

good as another, they will replace each other randomly, and the imagining will have the quality of a daydream. Conversely, if there is a single, relatively rare occasion that stands out from the others, it will make the corresponding imagination robust. The experience of such *singularity* may be much like that of having solved a puzzle or detected a fact of nature. The occasion in question will stand out from the common ruck of imaginings just as a fact stands out from a fantasy.

Where information about the natural world is absent or ambiguous, singularity may be the best clue about how it functions – parsimony is a decent starting place for theories. But a belief that distinctly delivers good news and bad news will be productive of reward in its own right, regardless of its eventual accuracy. The emotional effectiveness of singular occasions may be experienced as a kind of factuality, more or less confounded with the factuality that comes from physical observation. In the most conspicuous cases, remembered events are experienced again on their anniversaries, especially when the anniversary is a round number; original works of art are felt to be more “real” than exact copies; and placebos (as in sect. 13) are effective in proportion to the expensiveness of the ingredients or the prestige of the healer. Even realistic beliefs get additional value by serving as occasions for emotional reward, as in the “drug effect” of money (Lea & Webley 2006). Conversely, faced with unwelcome urges such as hypochondria, phobic anxiety, or a sense of being dirty, a person searches for a favorable interpretation of the situation – whether she can feel well, or safe, or clean. This interpretation cannot be arbitrary; wishes have little impact. She must choose her belief on the basis of “facts” that she discerns in events beyond her control – a pill given by a doctor, a lucky charm or safety signal, or a “scientific” disinfectant. The belief may even become stabilized as a personal rule: in effect, “I will not give in to panic or disgust when this signal is present.” The same role of singularity can be seen in many other misbeliefs. For instance, delusions (sect. 9) tend to be based on a logical deduction or a remarkable coincidence, and religious faiths (sect. 11) depend on the singularity that comes from having had long histories of consensual agreement – hence their fear of heresies. It would be fruitless to try to decide whether such hedonically based beliefs are more or less adaptive than veridicality; evolution veered away from veridicality with the apes.

NOTE

1. The author of this commentary is employed by a government agency, and as such this commentary is considered a work of the U. S. government and not subject to copyright within the United States.

**False beliefs and naive beliefs: They can be good for you**

doi:10.1017/S0140525X09991178

Marco Bertamini<sup>a</sup> and Roberto Casati<sup>b</sup>

<sup>a</sup>School of Psychology, University of Liverpool, Liverpool L69 7ZA, United Kingdom; <sup>b</sup>CNRS Institut Nicod, Ecole Normale Supérieure, 75005 Paris, France.

m.bertamini@liv.ac.uk <http://www.liv.ac.uk/vp/>  
casati@ehess.fr <http://www.institutnicod.org>

**Abstract:** Naive physics beliefs can be systematically mistaken. They provide a useful test-bed because they are common, and also because their existence must rely on some adaptive advantage, within a given context. In the second part of the commentary we also ask questions about when a whole family of misbeliefs should be considered together as a single phenomenon.

If humans are biologically engineered to appraise the world accurately, how can we explain misbeliefs? After asking this question, McKay & Dennett (M&D) analyse various misbeliefs. Those

resulting from a breakdown in the system, and those that are by-products, do not threaten the claim of adaptiveness of the belief system. Positive illusions are the only bona fide example of misbeliefs. We shall integrate this account by first making a case for the adaptiveness of some mistakes in the conception of the physical world, and by discussing the possibility of a general egocentric bias in generating positive illusions.

The grand aim of Naive physics (NP) is to fully describe common beliefs about the physical world. Naive physics can be traced back to Gestalt psychologists such as Köhler, and to the seminal work by Lipmann and Bogen (1923). The term is also used in artificial intelligence and robotics (Hayes 1978). Despite its grand aim, interest in NP has focused on the discovery that people make some systematic mistakes about everyday phenomena. Examples include judgements about the pendulum motion (Bozzi 1958); predictions of motion of an object in terms of direction, path of motion, and acceleration (Hecht & Bertamini 2000; McCloskey et al. 1980); and predictions about what is visible in a mirror (Bertamini & Parks 2005). In the case of the pendulum, people consider as “natural” a movement that is actually artificially contrived. We can be sure that some mistakes are not cultural whims because they match scientific theories of the past (i.e., Aristotelian mechanics). NP beliefs are not necessarily approximations or simplified representations of the physical world (Cavanagh 2005). In some cases the implied physics is complex, for instance, when subjects deem as correct cast shadows that require light to bend around corners or to be projected from physically impossible locations (Casati 2008).

