Post-decisional accounts of biases in confidence
Joaquin Navajas¹, Bahador Bahrami¹ and Peter E Latham²

Most models of decision-making suggest that confidence, the ‘feeling of knowing’ that accompanies our choices, is constructed as the decision unfolds. However, more recent studies have noted that processes occurring after we commit to a particular choice also affect this subjective belief. This leads to the following question: when are we better judges of ourselves? If, after a decision, evidence continues to accumulate in an unbiased manner, then our confidence judgements should improve. Conversely, if post-decisional information processing is biased, our sense of confidence could be distorted, and so our confidence judgements should degrade with time. We briefly discuss recently proposed models of post-decisional evidence accumulation, and explore whether, and how, biases in confidence could arise.

Addresses
¹ Institute of Cognitive Neuroscience, University College London, London WC1 N 3AZ, United Kingdom
² Gatsby Computational Neuroscience Unit, University College London, London W1 T 4JG, United Kingdom

Corresponding author: Navajas, Joaquin (j.navajas@ucl.ac.uk)

Current Opinion in Behavioral Sciences 2016, 11:55–60
This review comes from a themed issue on Computational modeling
Edited by Peter Dayan and Daniel Durstewitz
For a complete overview see the Issue and the Editorial
Available online 18th May 2016
http://dx.doi.org/10.1016/j.cobeha.2016.05.005
2352-1548 (c) 2016 Elsevier Ltd. All rights reserved.

Introduction
Humans and other animals integrate noisy sensory input to infer the state of the world, and guide action and choice [1]. Action selection is accompanied by a ‘sense of confidence’, a subjective feeling about the validity of the choice [2]. Much of the psychology and neuroscience of decision making has focused on understanding the computations that underlie this subjective belief. Several different models for computing confidence have been proposed (signal detection theory [3,4], sequential sampling [5–7], Bayesian inference [8,9], heuristics [10], etc.) and they have been compared with explicit reports in humans [4,5,6,8,9] and with implicit estimates of confidence in non-human animals [3,7]. Until recently, most of these models assumed that confidence is a decisional process, that is, that it is computed by the same circuitry that drives choice or, at the very least, that it is constructed during the decision. This assumption rests on a vast corpus of neurophysiological evidence in rodents [3] and monkeys [7] showing that changes in stimulus reliability (e.g., the coherence of moving dots) modulate the firing of neurons that predict both choice accuracy and confidence [11].

In sharp contrast to this perspective, several more recent experiments have concluded that our sense of confidence is also determined by processes that occur well after we commit to a choice [12,13,14–16,17,18,19]. This observation leads to several questions: What are the consequences of such post-decisional processing of confidence? How does it affect the accuracy of this subjective belief? For example, should we trust our immediate (gut) feeling of confidence, or is it better to take our time and ‘gain perspective’? Here, we review state-of-the-art models of confidence and explore possible answers to these questions. In particular, we focus on how post-decisional processes affect our ‘metacognitive accuracy’, namely, the extent to which our confidence is consistent with our probability of being correct [20]. Far from being idle curiosity, knowing when we are better judges of ourselves could benefit us in several ways: it could help us cooperate effectively [21,22] and reduce aversive counterfactual thinking [23] that otherwise leads to negative emotions such as regret [24]. In addition, knowing the right time to gauge the validity of our choices is essential for minimizing distortions of confidence [25] wherein confidence is no longer predictive of objective accuracy (Box 1). These include overconfidence [26] and confirmation bias [27]; both are systematically observed in human choices, and both contribute to poor judgement and bad decisions [28].

