CellPress REVIEWS

Spotlight

Confidence in the Real World

Dobromir Rahnev[1,*]

Humans often assign confidence to multioption decisions, but most computational research only uses two-alternative tasks. In a new study, Li and Ma begin to reveal the mechanisms of confidence generation in multialternative tasks. This research should inspire further experiments on how humans assign confidence judgments in real-world situations.

Here is a little pop quiz: (i) did India gain independence before or after the Second World War? (ii) Which planet is closest to the Sun? (iii) Which country is Santiago the capital of? Did you make your guesses? Now, for each question, rate your level of confidence.

How do humans assign confidence to their responses? This question is the topic of investigation of a burgeoning literature focused on revealing the computational foundations of confidence reports. A critical issue that our little quiz reveals is that not all decisions for which we rate confidence are created equal. The three quiz questions above required you to choose between two, eight, and ~200 possible answers, respectively. Furthermore, the first question explicitly defined the two options, but most of us do not even know what all of the options for the last question are. Questions from the last type have been investigated for >50 years [1], but their complexity almost completely precludes a rigorous computational approach. Therefore, recent years have seen an explosion of research using two-choice tasks to achieve an increasingly sophisticated computational understanding of confidence [2–4]. While this research has made large strides in our understanding, some mechanisms deemed plausible for two-choice tasks may be implausible for, or not even apply to, 200-choice tasks. Therefore, truly understanding how confidence is generated requires that we expand beyond two-choice tasks but maintain the level of computational rigor that they have made possible (Figure 1).

Recently, Li and Ma [5] took an important first step in that direction by creating a paradigm that allowed precise computational modeling of a three-choice task. Subjects were asked to judge which of three clouds of differently colored dots generated a specific black dot and then to rate their confidence. To make the task intuitive, subjects were encouraged to think of the problem as deciding which of three groups of people wearing different colors a random person in a crowd belonged to. Li and Ma could then adjust the exact parameters of the clouds, which enabled them to build and compare specific computational models of how confidence is generated in this task.

The authors compared three main models. The first model postulated that confidence ratings reflect the probability of being correct, the dominant view that has emerged from the literature on two-choice tasks [6,7]. The second model postulated that confidence instead reflects the difference in the posterior probability of the top two options. Finally, the third model postulated that confidence ratings reflect the entropy (i.e., the overall uncertainty) of the decision. The data were best explained by the second model. This finding was robust across several different variations of the models and across three different experiments that varied the arrangements of the dot clouds and the availability of trial-by-trial feedback.

What do these results mean in practice? Consider a particular trial in which you estimate the posterior probabilities of the three options to be 50%, 45%, and 5%. Now consider a different trial where you estimate these probabilities to be 40%, 30%, and 30%. In both cases, you would choose the first (most likely option), but which decision would you be more confident in? The first and third models, which postulate that confidence reflects probability correct or entropy, predict that your confidence will be higher in the first decision (because the posterior probability of 50% is higher than 40%, and because the first distribution has smaller overall uncertainty). However, Li and Ma's results suggest that you will instead have higher confidence in the second decision because of the larger difference between the top two options (50 minus 45 is smaller than 40 minus 30). As suggested by the authors, an intuitive interpretation of this phenomenon is that confidence judgments reflect the probability that the chosen alternative is the best possible option rather than a very probable one.

Nevertheless, multialternative decisions in the real world have added complexity that Li and Ma's computational model may not be able to capture. For example, consider how you reported your confidence in the quiz questions (for those curious, the correct answers are after WW2, Mercury, and Chile). Li and Ma's model implies that your confidence was determined by the difference in the posterior probability of your top two alternatives. However, computing posterior probabilities may be difficult or impossible in situations such as the third quiz question,

