



# Introduction: Philosophical Perspectives on Confabulation

Sophie Stammers<sup>1</sup> · Lisa Bortolotti<sup>1</sup>

Published online: 28 August 2019  
© Springer Nature B.V. 2019

## 1 Why Confabulation?

The notion of confabulation has received considerable attention in the psychological and clinical literature, where confabulation (from the Latin ‘fabula’, story) is considered either as a symptom of psychiatric disorders featuring severe memory impairments (*narrow confabulation*) or the more general tendency people have to provide explanations for their choices and attitudes (*broad confabulation*). What is in common between the two notions of confabulation is that the story people tell is usually false, and not grounded in reality, even though there is no intention to deceive.

Confabulation is tightly linked to some core notions in the philosophy of mind and the philosophy of psychology, such as those of self-knowledge and rationality, and raises interesting epistemological questions about the coherence, accuracy, and justification of our reports. However, until recently and with some exceptions (Hirstein 2005; Bortolotti and Cox 2009; Strijbos and de Bruin 2015; Scaife 2014; Sandis 2015; Sullivan-Bissett 2015; Bernecker 2017; Bortolotti 2017), the notion of confabulation has been largely underexplored in philosophy.

This special issue aims to show that the phenomenon of confabulation is not only fascinating in its own right but also a powerful source of inspiration for philosophers with a range of interests and preferred methodologies. As part of ERC-funded project PERFECT, we created some opportunities for philosophers to meet and discuss recent work on confabulation. In May 2018, we organized a workshop in Oxford to bring together philosophers whose original research had already shed light on the nature of confabulation. A few weeks later, PERFECT sponsored a workshop on

the costs and benefits of confabulation organized by Elisabetta Lalumera at the University of Milan Bicocca where new papers were presented.

With the special issue, we intended to encourage further work in this area. Here we are happy to be able to introduce the nine papers we gathered, addressing the relationship between memory and confabulation; the role of confabulation in self-interpretation; the costs and benefits of confabulation; and the effects of confabulation on superstitious thinking, moral decision-making, and gaslighting.

## 2 From Memory to Confabulation

What is the relationship between memory and confabulation? In their contributions, Robins and Michaelian apply their theories of how memory works to cases of confabulation involving misremembering.

*Sarah Robins* discusses the relationship between mnemonic confabulation, a narrow form of confabulation focused exclusively on memory impairment, and broad confabulation, the form of confabulation that is widespread in the non-clinical population and includes the offering of ill-grounded explanations. First, Robins offers an account of mnemonic confabulation, situated with a causal theory of memory. On the causal theory of memory, in order for a person to remember the target event, say, the target event must have happened, the mental representation of the event must be accurate, and the content of the representation must have been produced in the right way, in that there needs to be a causal connection between the target event and the subsequent representation of it. Confabulation about an event means that either the event never occurred or there is no connection between the event as it happened and the mental representation of it. Second, Robins examines the features of broad confabulation, which includes explanations that are ill-grounded and fill a gap in a person’s knowledge about how certain relevant events occurred. Whereas mnemonic confabulation is false, broad confabulation needs not be false and can be true by accident. Whereas broad confabulation is

✉ Sophie Stammers  
s.stammers@bham.ac.uk

Lisa Bortolotti  
l.bortolotti@bham.ac.uk

<sup>1</sup> Philosophy Department, University of Birmingham,  
Edgbaston B15 2TT, UK

ill-grounded, as the explanation does not rely on evidence, mnemonic confabulation is grounded in the person's feeling that she remembers a certain event. And finally, whereas broad confabulation is something that a person can recover from, by learning what a well-grounded explanation of the relevant event is, people do not recover from mnemonic confabulation. For Robins, thus, there are good reasons to keep mnemonic confabulation out of the discussion of broad confabulation.

