons—offered as the causal basis of one's attitudes—a natural hypothesis is that behind confabulation there are mindreading mechanisms that explain the reasons for which people make choices, feel or believe this or that. [7] Indeed, confabulation is often taken to be a key piece of evidence supporting the view that our self-understanding and knowledge are interpretative rather than introspective, and that we use the same inferential mechanism to understand both others' and our mind (Carruthers 2011). Strijbos and de Bruin (2015, p. 298) describe this view as follows: "[w]hen giving answers to questions about our reasons for action, we […] come up with a folk- psychological story that makes it plausible why the type of action we performed is a reasonable response to the type of situation we faced." On this view, confabulation is simply the result of this interpretative process going awry. For instance, when commenting the results of their famous study on people's choices, Nisbett and Wilson (1977, pp. 248–9) explain: "[w]hen people are asked to report how a particular stimulus influenced a particular response, they do not do so by consulting a memory of the mediating process, but by applying or generating causal theories about the effects of that type of stimulus on that type of response." By "generating causal theories" they specifically mean theories about the types of considerations people find "representative" as motivating reasons (as basis for having certain attitudes, that is), as they explain in the following passage: "subjects may have been making simple representativeness judgments when asked to introspect about their cognitive processes. […] The knit, sheerness, and weave of nylon stockings seem representative of reasons for liking them, while their position on a table does not." Accordingly, confabulations are there to lay out the motivating reasons the agent had to do or believe something, motivating reasons that are concocted by applying folk-psychological interpretations to ourselves.

2.2 Causes and Explanatory Reasons

Not everybody interprets confabulations in this way, though. I mentioned at the beginning that there are two possible versions of the standard view of confabulation. The second version of the standard view is that confabulations are meant to provide the reasons *why* we choose or believe something. Reasons why are the bare causes responsible for the choice or belief in question. Causes can be mental or physical events, but the causes that bring about an action or form our attitudes may not match up with what the agent sees as reasons, or may not connect to those actions or attitudes *in the right way*. As a result, these explanations are not necessarily offered to rationalize but to *explain* and possibly excuse one's choice or attitude.

On this view, the mechanism behind confabulation is an abductive process employed in causal understanding of worldly events. For instance, Coltheart (2017) writes that confabulation is "a consequence of a general property of human cognition that is often referred to as 'the drive for causal understanding'". The epistemic shortcomings of confabulations are thus the product of a misused cognitive mechanism that searches for plausible causes of various events. Coltheart draws from Gopnik (2000) who has argued that generally humans find that lacking an explanation for an event "is, to varying degrees, an unsettling, disturbing, arousing experience […]. Conversely, finding an explanation for something is accompanied by a satisfaction that goes beyond the merely cognitive" (Gopnik 2000, p. 311). Coltheart's suggestion is that confabulations are the result of this human need to avoid saying "I don't know why" while providing abductive explanations for their attitudes and actions. As he explains, the epistemic shortcomings of these confabulated explanations are the result of a motivational factor: "[t]he sense of release achieved by arriving at an answer to that question overweighs any concern about the plausibility

[7] See Cassam (2014), Carruthers (2011), Nisbett and Wilson (1977), for example.

of that answer; it is enough that the answer be abductively legitimate" (Coltheart 2017, pp. 67–8).[8]

In sum, these two versions of the standard view ultimately converge in their description of confabulation as a (i) false and/or ill grounded statement, (ii) made—with no intention to deceive—in a dialogical-communicative context prompted by a question about action or attitude ϕ, and (iii) formed by an abductive or a mindreading mechanism which produces mistaken depictions of one's past mental/physical conditions which allegedly caused (and perhaps still do) one's ϕing. On this standard view, then, the original sin of confabulation is that in confabulation people—who are generally blind to the causes of their attitudes—actively produce a mistaken picture of their minds. More specifically, to cover for an underlying gap in self-understanding, confabulation "misrepresents the actual state of one's mind at some relevant time in the past", state of mind that is offered as the origin of the current attitude now subject to questioning (Strijbos and de Bruin 2015, p. 298). For many, that shows both a failure of introspection and a distorted self-knowledge more generally. It is not merely that we are ignorant of the causes of our attitudes, and that we are also unaware of being ignorant about that: in confabulating we actually come up with a fictitious narrative of how we came to have those attitudes and believe this narrative to be right (Lawlor 2003; Scaife 2014).

3 Against the Standard View

Although quite persuasive, I believe that the standard picture of confabulation is ultimately insufficient to really make sense of confabulations as it is putting too much emphasis on the past causes or motivations that confabulation allegedly misrepresents. Because it relies on this narrow description of the nature of confabulation, the standard view wrongly assumes that confabulations are *primarily* the result of some mindreading or causal mechanism, and that they *mainly* show that we lack self-knowledge.

To see where the standard view fails, it is important to review some key examples of confabulation:

A) Wilson and Nisbett (1978) famously report an experiment in which subjects are asked to choose one among four pairs of nylon stocking pantyhose. After making the choice they were asked "Could you tell me why you chose that one?". Surprisingly, as they failed to realize that the four pairs were identical, "[m]ost of the subjects promptly responded that it was the knit, weave, sheerness, elasticity or workmanship that they felt to be superior" (Wilson and Nisbett 1978, p. 124). According to Nisbett and Wilson, subjects in this study are clearly not aware that their choices are due to positioning effects.[9]
B) In a study conducted by Johansson et al. (2005), participants were presented with a series of pictures, asked to make choices about them, and then they are asked why they had made that choice. In some cases, their decision was changed (e.g. the pictures we swapped), and yet many participants did not notice it, and seamlessly went on to elaborate on the reasons behind their (apparent) choice.
C) In a recent study, Haidt presents to the participants a story in which two siblings make love. When first hearing this story, subjects in the study automatically judge the situation as morally wrong, and when prompted to say why, scramble to come up with reasons for their choice. These reasons are then shown to be weak and unsubstantiated by the story. Ultimately, subjects found themselves dumbfounded but refuse to change their minds. What they seem to ignore is the fact that their moral judgments are the result of emotional processes rather than reasoning.
D) Sullivan-Bissett (2015) tells the fictitious story of Roger, a member of a hiring committee, who fails to invite any of the female applicants for an interview. Roger's decision is—unbeknownst to him and contrary to his expectations—due to an implicit bias against women. When questioned, Roger sincerely explains that he did not select any female candidate because none of them seemed qualified for the job.
E) Frazer and Roberts (1994) studied a subject with Capgras delusion who believed that her son had been replaced by an impostor, while ignoring the real causes of that belief. In the study, she was explicitly asked to explain why she thought there was a difference between her real son and the impostor given that they looked very similar. Instead of answering "I don't know", her answer was that the impostor "had different-coloured eyes, was not as big and brawny as her son, and that her real son would not kiss her" (1994, p. 557).[10]

