



Confabulation, Explanation, and the Pursuit of Resonant Meaning

Sophie Stammers¹

Published online: 27 November 2018
© The Author(s) 2018

Abstract

People with dementia sometimes *confabulate*, offering sincere explanations of their situation which are not grounded in evidence. Similar explanation-giving behaviour occurs frequently in the non-clinical population. Some see this as evidence that clinical and non-clinical confabulations emanate from the same essential feature of cognition, a drive to provide causal theories (Coltheart in *Cortex* 87:62–68, 2017). Others maintain that clinical confabulations are not attempts to identify causal relations, but narratives which create and emphasise socially important meanings (Örülv and Hydén in *Discourse Stud* 8(5):647–673, 2006). We can reconcile these accounts, preserving the explanatory appealing features of both. I argue that humans have a tendency to imbue everyday explanations with *resonant* themes: ideas which strike a chord with us and render our experiences meaningful, and I explain how this is compatible with the drive to provide causal theories. Explanation-giving is a communicative, inherently social practice, and so it is not only shaped by the advantages it confers in the natural world; but also by the advantages it confers in the social world, and for related psychological functioning. As such, confabulation is an attempt to emphasise socially resonant themes, whilst also being continuous with everyday explanation giving.

Keywords Confabulation · Explanation · Narrative · Resonance · Dementia

1 Introduction

People with dementia sometimes *confabulate*, offering sincere, non-deceitful explanations of their situation which are not grounded in evidence, and which are often interpreted as part of the cognitive decline of the disease. However, sincere explanations not grounded in evidence also occur in a variety of non-clinical contexts (Nisbett and Wilson 1977, Cohen 2003). Max Coltheart has recently proposed an account (2017) which unifies clinical and non-clinical instances of confabulation, seeing them emanate from the same essential feature of cognition: a drive to explain which generates attempts to provide veridical causal theories, modelled on Alison Gopnik's notion of the theory formation system (henceforth; TFS) (2000). On this account, confabulation in all populations is the normal output of the TFS, an attempt to provide a veridical causal theory to explain a perplexing situation, which is unsuccessful.

Other authors who study confabulation in clinical populations have focused on a different aspect of the phenomenon.

Conway and Tacchi (1996) investigate the confabulations of a person with a traumatic brain injury, while Linda Örülv and Lars-Christer Hydén (2006) examine confabulation in dementia patients, and stress the narrative features of confabulation. They maintain that confabulations seen in these contexts are not attempts to identify causal relations, but narratives which create and emphasise socially important meanings.

How do these accounts fit together? Is non-clinical confabulation ever an attempt to create social important meaning? And can confabulations that play this role also be the output of the TFS? In this paper, I argue that we can reconcile these accounts, preserving the explanatory appealing features of both. Örülv and Hydén (2006) provide a powerful analysis which both explains and predicts the particular character of clinical confabulations, but I suggest that they identify a feature of explanation-giving which is in fact widespread in the non-clinical population as well. I argue that humans have a tendency to imbue everyday explanations with *resonant* themes: ideas which strike a chord with us and render our experiences meaningful, and I explain how this is compatible with the operation of the TFS. Explanation-giving is a communicative, inherently social practice, and so it is not only shaped by the advantages it confers in the

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

¹ University of Birmingham, Birmingham, UK

natural world (namely, providing veridical causal maps); but also by the advantages it confers in the social world, and for related psychological functioning. As such, confabulation can be seen as an attempt to emphasise socially resonant themes, whilst also being continuous with everyday explanation giving.

This result is of interest from a theoretical point of view, because it proposes a new account of explanation, extending the explanatory power of Gopnik's theory, enabling us to explain why, so often, people do not make use of veridical causal explanations. But further, an account which demonstrates that confabulation in clinical contexts is continuous with everyday explanation-giving behaviour is important and valuable for reducing the considerable stigma attached to mental illness, and in particular, dementia. I close by outlining some potential avenues for future research, namely, investigating whether there is a robust effect of using resonant narratives to facilitate the retention of otherwise disparate items in experience.

2 Examples of Confabulation

Let us begin by looking at some events recorded by Örvulv and Hydén (2006). I paraphrase two different perspectives on these events, starting with that of Martha, who is hosting guests in her living room:

LISCENCE (Martha): Seated on the living room sofas, host Martha turns to one of her guests, Catherine, and informs her that the TV belongs to her, and furthermore, that she pays for the licence.

COFFEE (Martha): Martha offers to serve her guests coffee. She makes to pour the coffee, but then establishes that everyone has had coffee already, so there is actually no need.

At least, this is how Martha understands the unfolding events. But further information reveals that something is up. Martha and Catherine have been diagnosed with Alzheimer's disease, and these events in fact take place in their care home, where Örvulv and Hydén are conducting an ethnographic study. Drawing on other contextual clues from their 6-months observation, Örvulv and Hydén have a different understanding of the situation:

LISCENCE (ÖH): Martha has (erroneously) come to believe that she is hosting guests in her own home. She is perplexed by the TV. She would never watch a TV without paying for the licence, but she is able to explain away the perplexity of the situation by concluding that the TV must belong to her, and that she pays the licence after all.

COFFEE (ÖH): Martha offers to serve her guests coffee, but when she makes to pour the coffee, she is unable to operate the unfamiliar coffee pot (which belongs to the care home). She is perplexed that she is unable to serve her guests, but is able to explain away the perplexity by establishing that everyone has already had coffee, and so there is no need for her to serve them.¹

According to Örvulv and Hydén, when Martha explains that the TV is hers, and that everyone has had coffee, she is *confabulating*. That is, she is offering a sincere account of what she takes to be that facts of the situation, but her account is not grounded in evidence, and doesn't accurately reflect the situation.