Biases in post-decisional processing
The most straightforward experimental evidence that subjects continue processing information after making a decision is that they often express a desire to reverse their initial choice [12,14]. These ‘changes of mind’ were observed both in simple perceptual decisions [12], and in a recognition memory task [14], and cannot be explained by models that disregard post-decisional processing. Given that evidence continues to accumulate after a decision, it would not be surprising if confidence changed as well. And indeed, confidence sometimes depends on the length of the inter-judgement interval, that is, the amount of time between making a decision and giving a confidence rating on that decision [16]. In line with these observations, recent studies have suggested that post-decisional neural signals correlate with [18] and causally drive [19] confidence judgements.
There have been several proposals to account for post-decisional evidence accumulation and for changes in confidence [13–15, 16, 17]. In our view, the most promising proposal involves a two-stage dynamic signal detection theory [15]. This is mainly due to its simplicity and applicability to a wide range of different scenarios, including perceptual choices [29], general knowledge questions [17], and value-based decisions [30]. In the first stage, a decision variable is accumulated, and choice is typically guided by the sign of that variable; the first stage ends at the time of the decision. In the second, post-decisional, stage, the decision variable continues evolving, and its absolute value determines confidence (Figure 1a). Post-decisional processing changes our confidence in the selected option, and might either confirm or reject the first choice. Different two-stage models differ primarily in how the agent accumulates evidence after choice [17]. Critically, the different hypotheses make different predictions for how post-decisional processing changes metacognitive accuracy.

For example, the decision-maker could continue accumulating evidence after choice in an unbiased manner. In that case, after making a correct decision, more evidence should provide further support for the choice and boost confidence. Conversely, if an error was made (e.g., due to noise in the process of evidence accumulation), post-decisional evidence will typically oppose the chosen option and, as a consequence, confidence in the decision will decrease. In either case, as more and more post-decisional evidence is accumulated, eventually the difference in confidence between correct and incorrect trials becomes large. Thus, longer inter-judgement intervals will both improve accuracy and confidence.

In general, accumulating post-decisional evidence is a good strategy to refine estimates of confidence, especially in rapidly changing environments where later samples carry more information than earlier samples. This is assuming that evidence is integrated without bias. However, several studies have shown that post-decisional processing could be biased, and so could distort confidence judgments [17, 26, 31, 32, 33, 34]. For example, evidence for the chosen option could be outweighed

---

**Box 1 Distortions of confidence**

From a normative viewpoint, an appealing property for a system that reflects the validity of its choices is to be ‘well-calibrated’, that is, to express high confidence only when it is likely to be correct and low confidence otherwise. Because confidence ratings are metacognitive judgements (i.e., decisions about decisions) this property is also known as having high ‘metacognitive accuracy’ [32]. Many experiments in real-life settings have shown, however, that humans are often miscalibrated. For example, we might ignore evidence contradicting the option that we chose (‘confirmation bias’ [27]), increase our confidence in predicted outcomes that seem to have a consistent pattern (the ‘illusion of validity’ [10]), underestimate our probability of being correct in hard scenarios, and overestimate it in easier situations the ‘hard-easy effect’ [28]). Among this rich repertoire of cognitive illusions, the most widespread is ‘over-confidence’ [28]. This bias is particularly harmful when it is present among experts, such as forecasters [33] and policy makers [34], and a deeper understanding of its cognitive origin may help us guard against it.

---

**Figure 1**

Two-stage models of categorical decisions. (a) Models a two-alternative choice between options ‘A’ and ‘B’. Evidence is accumulated by tracking a decision variable over time. At a given time, the decision is made, and the sign of the decision variable determines choice. After the decision, evidence continues to be accumulated during the inter-judgement interval, that is, until the confidence judgement is made. Solid black lines sketch two examples of unbiased processing in two different trials: one leading to the selection of option ‘A’ and the other one to choosing ‘B’. Coloured lines depict deviations from this optimal (unbiased) process. Red lines show the temporal evolution of the decision variable in the presence of a confirmation bias; blue lines show the evolution in the presence of post-decisional decay. (b) Serial dependencies can be modelled as an initial bias in the decision variable that is contingent on the previous choice. Solid black line: example of unbiased processing in one trial. Dotted black line: same example with serial dependence if the previous choice was ‘B’. This figure was inspired by the model described in Ref. [17], based on decisions that lasted ~500 ms. Similar models were implemented in other choices with a timescale of a few seconds [15].
Confidence in continuous variables

The models depicted in Figure 1 deal with two-alternative choices, and can be extended to categorical decisions with a larger number of options [35]. A very different problem occurs when participants need to estimate a continuous variable such as orientation [33**] or probability [8**]. Implementing a Bayesian perspective [8**,9**,36], subjective beliefs can be modelled as a probability distribution that evolves over time throughout the course of the decision. To determine this distribution, the decision-maker needs to track, at the very least, its mean and variance [37].