[8] Similarly, Gazzaniga (1985) believes that the left brain functions as an "interpretative module" which continuously tries to explain the world and our minds in terms of patterns of causes and effects.

[9] See Newell and Shanks (2014) for a new interpretation of the Nisbett and Wilson's study.

[10] Often pathological cases of confabulation are, from a clinical standpoint, the result of memory impairments (e.g. source monitoring issues, confusion about chronological order; see Hirstein 2005) or other neuropsychological disorders (Turner and Coltheart 2010). In all these pathological cases, we may find two types of confabulation: primary and secondary confabulation. Primary confabulation is the production of a false belief due, for instance, to a pathology of some kind (e.g. brain injury, delusion). Secondary confabulations are prompted as answers to a why-question concerning the subject's primary confabulation.

How should we interpret these confabulations? Are they offered as causal explanations or as the reasons for some action or belief? As we saw above, it seems that one popular interpretation is that confabulations are meant as purely causal explanations and are the product of an abductive mechanism that searches for causes of various events. Unfortunately, both these claims are at odds with the examples we saw above. For instance, in his explanation for why he did not choose to interview any female candidate, Roger—who clearly stands by his judgment—mentions what he takes to be his reasons behind that decision. Arguing that he is simply tracing a causal story, overlooks the fact that he is not just giving some explanation: he is offering reasons to justify his claim and choice. In Nisbett and Wilson's study, when the subjects were presented with the possibility that their choice was the result of a positioning effect, they denied it and felt "either that they had misunderstood the question or were dealing with a madman" (Nisbett and Wilson 1977, p. 244). Note, however, that there is nothing wrong with explaining a decision in terms of positioning effects: people at times do offer causal explanations for what they did by mentioning the brute, irrational, inscrutable forces that influenced their decisions. Not in this case, though: in Nisbett and Wilson's study subjects do not seem willing to see what they did as the result of *some* cause. The problem seems to be that the position of an item—although it conceivably *can* be among the brute causes of one's choice—is no *reason* for choosing that particular item over another. Indeed, since subjects here are confident they made a rational choice, to them "it seems outrageous" that their choosing an item "might be affected by its position in a series" (Nisbett and Wilson 1977, p. 252). It thus seems clear that in many cases those who confabulate are also standing by their judgments, and that confabulations should not be interpreted as proving mere causal explanations for choices.[11] Hence, the related hypothesis that confabulations are the outcome of an abductive mechanism will mostly deliver wrong assessments as it expects confabulating subjects to generally offer explanations to bridge a gap in knowledge of causal relations. In contrast, this is not what typically happens and in the examples of confabulation mentioned above, confabulations can hardly be understood as trying to provide mere causal explanations for their attitudes.

[11] For similar worries see Sandis (2015) and Keeling (2018). See Bortolotti (2017) for an analysis of the various possible interpretations of this study.

3.1 Confabulating is More Than (Self-directed) Mindreading

Opposing the causal-explanatory view, a more plausible hypothesis is—as we saw—that the confabulatory episodes A–E don't simply point to the causes but to the causes that are also motivating reasons, namely the basis of choices and attitudes. This is still a primarily psychological view of confabulation, but here confabulations are not just causes: they are causes that rationalize one's attitudes as they are meant to describe psychological states that function as grounds or basis for current ones. There is a mindreading mechanism behind these (mistaken) rationalizations: we make sense of others' responses in terms of folk-psychological generalizations, and it seems that at times we direct that same mechanism toward ourselves. Confabulations are typically poor, self-directed folk-psychological explanations.[12]

Although it is true that confabulations are talking about psychological states that ground current choices and beliefs, this (motivating-reasons-as-a-result-of-mindreading) view wrongly focuses only on a backward-looking, causal approach of this grounding relation, and overlooks the normative role confabulations have. In particular, in the next few pages my goal will be to convince the reader of the truth of the following two claims: confabulations are not always meant to trace the causal development of one's attitudes, and at least in the examples above they are not meant to simply tell us something about our psychology, but they also speak of the normative status of our attitudes. I believe that the view that confabulation is self-directed folk-psychological explanation produced by mindreading fails to make sense of these two claims, and thus we need a better, more comprehensive account of the mechanism behind confabulation.

For starters, the notion of a basis (or motivating reason) adopted by this standard picture of confabulation is too restrictive. The view that confabulations are the result of a mindreading mechanism goes hand in hand with the view—called the "standard view" of motivating reasons—that motivating reasons are mental states that cause (in the right way) other mental states to occur. Not surprisingly, on this view, confabulations are "ill-grounded claims about the causes of [one's own] attitudes and choices" (Bortolotti 2017, p. 235).