*Kourken Michaelian* defends a simulationist theory of confabulation, based on a constructive conception of memory, from objections raised by proponents of the causal theory. We saw that for the causal theory of memory, remembering requires a causal connection between the target event and the representation of such event. When mnemonic confabulation occurs then, either the target event never occurred or there is no connection between the target event and its representation. According to the simulationist theory of memory, instead, to remember does not require an appropriate causal relationship between target event and representation. Rather, memory is a special instance of imagination, and just as imagining an event does not require that the imagination is causally connected to the event, neither does remembering it. For the simulationist, to remember is to imagine in a reliable way. To confabulate, instead, is to imagine in an unreliable way. A philosophical theory of confabulation, for Michaelian, needs to be able to distinguish confabulating from remembering and also from other forms of misremembering. So, in his paper Michaelian provides such distinctions, while responding to challenges raised by causalists against the simulationist approach. Whereas Robins, as we saw, distinguishes confabulation from remembering on the basis of the absence of the causal connection between target event and representation, Michaelian elaborates on the notion of reliability that characterizes the process of remembering, and describes in what ways reliability is compromised in cases of confabulation.

### 3 Confabulation, Interpretation and Self-Construction

Does confabulation constitute a failure to know something important about oneself? Two papers challenge traditional views on this issue, with De Bruin and Strijbos focusing on understanding the threat to first-person authority; and Bergamin viewing confabulation as tied up with an interpretive faculty essential to understanding ourselves and our world.

*Leon de Bruin* and *Derek Strijbos* address the view that findings on the prevalence of confabulation show that people lack self-knowledge (e.g., Scaife 2014), and that this undermines first-person authority—the special authoritative status we have when attributing mental states to ourselves,

compared to attributing them to others. De Bruin and Strijbos argue that even on folk-psychological understandings of the mind, expectations of perfect alignment between how we act and what we say about our actions are unrealistic. They advocate for a new understanding of first-person authority which depends not simply on whether there is a gap between our self-ascriptions and our actions, but on a capacity for self-regulation: the extent to which we are able to bridge such gaps by aligning our self-ascriptions with our actions and vice versa. They claim that existing empirical research on confabulation does not address whether or not this capacity is undermined. For de Bruin and Strijbos, self-regulation is an ongoing, socially situated practice, which involves taking on board input from social peers who notice potential misalignment in order to adjust one's self-ascriptions and actions. Success in this domain is dependent on self-directed attitudes of flexibility, openness to new perspectives, inquisitiveness, etc.; whilst failure often involves self-directed attitudes of inflexibility, rigidity, lack of interest, etc. For de Bruin and Strijbos, first-person authority is undermined when the latter self-directed attitudes are present and the former diminished, leading us to systematically fail to re-align our ascriptions with our actions and vice versa.

*Joshua Bergamin* draws on work by Dreyfus (2005) and McDowell (2007). He outlines an interpretive faculty that is essential to human thinking, which takes our immediate, pre-reflective experience as input, and transforms it into conceptual thought. Experience is interpreted in accordance with an existing web of shared cultural beliefs and our own personal narrative. This faculty enables us to construct and to believe complex narratives to explain our actions, even when our reasoning is only incidentally related to the real drivers of our behaviour, such as in the case of confabulation. Bergamin suggests that in so far as this faculty enables us maintain a sense of agency over our choices and actions, it plays an important role in maintaining mental health. However, our interpretive faculty can become overactive, leading our narratives about our experiences and actions to become extremely dissociated from reality. Bergamin argues that is how we should understand schizotypal-spectrum disorders. Whilst it is part of the ordinary operation of the interpretive faculty to search for patterns, Bergamin draws on evidence to suggest that 'schizotypal' minds are particularly prone to interpreting patterns which aren't really there, and argues that this account enables us to explain some of the hallmark sensory symptoms of, for example, schizophrenic psychosis, such as hallucination. Bergamin's account is bolstered by the observation that the content of schizophrenic delusions is often culture-specific.

#### 4 Benefits of Confabulation in the Context of Broader Cognitive Mechanisms

Stories that depart from reality have epistemic costs, but do they have any benefits? Stammers and Bergamaschi Ganapini explore confabulation in the context of broader cognitive mechanisms or tendencies which confer important biological and social benefits.