Many have argued that confabulation is the result of disruption to memory processes occurring in cognitive pathology (Berlyne 1972; Moscovitch 1989). However, people with no known pathologies can exhibit similar behaviour, offering sincere accounts of a situation, which are not grounded in evidence. For instance, Nisbett and Wilson (1977) demonstrate that people tend to choose the item on the right-most side of an array of broadly similar clothes, even when the order of the clothes is randomised. When people explain their choices, they do not cite the position of the item, referring instead to factors like the colour or the fabric quality. Because we see the position effect emerge despite randomisation, Nisbett and Wilson argue that the reasons given by participants are not based on considerations which moved them at the time of decision, yet these reasons are sincerely given. In another study Cohen (2003) provided participants with one political party's welfare policies, but framed them to look as if they were policies of a party on the other side of the political spectrum, then asked people whether they supported the policies. People tended to support policies framed as of their preferred party (thus actually supporting policies of the opposing party), but claimed that they made their choices, guided by their own philosophy of governance, on the basis of the policies' content.

Further to explanations about past choices, other studies demonstrate that a similar effect can occur when participants attempt to make sense of current, unfolding experiences. For instance, Barnier et al. (2010; in Coltheart 2017) demonstrate that healthy participants may be hypnotised with the suggestion that they will see a stranger in the mirror. When participants see their own reflection, they explain that there must be a room on the other side of the mirror, or that the experimenters have found someone who looks like them. Hirstein (2006, pp. 13–14) also reports on a case in

¹ Örvulv and Hydén reach these interpretations of Martha's situation through compiling contextual clues from a number of episodes during their many months of observation, as will be discussed further in Sect. 4.

which a person is hypnotised with the suggestion to lift their arm when they hear the word ‘money’. When they hear the word and raise their arm, they explain that they just felt like stretching.

In the above studies in which people explain a previous choice they made, a factor other than that identified in the participants’ explanations drives choices, and so explanations are not grounded in evidence of what actually brought about choices. But they are nonetheless sincerely given.² In the hypnosis studies, people do not recognise that hypnotic suggestion prevents them from accessing an accurate explanation of their ongoing experience, but this does not prevent them from trying to make sense of what they are presently seeing or doing. So, behaviour similar to that just discussed in the context of Alzheimer’s can also occur without pathological memory deficit.

Recognition of this has led some to argue that the cases above and memory pathology cases share sufficiently many features so as to be addressed together (Hirstein 2006; Bortolotti and Cox 2009; Coltheart 2017; Bortolotti 2018). For instance, Lisa Bortolotti and Rochelle Cox (2009), distinguish “narrow” approaches to confabulation, which make reference to cognitive pathology that accompanies confabulation in clinical contexts, from “broad” approaches which draw attention to the epistemic and functional features common in both the clinical and non-clinical cases just discussed, namely that they are sincerely given accounts which are not grounded in evidence.³ Recently, Coltheart (2017) has provided a compelling defence of the notion that there is a broader phenomenon of confabulation in play in both clinical and non-clinical contexts, which we’ll explore in more depth in the next section.

3 Confabulation as (Attempted) Causal Explanation

In the spirit of the “broad” understanding of confabulation, Coltheart (2017) has recently offered a new theory that unifies clinical and non-clinical instances, seeing them emanate from the same essential feature of cognition. Conversational context requires the individual to explain why they have behaved in the way they just did. In these cases, either an experimental confederate or a clinician asks the individual for an explanation. According to Coltheart, these individuals are unable to offer true explanations, because in each case, there is no answer to the question (2017, p. 66). One might contend that this isn’t strictly true, for individuals do not act at random. For instance, Nisbett and Wilson (1977) suggest that participants choose because of a position effect, which is surely one possible answer to the question of why they acted. But interlocutors usually want to hear a particular sort of answer, an answer in which we explain that the choice was brought about by our recognition that the chosen option was superior to others, thus justifying the choice. So, I think we can read Coltheart as meaning that there is no answer which both justifies the choice and explains how the choice came about. Nonetheless, explanations are given sincerely (2017, p. 66).⁴

Drawing on work by Gopnik (2000), Coltheart maintains that there is a general property of human cognition, which he calls “the drive for causal understanding”, which leads us to readily come up with explanations of why things happen, even when our explanations are not grounded in evidence and fail to identify the relevant causes. Recall that Gopnik argues that humans have a special representation system, the TFS. In her exposition of the TFS, Gopnik makes an analogy with the visual representation system, which transforms retinal input into representations of objects moving through space. The TFS operates one level up, drawing on information from all perceptual systems, transforming it into causal maps, understood as “abstract, coherent, defeasible

² For other examples of confabulation in the non-clinical population, Brasil-Neto et al. (1992), Uhlmann & Cohen (2007), Hall et al. (2012), and Johansson et al. (2005).

³ Falsity, rather than an inappropriate relation to evidence, is often identified as the key epistemic characteristic of confabulation (e.g. Örvulv and Hydén 2006). But we ought to make room for the possibility of confabulation which does not have the appropriate relation to evidence, yet happens to be true (see footnote 6 for an example). I follow others (Bortolotti and Cox 2009; Hirstein 2006; Bortolotti 2018) in identifying confabulation as not grounded in evidence. One might think that there is a sense in which almost all explanations are grounded in evidence, in so far as they are given on the basis of an inference from experience. However, I also follow others (ibid) in speaking of an explanation of a choice that is “grounded in evidence” to mean an explanation that identifies the processes relevant to why that choice was made.

⁴ Coltheart’s focus here is confabulation that is provoked by a question directed at the confabulator, where the confabulation functions to provide an answer. By contrast, Martha’s confabulations were not provoked by questioning, but occur spontaneously. Originally proposed by Berlyne (1972) the provoked/spontaneous distinction is clinically significant, with provoked and spontaneous confabulations varying across diagnoses (e.g. Nahum et al. 2012). But both display the fundamental features of the “broad” understanding of confabulation (given sincerely; but not grounded in evidence). Further, Martha’s confabulations are not spontaneous in the ordinary understanding of the word, because they serve as answers to situations which Martha perceives to be perplexing, and which she desires to resolve. So, I will consider both standardly provoked confabulation, and confabulation such as Martha’s which is spontaneously produced, but not random, because it occurs when the confabulator perceives some feature of their environment in need of explanation.

representations of the causal structure of the world” (Gopnik 2000, p. 300). For Gopnik, a species that can model causal relations between entities has an evolutionary advantage over one that models only spatial information, the former being more successful at navigating, predicting and exploiting their environments, ultimately facilitating the achievement of goals related to survival, and increasing the likelihood they will successfully reproduce (2000, p. 309).