It is not obvious to me, however, that motivating reasons need be original causes and, as I will explain in a second, it is not clear that confabulations are always offered as

[12] One need not understand confabulation in terms of mindreading to claim that confabulators self-ascribe motivating reasons. For instance on Cox's (2018) account, confabulators self-ascribe motivating reasons but these self-ascriptions are not the product of mindreading and are not necessarily seen as causes. This approach goes in the direction of the view I will advocate for in the next section, and is not subject to the worries I raise in this section.

depicting those causes. Thus, as we try to understand what confabulations are, we want to adopt a view that does not tie us down to a particular notion of basing: we want a view that allows us to say that confabulations are not always looking at a causal past. More specifically, I worry that if we adopt a strictly causal approach to the notion of basis we will have a hard time explaining the following case of confabulation:

F) Moscovitch (1995) described the case of a 61-year-old with impaired autobiographical memory. The patient had been married for over 30 years and had four children, now all grown-ups. Because of his memory impairment, when asked how long he had been married, he answered, wrongly, "About 4 months". When challenged to explain his belief/claim on the face of the fact that he has all grown-up children, he explains that his children were adopted (which is not true), instead of simply saying "I don't know".

(F) has all the earmarks of confabulation: a false and ill-grounded statement is produced (i.e. "my children were adopted") without any intention to deceive, prompted by a question concerning an attitude held by the subject (i.e. "I have been married 4 months") who, however, ignores the causes that produced that attitude in the first place.

Contrary to what might be happening in other cases of confabulation, in this example, the confabulated claim "my children were adopted" is not offered as a consideration that played a role in forming the subject's belief that his marriage is 4 months old. It is not offered as the past *reason for which* he came to that conclusion as the subject does *not* offer the claim "my kids are adopted" to explain how he came to have the belief that he got married 4 months prior. Hence, neither version of the standard view can make sense of the content of this confabulation: indeed (F) is hard to square within a framework in which confabulations *are* causal motivations. As in many other cases of confabulation, in (F), confabulation is used to fill a gap in the agent's view of the world, only that this time the gap has little to do with the origin of his mental attitudes.[13]

A more plausible hypothesis here is that, in (F), the confabulated claim is offered because it fits into a coherent story offered to back up the other false claim ("we married 4 months ago"). In fact, in reference to this case Moscovitch (1995, p. 229) explains that "secondary confabulations may arise to explain (away) the internal inconsistencies of the primary confabulations that are sometimes apparent even to the patient." In the confabulator's mind, that is, the secondary confabulation "my kids are adopted" supports his primary confabulation "I have been married for 4 months" and undercuts a challenge moved to it (i.e. "how is it possible that you got married so recently given that your children are grown-ups?"). Hence, on his mind the two claims "I have been married for 4 months" and "my kids are adopted" can be seen as supporting each other and thus as *partly based* on each other (Evans 2013, p. 2946).

Unfortunately, the standard causal approach can't explain this form of basing. Luckily, there are views of the basing relation that can make sense of this situation. One of them, for instance, is the 'dispositional account of basing': "to base one belief on another is to be disposed to revise the one upon losing the other" (Evans 2013, p. 2955). Adopting this view, we can say that the subject's belief that he has been married for 4 months is presented by the subject as partly based on his belief that his children are adopted, even if there is no causal relation between these two attitudes.[14]

In the next section, I will present a view of confabulation that makes sense of the idea that confabulations are offered as basis for one's choices or attitudes while leaving it open that they might not be proposed as the causal origins for those attitudes. Before being able to fully explain my view, though, I need to raise a related issue, namely that both versions of the standard view of confabulation see confabulation as offering a psychological story, and therefore they see confabulation as ultimately revealing the shortcomings of our self-knowledge. However, even the version of the standard view that understands confabulations as offering motivating grounds for attitudes, overlooks that confabulations are generally presented as *good* (or proper) grounds. In other words contrary to the standard view, very few confabulations seem in the business of simply offering a rationalizing psychological story: confabulations are usually offered to back up attitudes or choices we endorse, and this is clear in the examples A-F above.

Here is what I mean. Motivating reasons can rationalize even if admittedly mistaken: I can rationalize my action (e.g. I went to the store to buy milk) by explaining it in terms of mistaken mental states that motivated it (e.g. my thought that we were out of milk). In many of the examples of confabulation mention above, however, subjects are not trying to rationalize or make sense of what they did/thought in light of their psychological states or past considerations. They seem often fairly confident that they are in the right and that their reasons for doing or believing still hold, and

[13] The gap-filling nature of confabulation is discussed in, e.g., Berrios and Hodges (2000), Hirstein (2005, p. 30), Wheatley (2009) and Sullivan-Bissett (2015, p. 552).

[14] Alternatively, we could also adopt a definition of basis that is still causal even if it does not refer to the way an attitude or choice was causally *formed*: a psychological state A may *causally sustain* psychological state B even if A did not take part in B's formation (Dretske 1999).

are often completely oblivious to the fact that these statements are false and poorly supported.[15] In other words, those who confabulate are not only trying to offer *some* basis for their actions or attitudes: they are offering reasons that are supposedly *enough* to normatively support those actions/attitudes. Hence, confabulations are not meant to merely rationalize but also to *justify* those attitudes and choices.[16]

If mindreading were the main mechanism behind confabulations, in inferring our mental states as the motivating reasons for what we do or believe, we could be applying folk-psychological generalizations concerning why people react the way they do. For instance, we might explain a certain set of decisions (e.g. choosing a certain career path) as the result of desires of a certain kind (e.g. having a remunerative job). We could do that even if we do not think those desires constitute justifications for those decisions. However, once I turn those folk-psychological generalization toward myself—as the inferential account of mindreading suggests—then I have to take into account what I *took*—at the time my choice was formed—to be a good reason. If I went to the store because I desperately wanted to buy some regular milk (while forgetting that I am severely lactose intolerant and should not drink milk at all), my desire at that time was for me a good reason to go (i.e. a consideration I took to justify my action). Unfortunately, now that I remembered my lactose intolerance, my desire for it is not a good reason anymore. And yet if all I am doing is applying my mindreading mechanism to myself, I would have to conclude that I went to the store to buy some milk. That is, in applying mindreading to myself, I would be mentioning what, at that time, I took to be a good supporting reason for my action, not necessarily what I now take to be a sufficiently good reason.