*Sophie Stammers* compares two accounts of confabulation: *causal* accounts which unify clinical and non-clinical instances of confabulation by modelling both as outputs of a theory-formation system that drives us to produce causal theories; and *narrative* accounts of confabulation, which model the confabulations that arise in cognitive pathology such as dementia as attempts to create and emphasise socially important meanings. Motivated by similarities between everyday confabulation and clinical confabulation, Stammers investigates how such accounts might fit together, and maintains that we can preserve the explanatory appealing features of both. She argues that explanation-giving is a communicative, inherently social practice, and therefore we should expect that it is shaped not only by the advantages it confers in the natural world (providing causal maps); but also by the advantages it confers in the social world. She posits that humans have a tendency to imbue everyday explanations with resonant themes, which render our experiences meaningful. This tendency is developed and maintained through cultural practices, such as the convention of meaningful storytelling which is learned through social-interaction with caregivers, that we are inducted into in early childhood. Stammers holds that this tendency has psychological benefits, enabling us to construct our identity and increase our resilience. It also has social benefits, increasing the pursuit of shared goals and strengthening social bonds. Stammers argues that confabulations in dementia which emphasise socially important meanings should be seen as continuous with non-clinical confabulation, and with everyday explanation-giving behaviour in general, because such behaviour is replete with attempts to emphasise socially resonant meanings.

*Mariana Bergamaschi Ganapini* presents the case against standard views of confabulation, on which confabulations are (mistaken) accounts of the *causes* or *motivations* for action. She holds that these sorts of accounts are wrong to model confabulations as primarily the result of some mindreading or causal mechanism, and, as a result, overlook the normative role of confabulation. For Bergamaschi Ganapini, confabulations not only describe reasons which make sense of some choice or action, but, chiefly, provide (putative) normative reasons for why such action should be taken and which (allegedly) *justify* acting that way in the future. She draws support from empirical

work such as that by Haidt (2001) on the confabulations that can arise when people defend their moral convictions. She argues that even when confabulating subjects are presented with evidence which undermines their claims, they fail to recognise their error and to update their attitudes accordingly, and instead continue to search for reasons that are meant to justify their moral convictions, rather than explain them. She argues that we can best make sense of these findings by holding that subjects are driven to confabulate not by a mechanism specialized in tracking causal relations, but one specialized in producing arguments. Bergamaschi Ganapini follows Mercier and Sperber (2011) in holding that this more general argumentative reasoning mechanism functions to prepare us to be better at communication amongst our peers. This mechanism has the benefit of enabling us to project an image of ourselves as normative, rational agents, which can enhance our reputation and allow us to influence others.

#### 5 Further Applications of the Notion of Confabulation

In the previous sections, we saw how philosophical insights into the notion of confabulation can help us clarify our theory of memory and refine our conception of self-knowledge as interpretation. We also saw that confabulating has many interesting effects, with positive and negative implications for epistemic agency and socialisation. In this last section, we review some previously unexplored connections between confabulation and the thinking that characterizes superstitious behavior, moral decision making, and gaslighting.

*Anna Ichino* examines the relationship between confabulation and superstitious attitudes, arguing that they share some core features and can be explained by appealing to the same psychological mechanisms. Ichino notices that most examples of superstition can be described as confabulations that are about the world (i.e., they are not necessarily about the self), and that are spontaneous (i.e., they are not the response to a request for information). For instance, ‘My wise pen, who attended all the lectures, will help me in the exam’, is a claim that is ill-grounded, is offered with no intention to deceive, fills a gap in our understanding of how things work, and responds to the need to identify meaningful connections between events. What kind of mental attitudes are superstitions and confabulations? According to Ichino, the best way to make sense of their functional characteristics, and of how they are formed and maintained, is to regard them as *non-doxastic* attitudes. This is because superstitions and confabulations are not sufficiently sensitive to evidence to be beliefs, and typically fail to integrate with other things we believe. In the end, Ichino considers the various implications of identifying superstitions and confabulations as

related phenomena and as non-doxastic attitudes, both for philosophy and for the cognitive sciences.