Crucial to the proposal is the distinctive *phenomenology* of explanation-giving: the satisfaction we feel when we experience what Gopnik calls the “aha!” moment of explanation, when we take ourselves to have identified the relevant causal relations and perplexity is diffused (2000, p. 310): The satisfaction we experience when we produce explanations motivates us to keep coming up with them (analogously to the way that orgasm motivates creatures who experience it to have sex and reproduce), so that the TFS is in frequent use, and cognition is well stocked with causal maps of our environment.

What about bad explanations, those that do not have the appropriate relationship to the evidence, which result in non-veridical causal maps? For Gopnik, whilst the TFS aims at producing veridical causal maps, it is not guaranteed to do this, and might often misidentify causes. Analogously, the visual system often produces non-veridical representations (e.g. illusions, hallucinations), but this does not undermine that the selected-for function of the visual system is to produce veridical representations. Likewise, the fact that the TFS often produces what Gopnik calls “normatively deficient” explanations, resulting in non-veridical causal maps, does not undermine that its selected-for function is to produce well-grounded explanations, which get the causal structure of the world broadly right (Gopnik 2000, p. 314).⁵ Bad explanations merely demonstrate that the TFS does not function as it should on some occasions.

According to Coltheart, when people confabulate (in both clinical and non-clinical contexts), the TFS is operating, and they are motivated to achieve that same feeling of the satisfaction of explanation which drives all explanation-giving behaviour. Coltheart maintains that clinical confabulations are “cognitively normal responses to cognitively abnormal situations” (2017, p. 68). Thus, even in when they occur in clinical cases alongside cognitive deficits, confabulations fall within the spectrum of normal human behaviour.

This is unlike previous accounts which stress the narrative features of confabulation, which we’ll now explore in more depth before examining whether they are consistent with one another.

⁵ Gopnik’s account tends to run together (epistemic) norm talk and biological function talk. For an explicit account of the compatibility of these notions, see Sullivan-Bissett (2017).

4 Confabulation as Narrative

Örülv and Hydén criticise past work on confabulation for concentrating on pathology and cognitive deficit (2006, p. 649). Focusing on cases in dementia, they identify what they call some “productive” (beneficial) aspects of confabulation which are not yet identified or incorporated into Coltheart’s proposal. Specifically, they argue that confabulation scholars should recruit “a contextualized description of confabulation in order to understand it as a social and discursive phenomenon with productive features and certain functions rather than merely as a physiological and cognitive phenomenon” (ibid). Their proposal is that confabulations are discursive events with distinctive narrative features, the aim of which is not simply to provide a plausible causal explanation, but to impose meaning on otherwise perplexing experiences. Of course, almost all communicative acts have meaning of some sort, but Örülv and Hydén seem to have something more specific in mind: personally and socially significant meanings which link to the identity and life-history of the confabulator.

We can revisit their example of Martha to see this in context. Örülv and Hydén reach their understanding of Martha’s situation as described in LISCENCE and COFFEE from Sect. 2 by drawing on further clues from Martha’s conversations with other residents. Martha makes numerous attempts to offer many other residents (and their families) coffee besides Catherine; she is confused as to why there is a card addressed to another woman on the table; she is baffled when the nurses do not ask for her permission before using the lavatory. Örülv and Hydén also identify several reminiscent conversations in which Martha tells other residents how she had taken pride in welcoming guests (particularly those in need) into her home, sharing food and showing hospitality (2006, pp. 655–656). They conclude that in the events captured in LISCENCE and COFFEE, Martha has come to believe once again that she is the gracious hostess, as she genuinely had been on many occasions throughout her life, a role that was clearly important to her (2006, p. 655). As such, her confabulation in COFFEE, concluding that everyone has already been served, rather than acknowledging that she cannot work the unfamiliar coffee pot, allows her to preserve an interpretation of the situation that has important social meaning for her.⁶

⁶ Martha’s claim that everyone has already coffee is in fact true; prior to the COFFEE episode, the nurses did indeed serve everyone. But attending to the precise unfolding of events, Örülv and Hydén suggest that Martha does not arrive at this claim through appreciation of this evidence, because she makes the claim directly after failing to operate the unfamiliar coffee pot, and this, in the context of other episodes in which Martha acts out the hostess roll, leads to the interpretation that Martha’s claim is a confabulation. Note that an account which centres falsity as the key epistemic feature of confabulation cannot characterise Martha’s claim as confabulatory, and so this is a concrete exam-

Likewise, the confabulation in LISCENCE reflects another socially important theme for Martha. Örvulv and Hydén document a number of conversations in which Martha talks of past events in which she was self-supporting financially, paying for things herself, whilst also saving large amounts to provide for her family, as people in both her personal and professional life took advantage of her generosity (2006, pp. 662–665). Establishing that she would never watch a TV without paying for the license, therefore, is an opportunity for Martha to call to mind and affirm these important aspects of her social identity.

Others have highlighted these productive features of clinical confabulation. Conway and Tacchi (1996) contend that the content of confabulated memories often concerns issues of personal significance to the patient, particularly regarding features of their identity and relationships to others. They discuss the case of “OP”, a brain injury patient with post-traumatic amnesia who confabulated elaborate stories about family holidays which could not have happened, but which served to provide support throughout the disorientating experience of care and treatment after the accident, during a time when her family in fact reduced their contact with her. OP’s confabulations enable her to recast negative events in a better light, emphasising social meanings that define her sense of self, enabling her to “rewrite her personal history so that both the remote and recent past provide support for her in a difficult period when little external support is present,” (Conway and Tacchi 1996, p. 333).