Confabulations, however, present what subjects see as sufficiently good reasons *now* too—which is not the result you would necessarily get if you were simply applying mindreading and folk-psychology. Confabulations are what agents take to be their reasons for ϕing—and by "reasons" they mean not only what made sense of their ϕing (in the past) but also what actively justifies ϕing (in the present and in the future).[17] And this is portrayed also in the many examples of confabulations above, where confabulations are representing (putative) good reasons or justifications, i.e. facts that favor a certain response, and not (simply) the subject's basis for that response.

The experiment by Haidt and collogues is a good example of that. In that study, subjects were asked why they thought that the two siblings' incestuous act was wrong. As their answers were shown to be weak and unsubstantiated by the story, they kept changing them (and ultimately found themselves incapable of providing a good answer). Here's Haidt's (2001, p. 814) initial description:

> Most people who hear the above story immediately say that it was wrong for the siblings to make love, and they then begin searching for reasons (Björklund et al. 2000). They point out the dangers of inbreeding, only to remember that Julie and Mark used two forms of birth control. They argue that Julie and Mark will be hurt, perhaps emotionally, even though the story makes it clear that no harm befell them. Eventually, many people say something like, "I don't know, I can't explain it, I just know it's wrong."[18]

If subjects were only trying to rationalize their moral judgments offering a causal-motivational account of what brought them to form them, it is surprising that they would change their story when those reasons are challenged: if I come to think P because I believe X, even if I realize that X is no good (as a reason), my causal story won't change. I would have to say something like: "I thought P because I originally believed X, but now I realize P is unsupported, so I don't believe P anymore". And yet, this is the opposite of what happens in this case: in the experiment, subjects refuse to change their minds, and try to resist the objections by offering reasons that are meant to justify, not explain or rationalize, their moral convictions. This suggests that confabulations are there to offer sufficient justification for attitudes and choices, not merely motivating reasons. And this is something the mindreading account of confabulation fails to account for.

[15] Fiala and Nichols (2009) argue that that typically confabulators are not so confident about what they're saying. For contrary evidence, see Johansson et al. (2005) and Hall et al. (2012).

[16] More specifically, in confabulation subjects present *all things considered* normative reasons, reasons that are, in their minds, sufficient to justify their action or attitude. For instance, in Frazer and Roberts' study on Capgras delusion, the subject does not mean to offer *some* reason to support her claim that her son has been replaced by an impostor, but tries to offer sufficient support for that claim (Frazer and Roberts 1994, p. 557).

[17] The claim that confabulations are trying to provide justifications is discussed by Sandis (2015) and Bortolotti (2017). Greene (2014) takes confabulations to be aimed at justifying one's moral judgements.

[18] In his paper, Jonathan Haidt considers the mechanism by which we arrive at moral judgments, asking whether our moral judgments are made intuitively or as a result of reasoning. He claims that moral judgments are made by reliance upon unconscious intuition and that reasoning is merely ex post facto. In line with Haidt's view, Greene argues that "[o]ur automatic settings gives us emotionally compelling moral answers, and then our manual modes go to work generating plausible justifications for those answers" (2001, p. 300). For a dissenting voice offering a different interpretation of Haidt's experiment, see Royzman et al. (2015).

4 An Argumentative View of Confabulation

In the reminder of the paper, I would like to offer a unifying account of the underlying mechanism (and motivation) behind confabulation. The starting point of my analysis is that, in answering questions such as "why did you ϕ?", confabulations are arguments rather than self-reports. That is, in making confabulating statement R, an agent A offers R as part of an argument for her ϕing. Accordingly, my proposal is then that whatever malfunction (e.g. defective source monitoring) may be causally responsible for specific confabulatory occurrences, our reasoning mechanism—whose biological function is arguably not to form correct attitudes, but to argue—is the key element generally responsible for the occurrence of some typical cases of confabulations (both pathological and normal). Accordingly, the view proposed here is that a confabulation is (i) a (usually) false or ill supported set of statements, (ii) offered—with no intention to deceive—as an argument to support some decision or attitude ϕ, (iii) produced by the argumentative reasoning mechanism usually in a situation in which the subject is not aware of the causes that formed ϕ in the first place. I take conditions (i–iii) to be jointly sufficient, although perhaps not necessary, to generate confabulations.[19] I also believe that conditions (i–iii) apply to and make sense of all the examples of confabulation mentioned above.

In the previous section I attacked the view that sees confabulations as simply offering a (faulty) psychological story about the origins of our attitudes. In contrast, I claimed that usually confabulations are offered to play the role of good, sufficient basis for attitudes and choices. In other words, confabulations' primary function is normative rather than psychological. More specifically, confabulations are *arguments*, i.e. sets of "propositions of which one is claimed to follow from the others, which are regarded as providing support or grounds for the truth of that one" (Copi and Cohen 1990, p. 6). To be clear, arguments are not simply sets of propositions connected by inferences, but consciously entertained representations in which "[t]he premises are seen as providing reasons to accept the conclusion" (Mercier and Sperber 2011, pp. 57–8). The premises of an argument, in general, are meant to both justify and ground the accepted conclusion, that is they are meant to show that the conclusion is not only justifiable but is also an attitude that is in fact based on sufficient normative reasons. If, as I claim, confabulations belong to arguments, that means that the conclusion of the argument a confabulation is part of, is the content of the attitude the confabulatory statements are meant to justify and provide a basis for. Of course, as argued above, an attitude can be a basis for another attitude even if it did not take part in its formation. Argumentation is generally intended to provide *ex-post*/doxastic justifications (Goldman 1979), namely justificatory reasons that ground one's conclusions but not *necessarily* explain how those came to be from a psychological standpoint. Don't get me wrong: as we perceive (and want others to perceive) ourselves as rational and reliable, we may offer justifications also as psychological explanations. But I don't think this point is key to explain what confabulations typically are.[20]

Take for instance a case of Frazer and Roberts' study of a subject with Capgras delusion. When her delusion is challenged, the subject confabulates false statements that supposedly speak in favor of believing that her son has really been replaced by an impostor. Oblivious as she is to the ill-grounded nature of those statements, she takes them to be very valid reasons. Has she offered these claims as revealing causal motivators too? Possibly, but the key here is that her confabulated reason-giving is there to epistemically support a claim she takes to be unmistakably true. Had she admitted she had no support for her delusional thinking, she would have been unable to hold her ground and reassert the truth of her delusion. It seems she wants to avoid that at all costs, and so she confabulates. As Turner and Coltheart (2010, p. 350) explain, these types of confabulations are primarily offered to "serve the purpose of defending an initial claim" by providing the premises of an argument for it.