Figure 2a sketches this process in the absence of biases. As more evidence is accumulated, the estimated mean converges to the true value while variance decreases. Confidence, in this scenario, should reflect the uncertainty encoded by the probability distribution; namely, its inverse variance or precision [8**,38]. A recent study has shown that human subjects do indeed learn to estimate probability (a continuous quantity) similarly to an ideal Bayesian observer, and report their internal precision as confidence [8**].

To the best of our knowledge, there are no studies that manipulate the length of the inter-judgement interval in these types of tasks. But, assuming that evidence continues to be accumulated after choice (as in categorical choices [13*,15,17*]), then unbiased processing would predict both more accurate and more confident estimates. Biased processing, on the other hand, could lead to either a reduction in variance (corresponding to confirmation bias; red trace in Figure 2b) or an increase in variance (corresponding to underconfidence; blue trace in Figure 2b).

Origins and function of post-decisional biases

Empirical evidence suggests that post-decisional processes affect our sense of confidence and influence subsequent decisions. Some of these processes clearly arise from a finite cognitive capacity. For example, in studies in
which sensory stimulation is turned off after choice, post-decisional decay in accuracy may be due to the transient nature of working memory [39]. The lack of perceptual input in these studies may lead to greater uncertainty and a reduction in confidence. Several studies have also tested conditions in which the stimulus remains available after choice [13,15,17]. It would be interesting, however, to see more experiments testing both conditions and manipulating the availability of perceptual evidence during the inter-judgement interval (e.g., [13∗]). This would make it possible to determine which paradigms are likely to elicit post-decisional decay and which ones lead to confirmation bias.

Variations in post-decisional bias can be attributed to individual differences in metacognition. Previous research showed that healthy adults differ in their metacognitive accuracy [40], in their tendency to be under or overconfident, and in the shape of their distribution of confidence ratings [41]. These features were linked to individual differences in brain structure [40], function [42], and personality trait [41]. For example, scoring high in optimism correlates with the tendency to be overconfident [41]. It would not be surprising if these individuals were also more prone to post-decisional biases that inflate confidence such as confirmation bias, but experiments testing this have not been performed yet.

Confirmation biases could also be a consequence of finite cognitive resources which results in the use of heuristics [43,44]. One proposal posits that humans can contemplate only one hypothesis at a time, and that they implement a ‘positive-test strategy’ [45]. This approach assumes that a given hypothesis is true and rejects it only if there is sufficient evidence against it. Positive-test strategies are much more liberal than most statistical tests, which assume exactly the opposite (the ‘null hypothesis’) precisely to avoid false positives. Other studies emphasise motivational aspects of the confirmation bias, such as our desire to believe in propositions that we would prefer to be true. For example, people may hang on to beliefs that are categorically wrong to minimise cognitive dissonance [46], even in the light of overwhelming evidence against them [47]. Yet another explanation argues that decision-makers are pragmatic, and that confirmation bias might be optimal in certain real-life scenarios [43]. According to this view, humans might not be so concerned about determining the veracity of different hypothesis as they are about minimising the odds of making a costly mistake. If the negative consequences of assuming that a particular hypothesis is false are larger than the positive ones associated with accepting it as true, then the strategy that maximises reward would also exhibit a confirmation bias (see [43,44] for real-life examples of this situation).

Finally, serial dependencies (Figure 1b) can lead to bias in laboratory experiments (where evidence is often independent and identically distributed), but they may be a good strategy in more realistic conditions, where noise is structured differently [33∗,34]. Because the statistical properties of the physical world are temporally stable (for example, low-level properties in a natural scene do not vary randomly over time, making the past a good predictor of the future), the brain might be tuned to exploit these regularities in the environment. This principle was demonstrated in the visual system both in the processing of orientation [33∗] and face identity [34]. Future research should explore whether this effect is also present in other sensory modalities, and whether it affects our sense of confidence.