Similar considerations lead Örvulv and Hydén to conclude that:

Confabulatory utterances such as in [LISCENCE] should be approached as narratives told to an audience in order to make a certain point. They should be seen as having a logic of their own in accordance with a plot and an overall storyline, rather than as statements about the world that are to be perceived as either true or false. The confabulating person’s way of understanding the situation is based on a narrative logic rather than on a logic of true and false. It does not concern the world as it is, but the world as it makes sense—and the person acts accordingly (2006, pp. 669–670).

There is some unclarity in the articulation above that because confabulations are narratives, they are not to be perceived as either true or false. Narratives are, minimally, (sets of) propositions about the world, and so they are at least

Footnote 6 (continued)

ple of why accounts which stress the importance of *how the claim is arrived at* (e.g., that it is appropriately grounded in evidence of what drives the relevant choice or action) should be preferred.

truth-evaluable. So, Örvulv and Hydén are more charitably interpreted as meaning that confabulations should not be understood simply as attempts to generate theories about the world, but instead to tell a story that imposes sense on an otherwise perplexing experience, achieved by emphasising ideas that have some special importance to the confabulator.

We have here a different perspective on confabulation to that provided by Coltheart (2017), and it has not yet been established whether confabulations which have this specific function, to create meaning, can still be considered the output of the TFS. The TFS, recall, aims to produce veridical causal theories, an aim that is supposedly explained by natural selection—because a capacity to produce veridical causal theories confers a selective advantage. Creation of meaning is not yet part of the picture. But, for Örvulv and Hydén, meaning creation is a systematic function of clinical confabulation, one that individual confabulations can successfully perform, without being veridical causal maps. Are they produced by a system distinct from the TFS after all, one that functions specifically to create meaning? I turn to this question in the next section.

5 The Pursuit of Resonant Meaning

In the following, I suggest that sometimes—in fact, often—the TFS produces explanations that appear to be attempts to supply causal theories, but which have a specific character not yet articulated on Gopnik’s account. That is, in presenting what is ostensibly a causal theory, they also aim to emphasise particular meanings, which often have very little to do with the relevant causal relations. This is not restricted to clinical confabulation, and nor is it an accidental quirk of the TFS. Whilst causal maps can serve their function in private, explanation-giving itself is communicative, and so it is an inherently social practice. Explanation-giving is therefore not only shaped by the advantages it confers in the natural world (e.g. successfully modelling causal relations); but also by the advantages it confers in the social world.

Örvulv and Hydén’s narratives are a subset of larger set of narratives with a particular property that I shall call *resonance*, which we’ll now explore in detail.

5.1 Resonance

A narrative is resonant when it “strikes a chord” or “rings true” for the teller and audience. This is close to the idea of “narrative fidelity” proposed by Walter Fisher (1984), in which we value stories with themes that we recognise from our own lives, which capture some of the essence of the human condition. Fisher draws on the epic of Gilgamesh as an example of a story with narrative fidelity, which he claims, even at over 4000 years old, contains many ideas

and themes we recognise as relevant and important to human experience today (1984, p. 281). Fisher acknowledges that fidelity has some epistemological constraints, in so far as a story should not be too far from what we can imagine as corresponding with reality. But fidelity is also given in the “mythic” aspects of a narrative, by which Fisher is plausibly interpreted to mean how the story engages us with stirring but recognisable themes, as well in its promotion of moral norms and behaviour, which Fisher says uplifts us (1984, p. 280).

We will not take the notion of narrative fidelity *tout court*. It’s a common criticism of narrative fidelity that it is unable to account for people who find distinctly *immoral* narratives compelling—*Mein Kampf* has struck a chord with many, for example (Warnick 1987). More generally, not everyone finds the same stories compelling to the same extent. With this in mind, my proposal of how we should understand resonance is to take Fisher’s notion of narrative fidelity, but strip away the normative baggage. Resonance has the same phenomenological component, that feeling when a narrative “strikes a chord”, and elucidates meaning, but we cannot so neatly account for when we’ll feel it by pointing to some universally agreed upon set of moral principles. Instead, I propose an analogue to Quinean holism about verification, on which the resonance of any given narrative for a subject *S* is determined by its various connections to other experiences and narratives that *S* finds resonant. Given that many experiences are common across the human condition, there is still room on this picture for many people to find many of the same ideas similarly resonant.

One might wonder whether resonance is just the same thing as Gopnik’s “aha” moment (2000), but the evocation conditions are different. Gopnik’s argument is that “ahas” happen when we aim to provide causal theories, whereas I am suggesting that resonance arises when we interpret and assign meanings, through recognition of connection to other ideas and experiences, regardless of whether we are also searching for causes. We can find fictional narratives resonant, even though we do not take them to be veridical maps of reality. We can also find narratives which present themselves as descriptions of reality as resonant independently of whether they are grounded in evidence. As such, the experience of resonance does not always promote the discovery of causes. We’ll return to the precise way that this proposal extends Gopnik’s idea shortly.

5.2 Existing Literature and Examples

Let’s explore how the idea that we often imbue our descriptions and explanations of aspects of ourselves and the world with resonance is supported by existing research into narrative. A number of authors in the humanities argue for the centrality of narrative to interpreting past and

current experiences. It has been argued that we consolidate who we are through the stories we tell about ourselves (Ricoeur 1984–1988); and that recalling and communicating our experiences is not so much a process of introspection as it is interpretation, a chance to find meaning and significance in our lives. (Freeman 1993). As well as finding meaning in past events, we use narratives in real-time to interpret unfolding experiences, to connect and situate ourselves in our current environment (Fisher 1985).