4.1 The Mechanism Behind Confabulation

In the study conducted by Johansson et al. (2005), participants were ready to offer reasons for choices they never made (though they were told they did), and adopt those choices as theirs as a result. This case of confabulation is quite unique as subjects here not only prefer to confabulate rather than to admit they have no reason for their choice (as it happens in other cases), but they also prefer to confabulate rather than second-guessing whether they *ever* made that choice in the first place. I believe this suggests that, at least in some contexts and situations, the need to present oneself

[19] Nothing I say below precludes the possibility that some confabulations *may* be the result of mechanisms other than reasoning. However, I take it that my account can explain examples A to F better than any other account and it offers a comprehensive, general account of the phenomenon of confabulation.

[20] When I offer an argument for an attitude I already hold, its premises, namely the reasons that supposedly justify its conclusion, can also at times function as the causes for why I came to that conclusion in the first place. That said, however, in some cases the premises of an argument are not meant to figure as parts of the causal process leading to the formation of an attitude. As I mentioned before, basis need not be causal. Thus, I can offer an argument to justify a conclusion I hold while admitting that I did not form that conclusion because of those causes.

as having an argument is more important than discerning whether that argument supports a choice we really made (as opposed to a choice *others* think we made). But why do we care so much about showing off our argumentative skills? I believe the answer is: to foster our status as trustworthy communicators, to project an image of reliability and rationality, and possibly to convince others to endorse our views.

To see why, we need to take a little detour into Dan Sperber's account of communication and the theory of reasoning he has recently worked out with Hugo Mercier. According to Sperber, communication is the result of a complicated balance between two conflicting interests: the receiver of information wants to acquire reliable, true information, whereas the sender wants the receiver to trust her. Before accepting the information (or decision) coming from the sender, the receiver will evaluate its content and its source. Meanwhile the sender will offer the receiver reasons to convince her. Since at times informants may try to deceive and misinform, receivers have to have ways to filter out potentially false information. That is why we—*qua* receivers of information—are endowed with mechanisms that check the reliability and plausibility of what we are told. Indeed, there is now mounting evidence that we have a suit of folk-epistemological cognitive mechanisms ("epistemic vigilance") with the function to check the reliability of the communicated information by screening the authority and knowledge of the testimonial source, while evaluating the content of her testimony based on other things we know (Sperber et al. 2010).

When it comes to our role as senders of information, in contrast, we produce arguments to convince others and provide them with reasons for a given conclusion. For Mercier and Sperber, argumentative reasoning is the mechanism that creates these arguments. As they put it, "[t]he mental action of working out a convincing argument, the public action of verbally producing this argument so that others will be convinced by it, […] correspond to what is commonly and traditionally meant by *reasoning*" (Mercier and Sperber 2011, p. 59).

On this view, reasoning is the result of a meta-representational mechanism that evaluates the normative "relationship" between premises and conclusion. The inner workings of our reasoning mechanism are not accessible though introspection, and the mechanism itself delivers its outputs—namely, arguments—as a result of intuitions (Mercier and Sperber 2011). Although these intuitions usually track rational norms, they are often skewed in favor of the reasoners. That is, in reasoning, we come up with claims that look like normative reasons for things we already believe/intend/desire with little interest for whether they in fact support them. In fact, on Mercier and Sperber's view of reasoning, reasoning has a biological function, and its function is not to discover the truth or make sound decisions, but to produce arguments to support *one's* conclusions and to convince others to endorse those conclusions too.[21]

Similarly, it seems quite plausible that confabulations are there to play the role of normative reasons to support the agent's beliefs, choices and actions. And this is compatible with the fact that in many instances of confabulation we offer (putative) normative reasons as sufficiently good basis for a choice, an action or a mental attitude we have. Reasoning is the culprit here, namely it is the mechanism that—in absence of alternative explanations—produces confabulations to project a normative and rational picture of our minds.

4.2 Strengths of the Argumentative View of Confabulation

Why should we accept this view? For starters, the argumentative view has a broader explanatory power than its competitors: it explains what some of its competitors can account for (confabulations are offered as psychological basis for attitudes) while also delivering better descriptions of the actual cases (confabulations are primarily offered as good, sufficient basis for attitudes). In the sense of 'argument' I use here, arguments are meant to offer what philosophers call *ex-post* or doxastic justifications: claims that not only support but also ground one's conclusions. It turns out that each time we offer an argument, the premises are intended as good and sufficient basis for our conclusions. Similarly, confabulations are arguments, in which we primarily offer what we take be good and sufficient grounds for attitudes we have. At times, these grounds could be intended also as the origins for those attitudes (as advocated by my opponent), but the emphasis is on the normative aspect not the psychological one.[22] Because the emphasis is on the normative I believe reasoning is usually responsible for confabulations, not mindreading.[23] This more comprehensive approach is

[21] And Mercier and Sperber are not alone in thinking that reasoning has a social-argumentative function. Their view of reasoning is indeed compatible with Haidt's (2001) social intuitionist model. For Haidt, the function of reasoning (and reasons-giving) is social: it's to convince others and shape their minds and views of morality, while defending those view and avoiding being forced to give them up.

[22] Sperber and Mercier claim that argumentation has both a backward and forward-looking role: "[i]n standard cases of argumentation, […] the same reasons have both retrospective and prospective relevance", namely they can be used to explain what led me to form an attitude while also justifying that attitude (Mercier and Sperber 2018, p. 176).