Even if these mistakes are the manifestation of (implicit) mental models (McCloskey 1983), where do these models come from? Typically NP beliefs are resilient and non-revisable, thus pointing to some modular underlying mechanism. Some NP beliefs are grounded on evidence provided by the visual system. The belief that a pendulum looks unnatural when it moves, for example, originates from how people perceive motion (Bozzi 1958; Pittenger 1989). Aspects of how people reason are also important, as exemplified by the reliance on prototypes of actions (Yates et al. 1988) and heuristics (Proffitt 1999). Mistaken beliefs that originate from properties of perceptual or reasoning mechanisms could be classified as evolutionary by-products. On the other hand, one can ask the question of why these as opposed to other by-products occur. System limitations should also be considered from an evolutionary standpoint. For example, if waitresses make larger mistakes than housewives in the water-level task (the orientation of water in a tilted glass) this may be because the glass as a frame of reference is more important to them in their job than it is to other people (Hecht & Proffitt 1995). This may seem paradoxical but it suggests that attention to a local frame of reference, which is crucial for a task, makes it harder to learn about more abstract frames of reference. Context is, therefore, critical here. At least some NP beliefs, we surmise, are examples of systematically mistaken adaptive beliefs. In spite of their wrongness they provide contextually useful representations.

We are not claiming that each specific NP belief is an adaptation. Our perceptual system and our thoughts may lead us to them as a response to a situation. This brings us to the second point of our commentary.

Adaptiveness itself is hard to assess. Veridicality is not sufficient as a criterion. Just like percepts, most beliefs are *prima facie* veridical (they do not interfere with our interactions with the world) but compliance with logic or the laws of physics is not what they (beliefs as well as percepts) have evolved towards. An adapted organism is one that has accumulated characteristics that maximise fitness, not knowledge per se. Positive illusions are adaptive because they lead people to engage in adaptive behaviours. Whatever the mechanism, positive views of one’s medical condition and of one’s ability to influence it lead to increased health. Quite possibly the effects are not directly in terms of guiding deliberation and choice, rather they are ancillary

effects, such as triggering emotional adjustments and immune reactions. The evidence about biased responses concerning the self is vast, and controversial. It spans items as diverse as: self-serving biases and positive illusions (Taylor & Brown 1994b), implicit egotism (Pelham et al. 2005), narcissism (Nuttin 1985), self-enhancement (Sedikides & Gregg 2008), and self-resemblance and trust (DeBruine 2002), among others.

But are these beliefs specific adaptations or are they facets of a powerful but unspecific underlying mechanism, which we may call “looking after number one”? We think the jury is still out. If specific beliefs originate from specific adaptations, then it should be possible to find not only examples of “positive” illusions about oneself, but also of “negative” illusions about oneself that are, under different circumstances, adaptive. We would, therefore, need an example of a trait that is both generally perceived as positive (e.g., height) and yet such that people tend to see themselves as lacking because the resulting underestimation has a specific adaptive effect. If, on the contrary, we only have examples of overestimations (i.e., errors in the direction perceived as positive) then the most economical hypothesis is that they are all related, and originate from the same generic bias in favour of the self. Another problem with the idea that specific beliefs are specific adaptations is the fact that biases in favour of the self exist also for neutral or non-beneficial aspects. For instance, preferences are influenced by presence in their formulation of the first letter of the name of the person expressing the preference (Nuttin 1985); compliance with a request increases when someone is told that they share a birthday with the requester (Burger et al. 2004); and people overestimate the size of their own head (more than other people’s heads) (Bianchi et al. 2008). It is unclear what the benefits are for these effects, and it seems more likely that they all originate from a generic (and adaptive) egocentric bias.