The idea that we use meaningful narratives as a means to interpret past and current events and experiences is also maintained through a body of empirical work in the social sciences (Fivush et al. 1995; Reese and Brown 2000; Neisser 1988; Engel 1999). For example, Fivush et al. (1995) show that parents who share embellished, resonant narratives about the past facilitate their children’s ability to recount their own past, and argue that autobiographic memory itself develops in service of our recounting our experiences to others in this way. Cunliffe and Coupland (2011) demonstrate that sports teams use narratives in their on-pitch discourse to characterise the unfolding successes and misfortunes of their game. For example, a “hero to villain to hero” narrative is recruited to redeem a player who scores after a second attempt. Of course, successful goals, tries, etc. are a combination of the player’s skill, and factors beyond their control (such as the trajectory of the ball as it is passed to the player taking the shot, the weather conditions and their effects on the pitch, and so on) but this does not necessarily prevent the team using narratives which focus chiefly on individuals’ characters to explain their fortunes in the game.

A further example comes from Poulson et al. (1979) who ran a study in which young children were asked to describe and recall a series of pictures taken from story books. In one condition the pictures were in their original order, with a clear narrative progressing throughout the sequence, whilst in another condition the pictures were scrambled with no obvious narrative. When recalling the pictures, 6-years-old tried to interpret the pictures as a coherent story: making inferences; attributing understandable thoughts, emotions and purposes to the characters; and using narrative conventions, such as setting the scene, using temporal connectives between events, and resolving events. Notably, this occurred even in the scrambled condition in which no obvious narrative was discernible. The authors say:

When the pictures were in order, they were quite successful [at telling a story]; when the pictures were out of order their attempts to tell a story frequently failed, yet they kept trying, so that the 6-year-olds produced in fact more story propositions in the scrambled condition than in the normal condition. (Poulson et al. 1979, p. 392).

One might wonder whether this study could just as easily be interpreted as further evidence of the standard operation of the TFS. Perhaps some of the descriptive behaviors the 6-years-old exhibited could be construed as attempts to use causal theories as a way of interpreting events in an imaginative or fictional setting, such as explaining behaviour as the result of particular thoughts they attributed to the characters, and using temporal connectives to show one event was the result of one that preceded it. But it's harder to explain other behaviours as the provision of causal theories. For instance, scene setting and resolving storylines seem to have more to do with entertaining and engaging the audience in the thematic elements of the story, than simply uncovering probable causes.

I propose that many everyday explanations involve resonant narratives. Consider explanations of events that defer to there being "greater powers" at work, such as conspiracy theories, or explanations that invoke the operation of deities and other spiritual forces. Psychologists have argued that these sorts of explanations are attempts to impose meaning on experiences, to reinterpret (mis)fortune as the intentional work of characters and forces with particular agendas, rather than as the result of more mundane conditions and circumstantial luck (Hood et al. 2009). One example is using astrological phenomena (e.g. the position of planets relative to points on the earth's surface) to explain ups and downs in one's life. Take Robin, for example. Last week, he lost his debit card, bought milk that turned out to be sour, and failed his driving test. Causal explanations which cite a combination of physical and psychological conditions that satisfactorily explain these events independently of each other are available. Accordingly, the rip in his wallet that he had not got around to repairing had been steadily widening until his card finally fell out; his local shop experienced an unforeseen delay in which stock intended for the refrigerated aisle was left out in the sun; and he had not properly practised the manoeuvres required to pass the driving test, so the examiner made a reasonable call failing him.

However, Robin does not attempt to uncover the causal relations cited in these explanations, perplexed by what he sees to be a string of unlucky events which cannot just be down to chance. In this respect, Robin is much like other humans: typically bad at recognizing random sequences, and overperceiving patterns and significant relations which are not really there (Gilovich et al. 1985). In this case, Robin turns to his horoscope for this week. It suggests that Aries are in for a downturn in their fortunes (the lost card, the sour milk) and that Aries' natural confidence and tenaciousness will cause them to have a run in with level-headed Libras (Libras typically become examiners and judges, by the way). The horoscope explanation is preferable to Robin because it imbues the unlucky, but frankly mundane and unconnected events with a deeper significance, allowing

a narrative to unfold which weaves the events together; emphasising not undesirable traits of his personality as central to the story, as opposed to the more banal combination of an ordinary share of bad luck and a lack of effort and vigilance to explain the events.

Consider another example. For a recent birthday, my mother bought me a woollen jumper for wearing about the house, acknowledging that I get cold easily, which is true. The jumper has a series of furry tassels, which look a little bit like some toys I had recently bought my cats, and I joked that the garment was perfect, because it was like it had built-in cat toys. The day went by, and later on, the conversation returned to the jumper. At this point, my mother explained that she had bought the item for me because it is like it has built in cat toys. All the contextual clues suggested that her explanation was sincere, that she'd forgotten that in fact I had made the original observation that the garment included cat toys, and now, she genuinely took this to be (one of) the reason(s) she'd gifted the jumper to me.

One might wonder why she'd do this, but further information on the context of our relationship is revealing. We don't have as much to talk about and to regularly touch base on as we did in the past, but we're both real animal lovers, and sharing stories about our pets is often a significant part of our conversations when we do catch up. The idea of giving a jumper with built in cat toys to a person who is known for being very fond of their cats captures both of our imaginations, and resonates with us in a way that gifting a jumper for warmth only just doesn't as much.

The explanations put forward by Robin and my mother attempt to identify the causes of misfortune or reasons for giving respectively (even though neither are sufficiently supported by the evidence) and can be thought of as outputs of the TFS. Better explanations exist which succeed in identifying the relevant causes, although Robin does not choose it in the first place, and my mother does not stick to it. Why? Gopnik's account as it is does not have the appropriate apparatus to give an answer. I think we have reason to extend it as follows: when offering explanations, people are not simply motivated to search for causal relations, they are also motivated to supply explanations which emphasise themes which they find resonant, and which facilitate the attainment of social goals, and related psychological functioning, as we'll now see.