[23] Some have argued that mindreading itself is not just about describing the causal psychological forces that motivate us, but also importing systems of rational and moral norms (McGeer 2007; Zawidzki 2013) Similarly, Strijbos and de Bruin (2015) advocate for a view of confabulation inspired by the idea that self-attribution is typically the result of (normative) "mindshaping" rather than self-directed mindreading. My account of confabulation is, I believe, compatible with this approach.

preferable as it explains what is going on in the examples of confabulation mentioned above, while its competitors can't do that.

Reasoning and mindreading are not extraneous mechanisms, to be sure. Reasoning is a metacognitive mechanism with some mind-reading knowledge built in, and as such it can explain why in confabulations we talk about mental states grounding other mental states (either as causes and as dispositional grounds). For instance, engaging in epistemic vigilance—which is one of the features of reasoning when directed to others—requires that we are able to understand others' mental states before we evaluate them epistemically. However, when self-directed, reasoning attributes mental states to oneself referring to rational and normative standards for attitudes. As a result, they don't just offer causes or basis for attitudes but good and sufficient basis to justify those attitudes. In other words, the difference between mindreading and reasoning is that mindreading explains and rationalizes attitudes: reasoning tries to prove that these attitudes—when they are mine—are justified and therefore ultimately correct (true, appropriate) attitudes. And since this is what we see happening in many of the examples of confabulation, I believe the argumentative view of confabulation is better equipped than its competitor to make sense of confabulation.

4.2.1 The Contexts of Confabulation

The argumentative view of confabulation explains the *context* of occurrence of confabulations and the fact that confabulations are usually prompted by challenges and questions. These are the right settings for reasoning to kick in and fulfill its function. Of course, this does not mean that we use reasoning only in social contexts. Indeed, reasoning is also adopted preemptively to address possible challenges. Even in cases where nobody is watching, reasoning intervenes to take sides. Hence, we can see confabulation also in situations in which the subject's confabulation is not directed to any audience in particular (Wilson et al. 1989).

In contrast, when the focus is not on an attitude, choice or action of mine, then I have no interest in defending it, and so no confabulation needs to take place. Indeed, when assessing others' claims, choices and ideas, we seem to be are well aware of people's cognitive limits (Pronin et al. 2002), and we know that what drives them is often not rationality but unacknowledged forces, difficult to introspect (Malle et al. 2007). As Keeling (2018, p. 6) recently pointed out to object to the idea that confabulation is self-directed mindreading, "if the same inferences underpin both self- and other-ascriptions, this raises the question of why confabulations follow a certain pattern we do not see in other-cases." That is, it is surprising that when evaluating our own attitudes, all our insightful understanding of the human mind often magically disappears, and confabulation takes its place. Now, Malle et al. (2007) suggests that what drives the asymmetry between first and third-person assessment is a motivational factor: we are prone to offer a flattering portray of ourselves and our choices, that's why we are blind to our possible shortcomings and talk of ourselves in terms of 'reasons'. Based on the view offered here, part of the asymmetry is also due to the fact that the *mechanism* behind the two assessments is different: we often use mindreading when we want to explain others' minds, whereas when we try to *account* for our own actions and attitudes argumentative reasoning usually takes over.[24]

4.2.2 Confabulations' Epistemic Shortcomings

Confabulations are usually false and ill-grounded, and although they are offered as normative reasons for attitudes, they in fact often offer poor support for those attitudes. Remarkably, however, those who confabulate do not feel the need to check and second guess them. The argumentative view easily explains these *epistemic shortcomings*: if the point of reasoning is to convince others and defend oneself, then it is not surprising that reasoning looks for reasons in support of the agent's viewpoint. When evaluating and forming arguments, we fall prey of the well-known confirmation bias (Nickerson 1998, p. 175; Mercier and Sperber 2011). We are mostly prone to accept information that is in line with what we already believe, and are blind with respect to the epistemic shortcomings of claims, views, or studies that support what we take to be true. Now this confirmation bias and the function of reasoning explain, according to Mercier and Sperber, why our reasoning-skills seem so poor. Indeed, there is now a strong body of evidence that shows that humans often do not conform to principles of rationality and don't evaluate arguments based on those principles (Stein 1996). As Mercier and Sperber put it, reasoning "falls short of delivering rational beliefs and rational decisions reliably", and that, "in a variety of cases, it may even be detrimental to rationality. Reasoning can lead to poor outcomes not because humans are bad at it but because they systematically look

[24] When evaluating others' thoughts, ideas and actions in communicative contexts, we use reasoning and epistemic vigilance too (Mercier and Sperber 2011). Also, it is reasonable to expect that, in some cases, we are able to overcome the impulse to confabulate and prevent reasoning from producing (bad) justifications. In Nisbet and Willson's famous study only one subject—a student who was taking psychology courses—mentioned that his choice may have been due to a positioning effect. A possible interpretation of this unique instance, is that the student, being aware of the forces that often influences our decisions, was able to prevent reasoning from taking over.

for arguments to justify their beliefs or their actions" (2011, p. 72).[25]

On the argumentative view of confabulation, confabulation is the result of the biases and failings that plague reasoning more generally. Since reasoning grabs any plausible support for our attitudes that looks like a justification (especially in absence of explanations for those attitudes), it is not surprising that those alleged reasons will be false and poorly supported statements (and thus not real reasons at all). Because reasoning produces confabulations, and because reasoning is not interested in the truth, the result is that confabulations are epistemically problematic, and yet the confabulator seems to miss that.[26]

4.2.3 Why Do We Confabulate?

Many accounts of confabulation see confabulation as driven by a motivational component. This is usually seen as an attitude of some kind (e.g. desire, intention) that explains the content of the confabulation and its occurrence (i.e. why we confabulate instead of admitting ignorance) (Conway et al. 2009; Sullivan-Bissett 2015). In the literature, there have been attempts to explain the motivation behind confabulation and why subjects prefer to confabulate instead of admitting ignorance. They often do so to avoid embarrassment for having gappy memories or inconsistent beliefs (Sullivan-Bissett 2015, p. 557; McKay and Kinsbourne 2010, p. 291), and because "giving a confident answer is socially rewarded and advantageous as opposed to saying ''I don't know'''' (Bortolotti and Cox 2009: 961) while also creating a sense of self and stability (Ramachandran 1996, p. 351). Another suggestion is that confabulation is motivated by a desire, present in every case of confabulation, to fulfill the "obligation to knowledgeably explain our attitudes by reference to motivating reasons" (Keeling 2018).