Although conditions exist where biases improve decision making, they always distort confidence judgments. This is because, from a normative perspective, confidence should reflect the probability of being correct [20]. In this context, processing information with bias implies under or overweighting evidence for the chosen option (regardless of its validity) leading to suboptimal estimates of confidence (Box 1).

Concluding remarks
We focused on whether and how post-decisional processes influence our sense of confidence. In particular, we discussed a recent class of theories based on sequential sampling methods which allow decision-makers to continue accumulating evidence after choice [13,15,17]. This framework can account for a wide range of behavioural patterns, such as changes of mind [12], improvements in metacognitive accuracy with increasingly long inter-judgement intervals [13∗], and serial dependencies [33∗]. This framework also explains distortions of confidence (Box 1), such as under and overconfidence, as a consequence of biased processing taking place after choice.

One of the most intriguing aspects of two-stage models are their predictions for neural data. As an extension of accumulation-to-bound models, one would expect that neural signals indexing evidence accumulation (e.g., the firing of neurons in the macaque lateral intraparietal sulcus [11]) should continue evolving during the inter-judgement interval. This result has not yet been reported. One possible explanation could be that, until recently, neural signatures of evidence accumulation were found only in non-human animals, where confidence judgements are obtained indirectly (see [48] for a review of different techniques for indirectly measuring confidence in animals). Hence, testing this prediction might be more suitable for an experiment combining explicit reports in humans with M/EEG recordings as the analogous counterpart for the firing of intraparietal neurons [49,50]. In fact, a very recent study found that these signals indeed continued evolving after choice, guiding confidence...
judgments [51]. Further research is needed to identify the neural sources contributing to this process.

**Conflict of interest statement**

Nothing declared.

**Acknowledgements**

JN and BB are supported by the European Research Council StG (NEUROCODEG, #309865); PEL is supported by the Gatsby Charitable Foundation.

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


2. Grimaldi P, Lau H, Basso MA: There are things that we know that we know, and there are things that we do not know we do not know: confidence in decision-making. Neurosci Biobehav Rev 2015, 55:88-97.


This study asked subjects to state their decisions and confidence in an orientation discrimination task in which signal strength and variance were manipulated in a factorial manner. Decision accuracy was influenced by both factors but confidence ratings were best explained by a signal detection model that underweights stimulus variability and leads to overconfidence. This study shows how certain distortions of confidence (Box 1), typically observed in real-life decisions, are also present in simple perceptual choices.


In a learning task, subjects were asked to estimate transition probabilities between two stochastic events and to give confidence ratings on those estimates. Confidence was influenced by two different types of uncertainty: ‘expected’ uncertainty due to the stochastic nature of the environment, and ‘unexpected’ uncertainty given by changes in the stochastic characteristics of the environment. Overall, these results point out a crucial interplay between subjective confidence and probabilistic learning.


This study implemented Bayesian model selection to test whether confidence ratings in perceptual decisions reflected Bayes-optimal or simpler heuristic computations. Authors found evidence for the former in a standard experimental design where confidence was asked after choice. However, a less standard setting in which subjects stated their confidence and choice simultaneously prevented optimality. These results show how small changes in the experimental setting (such as asking confidence at different timings) can have large consequences in the interpretation and meaning of subjective confidence ratings.

26. Lichtenstein S, Fischhoff B: Do those who know more also know more about how much they know? Organ Behav Hum Perform 1977, 20:159-183.


33. Fischer J, Whitney D: Serial dependence in visual perception. Nat Neurosci 2014, 17:738-743. The visual properties of the physical world are generally stable across time; and it would be reasonable to expect that the brain has a mechanism to exploit this regularity. Indeed, this study shows that human visual perception is serially dependent: the visual system uses prior and present information to inform perception. This leads to a response bias in laboratory experiments where consecutive trials are independent. However, this might be a good strategy in more ecological scenarios where noise is structured differently.


45. Klaiman J, Ha Y: Confirmation, disconfirmation, and information in hypothesis testing. Psychol Rev 1987, 94:211.