5.3 Why Communicate Resonant Meanings?

There are psychological benefits to the interpretation and communication of resonant meanings. Telling meaningful stories about our experiences is one way that we construct our identity and our sense of ourselves (MacIntyre 1984; Schechtman 1996). This can in turn enhance our psychological wellbeing, increase our resilience, and increase our

success in practical achievements. For instance, constructing a personally meaningful story, in which we are the protagonists, improves our opinion of ourselves, and reduces the extent to which we stigmatise ourselves (Corrigan et al. 2013). Further, as Bortolotti (2018) points out, enhancing our sense of self makes us more likely to pursue our goals in the face of set-backs, rendering us more productive and more effective at problem-solving and planning (Alicke and Sedikides 2009; Hepper and Sedikides 2012).

There are also social benefits. Creating and maintaining resonant meanings in our conversations about our experiences and the world can actually enhance social interaction. Telling meaningful stories about experiences creates and strengthens social bonds and group cohesion, increasing the sense of community and belonging (Bruner and Feldman 1995; Fivush et al. 1995; Reese and Brown 2000; in; Kleinknecht and Beike 2004). It also develops shared values, which promotes understanding and working with each other, increasing the pursuit of shared goals and projects (Engel 1999; Linde 1993; Reese and Brown 2000).

A number of authors argue that the memory system itself is geared towards fulfilment of these psychological and social goals. Recall Fivush et al. (1995) claim that autobiographical memory develops to assist our recounting of our experiences to others, in order to support socialising. Conway and colleagues have argued that the information we are able to recall about ourselves and our past experiences is determined at least in part by particular social and psychological goals that are currently active (e.g. Conway and Pleydell-Pearce 2000; Conway et al. 2004). More recently, Mahr and Csibra (2018) argue that episodic memory evolved to support our assertions of epistemic authority in our interpretations of past events, as we communicate our take on the past to our peers.

We can now say why this practice of embedding resonant meanings in much of our communication occurs and is sustained, as well as how it makes use of Gopnik's TFS. Explanation giving is not just an opportunity to search for causal relations, but a chance to stake our claim on a situation in order to amplify ideas which resonate with us, where bringing those ideas to the fore facilitates the attainment of psychological or social goals.

Resonant explanations can have positive psychological effects. For example, Robin's preferred explanation, in which his experiences aren't the result of an undesirable lack of effort and vigilance, but are part of Aries' confidence and tenaciousness, boosts his sense of purpose and identity. This in turn might contribute to better psychological functioning, and have knock-on positive effects on socialisation, enabling Robin to present himself more confidently, believing that there is a deeper meaning to his experiences than simply bad luck and lack of effort, when he recounts this explanation to a friend.

Indeed, explanation is often a communicative practice, a chance to draw the attention of others to themes, the presence of which themselves enhance socialisation, for instance, by presenting us in a positive light, and making it more likely that our peers will want to continue shared projects with us; or by providing an opportunity for social bonding. We can think of the explanation given by my mother as an example of the latter, the gift of a jumper with apparently built in cat toys to a cat person has become a little piece of shared history that we refer back to and laugh about when we catch up. Martha's explanation that everyone has already had coffee is an example of the former, for it is designed to demonstrate to others that, despite being unable to operate the coffee pot, she, as the hostess, is nonetheless in control of the situation, and is taking care of her guests.

Even though the reasons why Robin and my mother fail to provide accurate explanations differ to why Martha fails to,⁷ all are motivated to provide an explanation which emphasises resonant themes, in order to facilitate the attainment of particular social goals, or related psychological functioning. So, imbuing explanation with resonance is not just something that occurs in clinical cases of confabulation, but in non-clinical cases, and explanation giving much more generally. As such, the TFS regularly produces theories which aim at emphasising resonant meanings.⁸

We need to say more about how this fits with Gopnik's evolutionary story about why the TFS functions as it does. Recall that for Gopnik, the TFS produces causal theories because its doing so confers an adaptive advantage (rendering us more successful at navigating and succeeding in our physical environment, thus surviving long enough to successfully reproduce). There is a strong and a weak hypothesis for integrating resonance into the TFS framework. The strong hypothesis holds that in so far as establishing resonant

⁷ For Robin and my mother, the information which would enable an accurate explanation to be given is not accessed due to a lack of epistemic vigilance—it perhaps could be accessed given further reflection, whilst for Martha, this information is thought to be irretrievable due to the progression of her disease.

⁸ One may wonder what distinguishes this proposal from views earlier discussed, that the memory system itself functions in response to social and psychological goals by encoding and retrieving memories which facilitate goal attainment (e.g. those of Mahr and Csibra 2018; Conway et al. 2004; Conway and Pleydell-Pearce 2000). The proposal is that TFS operates downstream of the memory system, drawing on the memory system's outputs when the explanandum is a past event, but the TFS also operates when we explain presently unfolding experiences about which there are no memories yet (such as Martha's confabulation about her present experience of being unable to open the coffee; or the participants' present experience of looking in the mirror in Barnier et al. 2010). The proposal is therefore able to model all explanation, whether it is of past or present experience, as a unitary phenomenon. I thank an anonymous referee for pushing me to clarify this point.

meaning enables us to navigate and succeed in our *social* environment, facilitating successful (survival until) reproduction, the production of resonant meanings also confers an adaptive advantage. But I will not defend this here. Even though we are a highly social species, and less complex social aptitudes often confer evolutionary advantages (the ability to read expressions, for instance, as in Shariff and Tracy 2011), providing meaningful explanations would likely be an indirect (and therefore reasonably weak) influence on the development of the TFS itself.

So, I defend the weaker hypothesis, that the tendency to imbue explanations with resonant meanings is not an evolutionarily selected for function of the TFS, but one that is maintained through cultural practices that we are inducted into in early childhood, which “hijack” the TFS, so to speak. Indeed, evidence suggests that the convention of meaningful storytelling is learned dialectically, through social-interaction with caregivers (e.g. Engel 1999, Fivush 1995, Nelson 1996). Further, it has been proposed that children learn that telling a meaningful story is supportive of social interaction before they recognise that others expect them to give accurate representations of events they have experienced (Neisser 1988).