Contrary to these views, I believe that the motivation behind confabulation is not a desire or intention, but a drive intrinsic to the biological function of our reasoning mechanism (Mercier and Sperber 2011). In the examples of confabulation we saw above, the argumentative reasoning mechanism delivered false and ill-grounded statements, intended to depict the agent's attitudes as supported and/or motivated by good reasons and sound arguments. My hypothesis is that the motivation behind this process is rooted in the biological function of our argumentative reasoning mechanism to prepare ourselves to be better at communication, by enhancing our reputation and influencing others (Mercier and Sperber 2011). In contrast, admitting one sees no clear epistemic or pragmatic value or support in one's choices and claims, is possibly very socially costly. Having good reasons for choices and attitudes and having arguments that are coherent and make sense, allow us to present ourselves as rational agents. Since good reasons are conducive to forming true judgments and making good choices, reasoning attempts to present ourselves as dependable (epistemic and practical) agents. This will likely enhance our reputation, making it possible for us to gain more credibility as communicators which will also enable us to bypass others' epistemic vigilance more frequently, influence them more, and possibly spread our values (Mercier and Sperber 2018).

5 Conclusion

In this paper, I made two related points. I offered a new account of the nature of confabulations, arguing that these are statements generally offered to justify attitudes, choices and actions. Then, I argued that confabulation is not the result of a failing self-directed mindreading because mindreading is not typically involved in confabulation. In contrast, confabulations are the product of our argumentative reasoning mechanism whose function is to advocate for the correctness and rationality of attitudes we already hold. Although it is possible that the phenomenon of confabulation reveals that we often lack knowledge of the real causes of our attitudes, I suspect that this failure in self-understanding is due to a more fundamental *rational* failure: those who confabulate are oblivious to the fact that they are offering faulty reasons to support mental attitudes for which they often lack sufficient justification to begin with.

Acknowledgements I presented an earlier version of this work at the international workshop "Confabulation and Epistemic Innocence" at the University of Milano-Bicocca, and I would like to thank the audience of the workshop for their helpful comments. I would especially like to thank Anna Ichino and two referees of this journal for their valuable feedback on earlier versions of this paper. Partial funding for the research leading up to this publication was provided by Union College's Humanities Faculty Research Fund.

Funding This study was partly funded by Union College's Humanities Faculty Research Fund (Grant No. #4036).

[25] Confirmation bias is especially strong when what we believe or choose has an emotional valence or is key to our identity, as shown by a series of experiments on motivated reasoning (Ditto et al. 1998, 2003). There we see also cases of belief-perseverance, namely the tendency humans have to refuse to give up some of their beliefs and convictions even when they have been proved to be unsupported, irrational or straight out false (Guenther and Alicke 2008, p. 706; Mercier and Sperber 2011, p. 68). I believe belief-perseverance and motivated reasoning explain why even healthy subjects who confabulate are often not willing to give up their own initial beliefs or choices when these really matter to them (see also Bortolotti and Cox 2009).

[26] Although reasoning is already prone to produce epistemically faulty confabulations, I suspect that in some cases, this tendency may be more severe due to the presence of pathologies. In pathological cases, confabulation is an attempt to justify beliefs, emotions or choices that are already extremely odd and implausible.

Compliance with Ethical Standards

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

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Alvarez M (2007) The causalism/anti-causalism debate in the theory of action: what it is and why it matters. In: Leist A (ed) Action in context. De Gruyter, Berlin

Audi R (1985) Rationalization and rationality. Synthese 65:159–184

Berrios GE, Hodges JR (eds) (2000) Memory disorders in psychiatric practice. Cambridge University Press

Björklund F, Haidt J, Murphy S (2000) Moral dumbfounding: when intuition finds no reason. Lund psychological reports Vol 1 no 2. Department of Psychology, Lund University

Bortolotti L (2009) The epistemic benefits of reason giving. Theory Psychol 19:1–22

Bortolotti L (2017) Stranger than fiction: costs and benefits of everyday confabulation. Rev Philos Psychol 9:227–249

Bortolotti L, Cox (2009) 'Faultless' ignorance: strengths and limitations of epistemic definitions of confabulation. Conscious Cogn 18:952–965

Bratman M (1987) Intention, plans, and practical reason. Harvard University Press, Cambridge

Cassam Q (2014) Self-knowledge for humans. Oxford University Press, Oxford

Carruthers P (2011). The opacity of mind: An integrative theory of self-knowledge. Oxford: Oxford University Press.

Coltheart M (2017) Confabulation and conversation. Cortex 87:62–68

Coltheart M, Turner M (2009) Confabulation and delusion: a common monitoring framework. Cognitive Neuropsychiatry 15:346–376

Copi I, Cohen C (1990) An introduction to logic. Prentice Hall PTR, New York

Cox R (2018) Knowing why. Mind Lang 33:177–197

D'Cruz J (2015) Rationalization, evidence, and pretense. Ratio 28:318–331

Davidson D (1963) Actions, reasons, and causes. J Philos 60:685–700

Ditto PH, Scepansky JA, Munro GD, Apanovitch AM, Lockhart LK (1998) Motivated sensitivity to preference-inconsistent information. J Pers Soc Psychol 75:53–69

Ditto PH, Munro GD, Apanovitch AM, Scepansky JA, Lockhart LK (2003) Spontaneous skepticism: the interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. Pers Soc Psychol Bull 29:1120–1132