## Extending the range of adaptive misbelief: Memory “distortions” as functional features

doi:10.1017/S0140525X09991397

Pascal Boyer

Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130.

pboyer@artsci.wustl.edu

<http://artsci.wustl.edu/~pboyer>

**Abstract:** A large amount of research in cognitive psychology is focused on memory distortions, understood as deviations from various (largely implicit) standards. Many alleged distortions actually suggest a highly functional system that balances the cost of acquiring new information with the benefit of relevant, contextually appropriate decision-making. In this sense many memories may be examples of functionally adaptive misbelief.

Memory illusions or distortions are a major area of recent research (Brainerd & Reyna 2005; Roediger 1996; Schacter & Coyle 1995). They are very diverse, ranging from intrusions in word-list recall to therapy-influenced imaginings of previous lives or systematic abuse.

Dramatic memory distortions seem to influence belief-fixation. For instance, in the illusory truth effect, statements read several times are more likely rated as true than statements read only once. People who repeatedly imagine performing a particular action may end up believing they actually performed it (imagination inflation). Misinformation paradigms show that most people are vulnerable to memory revision when plausible information is implied by experimenters. In social contagion protocols, people tend to believe they actually saw what is in fact suggested by the confederate with whom they watched a video.

Another major type of distortion is revision of prior mental states under the influence of newly received information or

changed contexts. People modify their autobiographical memories to fit implicit “theories of change.” They, for instance, think that one gets better at a particular task with practice and therefore revise their memories of past performance to fit the predicted performance curve (Ross & Wilson 2003). In a similar way, in hindsight protocols people revise memories of their own prior guesses (e.g., that London has 10 million inhabitants) after receiving feedback information. Most familiar is attitude-revision, in which subjects routinely mis-remember previously held and subsequently changed attitudes.

These distortions seem to result from the normal standard operation of memory systems. Yet they result in misbelief. Why is that the case?

Distortion is a normative notion, so what is the standard against which memory systems are failing? Surprisingly, this is generally left implicit in memory research. In contrast to, say, decision-making, in which human “biases” are described as deviations from normative models, there are no explicit standards in memory research. That is because an explicit standard for memory performance would require a description of memory functions, and traditionally memory researchers have not been overly preoccupied by functional considerations, with a few exceptions (Anderson & Schooler 2000; Nairne et al. 2008).

As a consequence, memory performance is evaluated against generally tacit, apparently self-evident commonsense assumptions – we can infer those assumptions from the very fact that some memory processes are treated as “distortions.” As mentioned above, it seems that they constitute deviations from a tacit and largely implausible view of memory systems. One assumption seems to be that memory as storage of information is not subject to the same cost-benefit constraints as the rest of cognition, so that information acquired should be stored rather than transformed, *pace* Bartlett (1932). Another assumption is that memory retrieval has its own function, independent from decision-making, so that one should, for instance, expect people to recall attitudes that did not lead to particular decisions.

But both assumptions are biologically odd. It makes obvious sense to consider memory retrieval as a biological function that comes at a cost and is therefore designed to maximize return on that cost (Dukas 1999). Also, it makes evolutionary sense to keep in mind that organisms do not develop cognitive abilities (e.g., retrieval of past experience) for abstract epistemic benefits (knowing what used to be the case). They retrieve information inasmuch as it helps fitness-enhancing decision-making in the present (Suddendorf & Corballis 2007).

Seen in this perspective, many cases of “distortion” appear highly functional. Consider misinformation and other situations in which memories are influenced by confederates’ suggestions. The possibility and need of acquiring vast information from conspecifics also creates the possibility of error and deception. For each item of information, memory and decision-making systems must, implicitly or explicitly, assess the costs and benefits of including information in a belief-box or, alternatively, of keeping track of the information’s “source-tag.” It is certainly plausible that, *in some circumstances*, it is too costly to keep the source-tags for many items of information if they are all used to build a coherent, usable account of one’s own experience. In the same way, repetition effects show that internal judgments of familiarity and fluency play an important role in decision-making. Intuitive epistemics here uses the external world regularity that *in some circumstances* true information is more frequent than false information. What matters for adaptive design is that the circumstances in question be such that this sort of decision-making does not lead to *excessive* vulnerability.

Now turn to attitude revision. In a functional perspective, accurate memory of past attitudes would be an odd proposition for a well-designed memory system. To preserve traces of past, now-irrelevant attitudes without compromising its computations, the system would need to quarantine them from on-line motivation and decision-making (Cosmides & Tooby 2000). The