*Anneli Jefferson* addresses the relationship between everyday confabulations (those confabulations that do not need to involve any memory impairment and are common in the non-clinical population) and rationalisations of our behavior which tend to present what we do as more acceptable than it actually is. She argues that typically confabulations fill an explanatory gap producing ill-grounded causal stories, whereas rationalisations are instances of motivated cognition producing justifications of behavior whilst also acknowledging that something can be at the same time a confabulation and a rationalization. Next, Jefferson asks whether confabulation and rationalization have the same moral benefits and costs when they are directed towards our behaviour. She identifies some benefits of rationalization in terms of self-enhancement and self-consistency: by helping us to form a positive self-image, and in particular to regard ourselves as good moral agents and as consistent in our moral behaviour, we may be more motivated to act morally in the future in line with our positive self-image. However, there are also risks in rationalizing our moral behaviour. Jefferson warns that, by justifying what we do as what is morally acceptable to do, we gradually revise our judgement about what is morally acceptable, and may not be able to recognize as problematic behaviour that does not meet the requirements of moral acceptability we started with. She concludes discussing whether confabulations have similar effects to rationalisations in this context, and she finds that the effects of confabulations on moral behaviour are less pronounced, whether they are positive or negative, because in confabulation the explanation element is more prominent than the justification element.

*Andrew Spear* argues that the potential for confabulations to count as epistemically innocent depends on the context in which confabulations occur and makes a case for confabulations failing to achieve epistemic innocence in the context of gaslighting. Epistemic innocence applies when confabulation has some significant epistemic benefits that could not be attained without it. Epistemic innocence does not deny that confabulations have epistemic faults, and typically are inaccurate or ill-grounded explanations. After providing some interesting cases of gaslighting and an account of the phenomenon of gaslighting, Spear argues that gaslighting involves confabulation, as perpetrators sometimes confabulate when they are not fully aware of their motives for gaslighting, and victims almost always confabulate when they need to square their own experiences and feelings with what perpetrators say. The confabulation involved in gaslighting does not meet the conditions of epistemic innocence according to Spear. It is true that often the confabulation presents itself as the only explanation for what is happening: for instance, when

victims of gaslighting trust and have a personal investment in the perpetrators, they will be motivated to look for explanations of their apparent epistemic failures that do not contradict what perpetrators say. However, in the context of gaslighting, there are no epistemic benefits in confabulating. The effect on perpetrators is that they seemingly obtain evidence for their victims' epistemic failures and against the legitimacy of the victims' resistance to the gaslighting; and the effect on victims is that their epistemic self-doubt, for which there is no good justification, intensifies.

## 6 Conclusion

In the special issue, contributors situate their arguments in the context of ongoing debates on confabulation in philosophy and cognitive science, as well as proposing further research questions emerging from their theses. We hope that philosophers and cognitive scientists will find this collection to be delivering informative and engaging insights into the recent progress of philosophical discussions of confabulation, and that they will be inspired to pursue further investigations of their own.

**Acknowledgements** For their work on the special issue and for the preparation of this introduction to the special issue, Sophie Stammers and Lisa Bortolotti acknowledge the support of the European Research Council under the Consolidator Grant Agreement Number 616358 for a project called Pragmatic and Epistemic Role of Factually Erroneous Cognitions and Thoughts (PERFECT).

## References

- Bernecker S (2017) A causal theory of mnemonic confabulation. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2017.01207>
- 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
- Dreyfus HL (2005) Overcoming the myth of the mental: how philosophers can profit from the phenomenology of everyday expertise. *Proc Addresses Am Philos Assoc* 79:47–65
- Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev* 108:814–834
- Hirstein W (2005) *Brain fiction: self-deception and the riddle of confabulation*. MIT Press, Cambridge
- McDowell J (2007) What myth? *Inquiry* 50:338–351
- Mercier H, Sperber D (2011) Why do humans reason? Arguments for an argumentative theory. *Behav Brain Sci* 34:57–74
- 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 Anal* 29:469–485

- Strijbos D, de Bruin L (2015) Self-interpretation as first-person mind- shaping: implications for confabulation research. *Ethical Theory Moral Pract* 18:297–307
- Sullivan-Bissett E (2015) Implicit bias, confabulation, and epistemic innocence. *Conscious Cogn* 33:548–560

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.