Dretske F (1999) Knowledge and the flow of information. CSLI P, Stanford

Dutton DG, Aron AP (1974) Some evidence for heightened sexual attraction under conditions of high anxiety. J Pers Soc Psychol 30:510–517

Evans I (2013) The Problem of the Basing Relation. Synthese 190:2943–2957

Fiala B, Nichols S (2009) Confabulation, confidence and introspection. Behav Brain Sci 32:144–145

Frazer SJ, Roberts JM (1994) Three cases of Capgras' syndrome. Br J Psychiatry 164:557–559

Gazzaniga M (1985) The social brain: discovering the networks of the mind. Basic Books, New York

Goldman A (1979) What is justified belief? In: Pappas D (ed) Justification and knowledge. Reidel, Dordrecht, pp 1–23

Gopnik A (2000) Explanation as orgasm and the drive for causal knowledge: the function, evolution, and phenomenology of the theory formation system. In: Keil JC, Wilson RA (eds) Explanation and cognition, MIT Press, Cambridge

Greene JD (2014) Moral tribes: emotion, reason and the gap between us and them. Atlantic Books, London

Guenther CL, Alicke MD (2008) Self-enhancement and belief perseverance. J Exp Soc Psychol 44:706–712

Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol Rev 108:814–834

Hall L, Johansson P, Strandberg T (2012) Lifting the veil of morality: choice blindness and attitude reversals on a self-transforming survey. PLoS ONE 7:e45457

Hieronymi P (2011) Reasons for action. Proc Aristot Soc 111:407–427

Hirstein W (2005) Brain fiction: self-deception and the riddle of confabulation. MIT Press, Cambridge

Hyman J (1999) How knowledge works. Philos Q 49:433–451

Johansson P, Hall L, Sikström S, Olsson A (2005) Failure to detect mismatches between intention and outcome in a simple decision task. Science 310:116–119

Johansson P, Hall L, Sikström S, Tärning B, Lind A (2006) How something can be said about telling more than we can know: on choice blindness and introspection. Conscious Cogn 15:673–692

Keeling S (2018) Confabulations and rational obligations for self-knowledge. Philos Psychol 31:1215–1238

Korsgaard C (1996) The sources of normativity. Cambridge University Press, Cambridge

Lawlor K (2003) Elusive reasons: a problem for first-person authority. Philos Psychol 16:549–564

Malle BF, Knobe JM, Nelson SE (2007) Actor–observer asymmetries in explanations of behaviour: new answers to an old question. J Pers Soc Psychol 93:491–514

McGeer V (2007) The regulative dimension of folk psychology. In: Hutto D, Matthew R (eds) Folk psychology reassessed. Springer, New York

McKay R, Kinsbourne M (2010) Confabulation, delusions and anosognosia. Motivational factors and false claims. Cogn Neuropsychiatry 15:288–318

Mercier H, Sperber D (2011) Why do humans reason? Arguments for an argumentative theory. Behav Brain Sci 34:57–74

Mercier H, Sperber D (2018) The enigma of reason. Harvard University Press, Cambridge

Moscovitch M (1995) Confabulation. In: Schacter D (ed) Memory distortion. Harvard University Press, Cambridge, pp 226–251

Newell BR, Shanks DR (2014) Unconscious influences on decision making: a critical review. Behav Brain Sci 37:1–19

Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. Rev General Psychol 2:175–220

Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. Psychol Rev 84:231–259

Pronin E, Lin DY, Ross L (2002) The bias blind spot: perceptions of bias in self versus others. Pers Soc Psychol Bull 28:369–381

Ramachandran VS (1996) The evolutionary biology of self-deception, laughter, dreaming and depression: some clues from anosognosia. Med Hypotheses 47:347–362

Royzman EB, Kim K, Leeman R (2015) The curious tale of Julie and Mark: unraveling the moral dumbfounding effect. Judgm Deci Making 10:296–313

Sandis C (2015) Verbal reports and 'real' reasons: confabulation and conflation. Ethical Theory Moral Pract 18:267–280
Scaife R (2014) A problem for self-knowledge: the implications of taking confabulation seriously. Acta Analytica 29:469–485
Scanlon T (1998) What we owe to each other. Harvard University Press, Cambridge
Schroeder M (2007) Slaves of the passions. Oxford University Press, Oxford
Schwitzgebel E, Jonathan E (2017) Rationalization in moral and philosophical thought. In: Bonnefon J-F, Tremoliere B (eds) Moral inferences. Routledge, New York
Smith M (1994) The moral problem. Blackwell, Oxford
Sperber D et al (2010) Epistemic vigilance. Mind Lang 25:359–393
Stein, E. (1996) Without good reason: the rationality debate in philosophy and cognitive science. Clarendon, Oxford
Strijbos D, de Bruin L (2015) Self-interpretation as first-person mindshaping: implications for confabulation research. Ethical Theory Moral Pract 18:297–307
Sullivan-Bissett E (2015) Implicit bias, confabulation, and epistemic innocence. Conscious Cognition 33:548–560
Turner M, Coltheart M (2010) Confabulation and delusion: a common monitoring framework. Cogn Neuropsychiatry 15:346–376
Turri J (2009) The ontology of epistemic reasons. Nous 43:490–512
Velleman D (1992) What happens when someone acts? Mind 101:461–481
Wheatley T (2009) Everyday confabulation. In: Hirstein W (ed) Confabulation: views from neuroscience, psychiatry, psychology, and philosophy. Oxford University Press, New York
Wilson TD, Nisbett RE (1978) The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. Soc Psychol 4:118–131
Wilson TD, Dunn DS, Kraft D, Lisle DJ (1989) Introspection, attitude change, and attitude-behavior consistency: the disruptive effects of explaining why we feel the way we do. In: Berkowitz L (ed) Advances in experimental social psychology, vol 22. Academic Press, Elsevier, pp 287–343
Zawidzki TW (2013) Mindshaping: a new framework for understanding human social cognition. MIT Press, Cambridge