Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

Optimistic and realistic perspectives on cognitive biases

Pete C Trimmer

Numerous papers have tried to explain cognitive biases, such as optimism and overconfidence, from an evolutionary perspective. The attempts have met with mixed success. I identify why some approaches are more successful than others in explaining sub-optimal behaviours. I conclude that some evolutionary explanations of cognitive biases can be successful; the relevant explanation will depend on the particular bias being studied. In particular, I highlight the need to incorporate internal costs when considering the evolution of mental mechanisms, and how this can provide adaptive explanations of sub-optimal behaviours.

Address
University of California, Department of Environmental Science and Policy, 1 Shields Avenue, Davis, CA 95616, USA

Corresponding author: Trimmer, Pete C (pete.trimmer@gmail.com)

Current Opinion in Behavioral Sciences 2016, **12**:37–43

This review comes from a themed issue on **Behavioral ecology**

Edited by **Andrew Sih** and **Alex Kacelnik**

For a complete overview see the Issue and the Editorial

Available online 21st September 2016

http://dx.doi.org/10.1016/j.cobeha.2016.09.004

Introduction and definitions

Since Kahneman and Tversky pioneered the study of *biases* in cognitive processes (e.g., [1]), the topic has been of great interest to psychologists and, more recently, behavioural ecologists.

Table 1 supplies various definitions of the term 'cognitive bias'. The common theme is of a bias (or distortion) to a cognitive process or mental representation. A cognitive bias could result in optimal behaviour (i.e., behaviour that maximises expected payoff, often measured simply in terms of surviving offspring) according to some definitions but not others. These differences have led to confusion, discussed below. An additional layer of complexity emerges in lab experiments when an individual behaves according to adaptively evolved responses, but could increase rewards if the lab setting was understood by the subject.

In recent studies of non-human species, 'cognitive bias' has been used very generally, to mean decisions that are influenced by emotions [2–4], even if the consequent behaviours are optimal. This can be confusing, as a bias is arguably *from* something, and it is not always clear under this definition what a bias is from (functionally).

Raghubir and Ranjan [5] identify several stages at which apparent cognitive biases could arise, including: perception, memory-retrieval, information integration, making a judgement, and behaviour. Neuroscience is generally not yet able to directly identify distorted cognitive representations at a mechanistic level (though progress is being made; [6•,7,8]), so identification of a cognitive bias requires inference from behaviour (e.g., [9•]). Note that the behaviour could simply be self-reporting of beliefs (although it is easy to mistakenly infer biases such as overconfidence from such reports [10]).

Many behaviours that initially appear irrational can be understood by taking background expectations and subsequent data into account from the perspective of the individual, as it is the expected payoff that determines whether a behaviour is optimal (rather than individual stochastic outcomes). The behavioural sciences have made great strides by using this 'optimality approach' to understand the distribution and timing of behaviours; for example, the marginal value theorem [11], drift-diffusion modelling [12]. However, not all behaviours are optimal; cognitive limitations mean that individuals will sometimes display sub-optimal cognitive biases even in their natural habitats. These behaviours are more challenging to explain from a theoretical perspective.

By dealing with sub-optimal behaviour in the natural habitat, I avoid the semantic subtleties of defining cognitive bias.[1] Such behaviours violate 'ecological rationality' [13] and, depending on interpretation, 'B-rationality' [14]. Rather than frequentist biases within a population (e.g., due to disorders [6•,15,16•]) or individual differences [17], I focus on the more difficult topic of cognitive biases at the species level. These may sometimes be understood at a holistic level by taking account of not only the behaviours themselves, but the mechanisms that would be necessary for alternative (or 'better') behaviours to be supported.

I first discuss apparent 'explanations' that are flawed.

Attempted explanations that do not address biased behaviours

Various papers have attempted to explain sub-optimal cognitive biases from an evolutionary perspective.

[1] My choice of definition would be a simplified version of that of Mineka & Sutton (1992): 'a selective or nonveridical processing of information'.

Table 1

The term 'cognitive bias' has somewhat different meanings.