CellPress REVIEWS

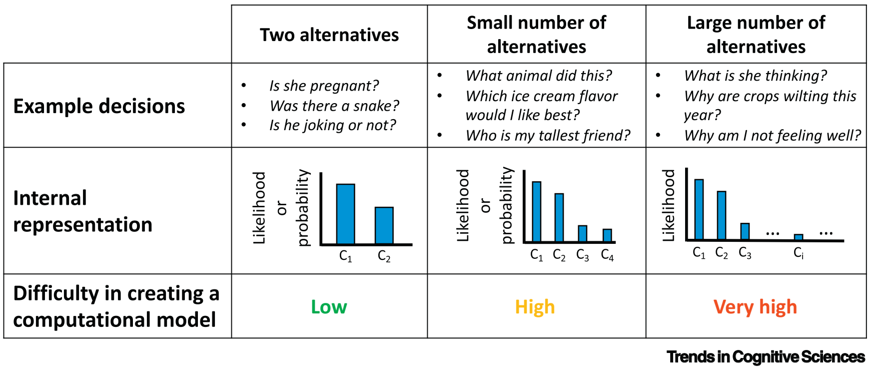

Figure 1. Three Types of Decisions with Discrete Alternatives. Decisions with just two alternatives, small numbers of alternatives (e.g., three to high single digits), and high numbers of alternatives (e.g., double digits or more) are all ubiquitous in real life. Confidence in two-choice decisions is easiest to understand computationally, but insights may not generalize to other situations. Thus, decisions with small numbers of alternatives provide a computationally tractable way for developing a general theory of confidence generation. Nevertheless, confidence for decisions with large numbers of alternatives may involve yet new mechanisms because posterior probability is difficult to compute for them.

where all possible answers are not even known in advance (Figure 1). Thus, confidence may often need to be based directly on the likelihoods rather than the posterior probabilities. In fact, Li and Ma showed that a model where confidence was based on the ratio of the likelihoods for the top two options performed almost as well as the winning model. In addition, contrary to Li and Ma's winning model, the exact values of the non-normalized likelihoods may also be important. For example, recent research [8,9] suggests that confidence is higher when the top two options both have high likelihoods (e.g., 'I was guessing between Mercury and Venus') than if all options have close to zero likelihood ('I just can't think of any country'). However, doubling the likelihood of all options results in the exact same posterior probabilities (because the computation of posterior probabilities involves normalizing the likelihoods) and, hence, does not influence confidence in Li and Ma's model. Finally, the mechanisms governing confidence may be different in purely perceptual tasks with brief stimulus presentations, where the confidence computations would strongly depend on the exact nature of the internal representation [10]. All of these possibilities should be directly explored in follow-up studies.

Thus, multialternative decisions likely involve additional mechanisms that are not considered in Li and Ma's winning model. However, their paper represents a critical first step and proves that rigorous computational modeling does not necessitate using a two-choice task. We may have just entered a new era where sophisticated computational models are constructed for tasks that mimic the multialternative choices ubiquitous in the real world.

[1]School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

*Correspondence:
rahnev@psych.gatech.edu (D. Rahnev).

https://doi.org/10.1016/j.tics.2020.05.005

References

1. Fischhoff, B. *et al.* (1977) Knowing with certainty: the appropriateness of extreme confidence. *J. Exp. Psychol. Hum. Percept. Perform.* 3, 552–564
2. Drugowitsch, J. *et al.* (2019) Learning optimal decisions with confidence. *Proc. Natl. Acad. Sci. U. S. A.* 116, 24872–24880
3. Odegaard, B. *et al.* (2018) Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proc. Natl. Acad. Sci. U. S. A.* 115, E1588–E1597
4. Denison, R.N. *et al.* (2018) Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proc. Natl. Acad. Sci. U. S. A.* 115, 11090–11095
5. Li, H.-H. and Ma, W.J. (2020) Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nat. Commun.* 11, 2004
6. Sanders, J.I. *et al.* (2016) Signatures of a statistical computation in the human sense of confidence. *Neuron* 90, 499–506
7. Meyniel, F. *et al.* (2015) Confidence as Bayesian probability: from neural origins to behavior. *Neuron* 88, 78–92
8. Peters, M.A.K. *et al.* (2017) Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat. Hum. Behav.* 1, 0139
9. Zylberberg, A. *et al.* (2016) The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *eLife* 5, e17688
10. Yeon, J. and Rahnev, D. The suboptimality of perceptual decision making with multiple alternatives. *bioRxiv* Published online March 25, 2020. https://doi.org/10.1101/537068.

Spotlight

Gender Equality and Gender Gaps in Mathematics Performance

Gillian R. Brown[1,*] and Kim Alexandersen[2]

In a recent analysis of mathematics performance (Eriksson *et al.* 2020), national gender egalitarian values were positively associated with an increase in the average mathematics scores of high-school boys relative to girls. This study highlights that progressive gender egalitarian values at a national level might not translate into equality of opportunity at an individual level.

The long-standing expectation that, on average, girls have lower mathematical abilities than boys has been undermined by several lines of empirical evidence; for instance, a meta-analysis has shown that the gender gap in average mathematics performance is close to zero [1], and the