This is compatible with Gopnik’s account, recall that her argument is that the TFS evolved because it gives us veridical information about the causal structure of the world in the long run, but it is consistent with this that the TFS regularly delivers non-veridical causal maps (2000, p. 315). Gopnik’s account as it stands lacks the theoretical apparatus to explain why, so often, resonant theories are preferred over veridical causal ones. This account can, because it highlights that explanation is an opportunity to further our psychological and social ends, and proposes that whilst the TFS still often delivers veridical causal theories, psychological functioning and social practices which promote resonant explanations have hijacked the TFS, leading to a tendency to produce resonant explanations as well as the evolutionarily functional outputs, veridical causal ones.

The circumstances in which the explainer finds themselves, as well as their present internal motivational states, will dictate whether the TFS operates produces a veridical causal explanation or produces a resonant explanation. For instance, a scientist versed in the customs of the discipline, who is highly motivated to provide an accurate model of the world, and whose immediate peers share this same motivation, may be less inclined to imbue her explanations with resonance.⁹ Conversely, someone seeking to build new

social networks and to connect with others at a social gathering might be drawn to explanations of their experiences which provide an opportunity for bonding, circumstances which promote resonance. So, which aim is operative, accuracy or resonance, for a particular explanation, depends on the individual’s circumstances and other active aims.

Let us turn now to the implications for confabulation.

5.4 Implications for Confabulation and Future Research

This account demonstrates that the approaches of (i) Coltheart (2017) on which confabulations are the outputs of a TFS which governs all explanations (thus modelling confabulation as continuous with everyday explanation giving); and (ii) Örvulv and Hydén (2006) who stress that confabulations should be understood as attempts to emphasise socially important meanings, can be consistent with each other—and moreover, that we can preserve the explanatorily powerful elements of both views. On my account, it remains true that confabulations are continuous with everyday explanation giving, and also true that confabulations aim to emphasise socially important meanings, because explanation giving in general is replete with attempts to emphasise resonant meanings, a practice explained by the social and psychological benefits of doing so. This gives us a new interpretation of the examples of non-clinical confabulation about consumer choices from Nisbett and Wilson (1977) and about policy support from Cohen (2003) as discussed in Sect. 2. Participants aren’t just giving an explanation which serves as a plausible causal theory, they’re giving an explanation which identifies personally significant principles which render their actions meaningful in the relevant circumstances—as humans do in a wide variety of contexts when telling others about themselves and the world.

This account paves the way for future research into the notion of resonance. One avenue for future investigation would be to examine whether using resonant narratives enables people to establish meaningful links between otherwise disparate items in experience with the effect of facilitating information preservation. Recall the Poulson et al. (1979) study in which children construct meaningful stories when retelling scrambled picture sequences. The researchers also found that the material which was recalled best was that which was well integrated into the story constructed by the child. Other uses of narratives appear to have a similar information preserving effect—one might, for instance, think of mnemonics as micro-narratives which create meaningful connections between otherwise disparate pieces of information. It is worth investigating whether something similar is going on in each of these cases, and whether there is a general, robust effect of information preservation through the use of resonant narratives.

⁹ We should be careful not to claim that results in science are therefore bias and motivation-free. Instead, the thought is that we build into our best sciences certain practices which motivate accuracy and de-motivate relying on resonance (e.g. use of valid and reliable measures; a peer-review process, etc.).

Such an investigation would be worthwhile in itself to further understand the significance of resonance. But there could also be important consequences for dementia sufferers and dementia care: should it be the case that there are information-preserving effects of resonance, then the epistemic considerations around confabulation and whether confabulated stories should be encouraged become more complex. These are of course all preliminary suggestions, and are worth exploring in full in future work, as I intend to do.

6 Conclusion

We don't have to choose between Coltheart's (2017) account of confabulation as continuous with everyday explanation giving, and Örvulv and Hydén's (2006) account that confabulations function to emphasise personally significant themes. We can view explanation-giving behaviour as both the output of the TFS, and as a means of imbuing our situation with resonant meanings, a tendency that developed as a result of the deeply social context in which explanations are given, and to facilitate related psychological functioning. Being motivated to give resonant explanations means that we are less orientated towards identifying underlying causes, and so we should not lose sight of the epistemic costs which accompany this tendency. However, this tendency confers significant social and psychological benefits, and future work should address the extent to which information preservation is a robust benefit of the tendency to imbue our accounts with resonance.

Acknowledgements The author acknowledges the support of a European Research Council Consolidator Grant, Pragmatic and Epistemic Role of Factually Erroneous Cognitions and Thoughts (PERFECT), Grant Agreement 616358. The author would also like to thank Katherine Puddifoot, Ema Sullivan-Bissett, Valeria Motta, attendees of the University of Birmingham Philosophy Department's Work in Progress seminar, attendees of the PERFECT 2018 Confabulation Workshop and the Confabulation and Epistemic Innocence workshop at Milano-Bicocca University, and two anonymous reviewers from the journal *Topoi*, for discussion of and comments on an earlier version of this article.

Funding This study was funded by a European Research Council Consolidator Grant, Pragmatic and Epistemic Role of Factually Erroneous Cognitions and Thoughts (PERFECT), Grant Agreement 616358.

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 performed by any of the authors.