Definition	Semantics/issues
Cases in which human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality [19••].	Any behaviour that is not reward-maximising could indicate a cognitive bias, even in non-natural lab settings. Thus, removing the term 'human' from the definition to study other animals, we find that an animal responding optimally (relative to its natural setting) may be regarded as cognitively biased.
An inaccurate view of the world [27].	Any behaviour that is not reward-maximising could indicate a cognitive bias, even in non-natural lab settings.
A systematic pattern of deviation from norm or rationality in judgement, whereby inferences about other people and situations may be drawn in an illogical fashion [55].	Although this definition does not specify what 'rational' means, adaptive behaviour may not be regarded as cognitively biased.
Effects of emotional state or trait on cognitive processes. [56]	Any behaviour can be inferred to show a cognitive bias if emotions are deemed to be a part of the decision-making process.
Any selective or nonveridical processing of emotion-relevant information ([57], based on a longer version, [58], specific to fear and anxiety).	It is unclear why the authors included 'emotion-relevant' in their definition, except that they were dealing with emotions in their paper.
Interesting cognitive biases obtain when beliefs depart systematically from those of an agent with Bayesian beliefs [26].	The authors point out that just about any decision-maker does not assign equal probability to every possibility — and that to do so would be a mistake.

However, several aspects have caused confusion, resulting in some authors attempting to explain sub-optimal behaviours (such as hazardous decisions, assumed to be caused by a cognitive bias of overconfidence) using models where the behaviour is optimal. A brief history helps to clarify the issues.

An influential paper by Haselton *et al.* [18,19••] identifies three classes of explanation for cognitive biases, which they term Heuristic, Error Management and Artefact biases. They describe *heuristic biases* as being due to information processing constraints (possibly due to phylogeny), resulting in mechanisms being used which fail to produce rational behaviour in systematic ways. *Error management biases* are produced by natural selection taking account not only of the probabilities of errors when taking particular actions, but the expected payoffs associated with those actions. Artefact *biases* are due to individuals being tested in non-natural settings, leading to non-rational processing of a problem.

The second of these, Error Management Biases, has received substantial attention. The mathematical basis of Error Management is signal detection theory [20], which shows how signal distributions can be combined with payoffs to set optimal thresholds for behaviour. Unfortunately, in signal detection theory the optimal threshold setting referred to as the 'bias'. This label refers to the setting in terms of the probabilities, rather than payoffs (or utilities).

Actions often differ in expected costs and benefits, so optimal decisions need not minimise probability of error. This is easily seen by example. Suppose that you have the opportunity of gaining some money by calling the outcome of the flip a coin. For correctly calling heads, you win £2; correctly calling tails, you win £1; you get nothing for calling incorrectly. Using a fair coin, the expected payoff is maximised by calling heads. Even if the coin were slightly bent, so the probability of heads was only 40%, the optimal behaviour is to call heads ($0.4*£2 > 0.6*£1$) even though the probability of calling correctly is maximised by calling tails. Looking *only* at the probabilities, rather than the expected payoffs, calling heads would appear biased.

This kind of signal detection 'bias' produces optimal behaviour, so it is not correct to infer a sub-optimal cognitive bias. This semantic confusion is exacerbated by it being easy to think in terms of local goals or probabilities, rather than overall utilities. For instance, Haselton *et al.* [18] identify overperception of sexual interest (typically by males) as a cognitive bias, which can be explained simply by signal detection theory, as they recognise. (Similar examples abound in the animal kingdom [21,22], including the possibility that some traits evolve to exploit existing cognitive biases in others[23,24].) Although an interesting topic, and one that falls under a wide definition of cognitive bias [2], this is not an explanation of sub-optimal behaviours (as identified by the authors [25] and others [26]).

In combination with the semantic issue of bias, some authors have assumed a sub-optimal mental mechanism (i.e., a sub-optimal component of the decision-making system) and then inferred a cognitive bias by showing that the individual would do better if the mechanism treated probabilities in a biased manner. Without any basis for assuming such a mechanism in the first place, this is a nonsensical approach to trying to understand cognitive biases that produce sub-optimal behaviour; Marshall *et al.* [27,28] correct some of this literature.

Others, such as Gigerenzer's ABC group, have suggested that we should not attempt to understand behaviours as optimal, arguing that optimal rules tend to be