

The Confabulation of Self

Joanna J. Bryson

40

The Confabulation of Self

Joanna J. Bryson

Confabulation is a technical term for a process typically ascribed to patients who have problems with their memory or their self awareness. We ask a patient why they have done something, and they tell us a narrative that sounds like a memory, but that we know to be false. So we say that the patient has confabulated. Their unconscious (but still diseased) mind has drawn together disparate stories in a desperate attempt to make their recent actions—and lives—make sense.

Such an example is described by Ramachandran and Blakeslee (1998). A patient denies that they have lost control of an arm after a stroke. To test whether the patient really believes they still have the use of two arms, they are handed a tray. The patient, who does indeed still believe both arms work, grasps only one side of the tray, allowing the contents to spill. When asked why the contents have spilled, the patient confabulates 'I didn't want them'. Oddly though, despite the lack of conscious access, the patient is remembering the fact of their paralysis. When their inner ear (part of the vestibular system) is stimulated with a shock of cold water, they report not only awareness of their paralysis but also of how long they have been afflicted. When their vestibular system recovers, they can recall the event of a cold ear, but not what they said concerning their paralysis.

Confabulation is both convincing and fluid, and it happens not only to the ill or injured. Experimental subjects falsely accused of breaking a computer by pressing a wrong key can come to believe the accusation, and even sign a confession (Kassin and Kiechel 1996). This internalisation of the projected guilt only occurs if the accused had been working rapidly (lowering their own certainty of events) and another witness is present that swears to have observed the event. In such a case, a subject can come to believe and even recall what seems to be the most likely explanation. Similarly, subjects fail to notice if a person they are giving directions to changes gender or race while some 'movers' (further stooges of the experiment) briefly come between the subject and the person to whom they are talking (Simons and Levin 1998; Levin et al. 2000). How likely is it for someone to change race or gender? The memory corrects itself without troubling the subject's conscious recall. These are also forms of confabulation: false recall of events known not to have occurred.

One explanation for such lucid and transparent constructions is simple—all narrative memory is essentially confabulation. Recalling narratives is not something at which we are all inherently skilled. This is why we value story tellers and those with good recollection, and why we use external devices like books to record our narratives for us. This is strikingly different from how well and easily we recall whether we have seen a picture before, or how fluently we speak. It is more like our capacity to catch and throw—something we practice for years, something for which you can find adults who still perform weakly, but ordinarily they do better than 12-year-olds, who are in turn better than the average six-year-old.

What most cognitive species are good at is, first, remembering contexts, and second, learning to associate them with ideal actions (Heyes 2012). This allows us to learn what to do when, to respond to opportunities, to behave quickly and intelligently. But for humans, this is not enough. We wish to make complicated, multi-step plans, learning not only actions but consequences. These consequences will create a new context, allowing us to produce a chain of actions and contexts to string complex and clever sequences into a grand design. We can use this same ability to deduce what must have been in the past. We seem easily to recall particularly unusual events as unique contexts, but what led to them? And what ensued? Did anyone ultimately benefit, and if so how did they act?

And why should we care? That question at least is easy. The past matters to the extent that it informs the present and predicts the future. If something bad happened, we want to know why so we can prevent it happening again. If something good happened, we want to be able to repeat it. If something extraordinary happened, we want to be ready if it happens again, if for no other reason than because humans offer prestige to those who seem knowledgeable.

MPS 11/9/2015 08:22

Formatted: Rrh

MPS 18/8/2015 21:53

Deleted:

MPS 18/8/2015 21:53

Deleted:

MPS 18/8/2015 21:53

Deleted:

MPS 18/8/2015 21:53

Deleted:

MPS 19/8/2015 22:39

Deleted: &

Victoria Hume 8/9/2015 08:03

Deleted:

MPS 18/8/2015 21:53

Deleted:

MPS 18/8/2015 21:53

Deleted:

MPS 18/8/2015 21:53

Deleted:

MPS 18/8/2015 21:53

Deleted:

Victoria Hume 8/9/2015 08:05

Deleted: a

Victoria Hume 8/9/2015 08:07

Deleted:

One problem of memory seldom acknowledged is that it takes time (Huk and Shadlen 2005; Sipser 2012). It takes time to think, it takes time to recall. It certainly takes time to consider different options and construct (or favour) a chain of events out of web of possibilities (Chapman 1987). And so, we forget. We only remember what is likely to be useful. But what's interesting is this—what's likely to be useful varies by context. And what we remember varies as well.

People are often shocked or frustrated by their own poor memories. One of the reasons for this is that our memories have not only a directly functional role, but also a more complicated meta-narrative role. Our memories are part of our identity. And while we might recognise that others behave inconsistently or seem to have contradictory goals or motives, we seldom see this in ourselves. One reason for this is because we in fact have several selves (a theory called constructivism; for a recent discussion see Lebow 2012). In particular contexts we have particular goals and beliefs. For example, when walking we may be annoyed by drivers, and while driving be annoyed by pedestrians. Thinkers from Hume (1739) to Minsky (1983) have speculated that a coherent self is an illusion, constructed for some form of convenience: 'we may observe, that what we call a mind, is nothing but a heap or collection of different perceptions, united together by certain relations, and suppos'd, tho' falsely, to be endow'd with a perfect simplicity and identity'.¹

Perhaps that convenience is planning the future—it's easier if at any one point you assume your goals are consistent.