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

References

- Alicke MD, Sedikides C (2009) Self-enhancement and self-protection: what they are and what they do. *Eur Rev Soc Psychol* 20:1–48
- Barnier A, Cox RE, Connors M et al (2010) A stranger in the looking glass: developing and challenging a hypnotic mirrored-self misidentification delusion. *Int J Clin Exp Hypn* 59:1–26
- Berlyne N (1972) Confabulation. *Br J Psychiatry* 120:31–39
- Bortolotti L (2018) Stranger than fiction: costs and benefits of everyday confabulation. *Rev Philos Psychol* 9(2):227–249
- Bortolotti L, Rochelle EC (2009) 'Faultless' ignorance: strengths and limitations of epistemic definitions of confabulation. *Conscious Cogn* 18(4):952–965
- Brasil-Neto JP, Pascual-Leone A, Valls-Solé J et al (1992) Focal transcranial magnetic stimulation and response bias in a forced-choice task. *J Neurol Neurosurg Psychiatry* 55(10):964–966
- Bruner J, Feldman CF (1995) Group narrative as a cultural context of autobiography. In: Rubin DC (ed) *Remembering our past: studies in autobiographical memory*. Cambridge University Press, Cambridge, pp 291–317
- Cohen GL (2003) Party over policy: the dominating impact of group influence on political beliefs. *J Pers Soc Psychol* 85(5):808–822
- Coltheart M (2017) Confabulation and conversation. *Cortex* 87:62–68
- Conway MA, Pleydell-Pearce CW (2000) The construction of autobiographical memories in the self-memory system. *Psychol Rev* 107(2):261–288
- Conway MA, Tacchi PC (1996) Motivated confabulation. *Neurocase* 2(4):325–339
- Conway MA, Singer JA, Tagini A (2004) Self and autobiographical memory: correspondence and coherence. *Soc Cognit* 22(5):491–529
- Corrigan PW, Kosyluk KA, Rüsç N (2013) Reducing self-stigma by coming out proud. *Am J Public Health* 103(5):794–800
- Cunliffe A, Coupland C (2012) From hero to villain to hero: making experience sensible through embodied narrative sensemaking. *Hum Relat* 65(1):63–88
- Engel S (1999) *The stories children tell: making sense of the narratives of childhood*. W. H. Freeman and Company, New York
- Fisher W (1984) The case of public moral argument. In: Porrovecchio MJ, Condit CM (eds) *Contemporary rhetorical theory: a reader*, 2nd edn. The Guilford Press, New York, pp 265–287
- Fisher WR (1985) The narrative paradigm: an elaboration. *Commun Monogr* 52(4):347–367
- Fivush R (1995) Language, narrative, and autobiography. *Conscious Cognit* 4:100–103
- Fivush R, Haden CA, Reese E (1995) Remembering, recounting, and reminiscing: the development of autobiographical memory in social context. In: Rubin DC (ed) *Remembering our past: studies in autobiographical memory*. Cambridge University Press, Cambridge, pp 341–359
- Freeman MP (1993) *Rewriting the self: history, memory, narrative*. Routledge, London
- Gilovich T, Vallone R, Tversky A (1985) The hot hand in basketball: on the misperception of random sequences. *Cognit Psychol* 17(3):295–314

- Gopnik A (2000) Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory formation system. In: Keil FC, Wilson RA (eds) *Explanation and cognition*. MIT Press, Cambridge, pp 229–323
- 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(9):e45457
- Hepper EG, Sedikides C (2012) Self-enhancing feedback. In: Sutton R, Hornsey M, Douglas K (eds) *Feedback: the communication of praise, criticism, and advice*. Peter Lang, London, pp 43–56
- Hirstein W (2006) *Brain fiction: self-deception and the riddle of confabulation*. MIT Press, Cambridge
- Hood RW, Hill PC, Spilka B (2009) *The psychology of religion: an empirical approach*. Guilford Press, New York
- Johansson P, Hall L, Sikström S, Olsson A (2005) Failure to detect mismatches between intention and outcome in a simple decision task. *Science* 310(5745):116–119
- Kleinknecht E, Beike DR (2004) How knowing and doing inform an autobiography: relations among preschoolers' theory of mind, narrative, and event memory skills. *Appl Cognit Psychol* 18(6):745–764
- Linde C (1993) *Life stories: the creation of coherence*. Oxford University Press, New York
- MacIntyre AC (1984) *After virtue: a study in moral theory*. Indiana University of Notre Dame Press, Notre Dame
- Mahr JB, Csibra G (2018) Why do we remember? The communicative function of episodic memory. *Behav Brain Sci* <https://doi.org/10.1017/S0140525X17000012>
- Moscovitch M (1989) Confabulation and the frontal systems: strategic versus associative retrieval in neuropsychological theories of memory. In: Roediger HL, Craik FI (eds) *Varieties of memory and consciousness: essays in honour of endel tulving*. Lawrence Erlbaum Associates, Hillsdale, pp 133–160
- Nahum L, Bouzerda-Wahlen A, Guggisberg A et al (2012) Forms of confabulation: dissociations and associations. *Neuropsychologia* 50:2524e2534
- Neisser U (1988) Time present and time past. In: Gruneberg MM, Morris PE, Sykes RN (eds) *Practical aspects of memory: current research and issues*, vol. 2. Wiley, New York, pp 545–560
- Nelson K (1996) Memory development from 4 to 7 years. In: Sameroff AJ, Haith MM (eds) *The 5 to 7 year shift: the age of reason and responsibility*. University of Chicago Press, Chicago, pp 141–160
- Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. *Psychol Rev* 84(3):231–259
- Örülv L, Hydén L-C (2006) Confabulation: sense-making, self-making and world-making in dementia. *Discourse Stud* 8(5):647–673
- Poulson D, Kintsch E, Kintsch W, Premack D (1979) Children's comprehension and memory for stories. *J Exp Child Psychol* 28:379–403
- Reese E, Brown N (2000) Reminiscing and recounting in the preschool years. *Appl Cognit Psychol* 14:1–17
- Ricœur P (1984–1988) *Time and narrative*. University of Chicago Press, Chicago
- Schechtman M (1996) *The constitution of selves*. Cornell University Press, Ithaca
- Shariff AF, Tracy JL (2011) What are emotion expressions for? *Curr Dir Psychol Sci* 20(6):395–399
- Sullivan-Bissett E (2017) Biological function and epistemic normativity. *Philos Explor* 20(1):94–110
- Uhlmann EL, Cohen GL (2007) 'I think it, therefore it's true': effects of self-perceived objectivity on hiring discrimination. *Organ Behav Hum Decis Process* 104(2):207–223
- Warnick B (1987) The narrative paradigm: another story. *Q J Speech* 73:172–182