I cannot compare the soul more properly to any thing than to a republic or commonwealth, in which the several members are united by the reciprocal ties of government and subordination, and give rise to other persons, who propagate the same republic in the incessant changes of its parts. And as the same individual republic may not only change its members, but also its laws and constitutions; in like manner the same person may vary his character and disposition, as well as his impressions and ideas, without losing his identity. Whatever changes he endures, his several parts are still connected by the relation of causation. And in this view our identity with regard to the passions serves to corroborate that with regard to the imagination, by the making our distant perceptions influence each other, and by giving us a present concern for our past or future pains or pleasures.²

Or perhaps the narrative of self is something that has evolved culturally because it promotes society—it is easier to maintain social order if people are responsible for their past actions, which after all they are likely to repeat in similar circumstances (Graziano 2014).

The self cannot be considered entirely an illusion. An individual has (ordinarily) one set of human DNA, one set of biological parents, one educational and one criminal history. An individual is more or less fixed in their adult height and a number of their acuties. Nevertheless, much of the self—the stories we tell, the people we love, our homes and other possessions, our friends, our goals, our needs, even our names—much of this changes over time. There is great convenience not just for others but for the self in being able to confabulate a version of the self to benefit the reasoning of that self (and others) about itself.

Now as we come to understand our selves better, we might hope to become better at reasoning about our other selves as well—the pasts we can't remember, the people we appear to be to others, the people we might become in the wrong or right circumstances. We might hope that this insight will improve our empathy and our judgement. But what if our ignorance is not the result of unfortunate chance biological limits, but rather reflects a set of optima, carefully honed by evolution, suited to the dynamics of our social situations and our mental well being? Our new reality of big data providing records of our every move, and of ever-improving scientific (or artificial intelligence) models for predicting and understanding our behaviour, make many fearful because of the loss of privacy from others (Bryson 2015). We fear exploitation by advertisers who might exploit our weaknesses, governments that might find ways to disenfranchise us, bullies and assailants that might manipulate or abuse us. But what about our privacy from reality, and from our selves? Are there limits to the extent to which a life should be self examined? Science tells us that optimism rather than realism is key to our mental well being (Taylor and Brown 1988), but also that our individual levels of optimism and pessimism are highly influenced by both our genes and our upbringing (Plomin et al. 1992). We will almost certainly experience the answer to this question in the near future, but whether we will accurately recall our lives in the past and know what we have gained or lost—that is less certain. More likely, our newly-confabulated selves will conform to the new norms of society and recall a skewed perspective on our current present, our future past.

Notes

- MPS 20/8/2015 12:32
Deleted: ; Huk and Shadlen 2005
- MPS 20/8/2015 12:33
Comment [1]: AQ: We find that a few cross-references have not been listed in the references. Please provide bibliographic details. We have highlighted these occurrences for your reference.
- MPS 19/8/2015 22:43
Formatted: Highlight
- MPS 18/8/2015 21:53
Deleted: ... [1]
- MPS 19/8/2015 22:44
Formatted: Highlight
- Joanna hates Microsoft 14/9/2015 10:19
Deleted: 5
- MPS 19/8/2015 22:44
Formatted: Highlight
- Victoria Hume 8/9/2015 08:07
Formatted: ... [2]
- MPS 18/8/2015 21:52
Deleted: ... [3]
- Joanna hates Microsoft 14/9/2015 10:20
Formatted: Indent: First line: 0 cm
- MPS 18/8/2015 21:53
Deleted: ... [4]
- MPS 19/8/2015 22:45
Comment [2]: AQ: Please confirm wheth... [5]
- MPS 18/8/2015 21:53
Deleted: ...the stories we tell, the peopl... [6]
- Joanna hates Microsoft 14/9/2015 10:22
Deleted: for
- MPS 18/8/2015 21:53
Deleted: ... [7]
- Victoria Hume 8/9/2015 08:06
Deleted: ...We might hope that this insigh... [8]
- Victoria Hume 14/9/2015 10:24
Comment [3]: AQ: could this be rephrase... [9]
- MPS 19/8/2015 22:45
Formatted: Not Superscript/ Subscript
- Victoria Hume 8/9/2015 08:07
Deleted:
- Victoria Hume 8/9/2015 08:07
Deleted: ...But what about our privacy f... [10]
- MPS 19/8/2015 22:46
Formatted: ... [11]
- Joanna hates Microsoft 14/9/2015 10:27
Deleted: !
- Victoria Hume 8/9/2015 08:07
Deleted:
- MPS 18/8/2015 21:53
Deleted: ... [12]
- Victoria Hume 8/9/2015 08:07
Deleted:
- MPS 11/9/2015 08:22
Formatted: H1, Line spacing: single

¹ David Hume, *A Treatise of Human Nature* (London: John Noon, 1739), 207.

² Hume, *Human Nature*, 261.

Victoria Hume 8/9/2015 08:10
Formatted: Font:Not Italic, English (UK)

[Ah, the citations are missing because I had to do this in word, not latex, so nothing checked for me. How old school! Note that Plomin was misspelled above so maybe you do already have that one?](#)

[Chapman, David. "Planning for conjunctive goals." *Artificial Intelligence* 32.3 \(1987\): 333-377.](#)

[Minsky, Marvin. *Society of Mind*. Simon and Schuster, 1988.](#)

[Plomin, R., Scheier, M. F., Bergeman, C., Pedersen, N., Nesselroade, J., and McClearn, G. \(1992\). Optimism, pessimism and mental health: A twin/adoption analysis. *Personality and Individual Differences*, 13\(8\):921–930.](#)