Becoming Confident in the Statistical Nature of Human Confidence Judgments

Jan Drugowitsch[1,*]
[1]Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: jan_drugowitsch@hms.harvard.edu
http://dx.doi.org/10.1016/j.neuron.2016.04.023

In this issue of *Neuron*, Sanders et al. (2016) demonstrate that human confidence judgments seem to arise from computations compatible with statistical decision theory, shining a new light on the old questions of how such judgments are formed.

There is little doubt that confidence is an essential component of decision making. I would not cross the road if I were not sufficiently confident in making it safely to the other side. I would aim my return in tennis just far enough inside the court to be confident that it actually lands inside while making it hard for the opponent to reach.

An essential ingredient in such decisions is that they involve some degree of uncertainty (Doya et al., 2007). My return in tennis, for example, won't always exactly go in the planned direction. In the presence of uncertainty, however, decisions become a matter of statistics: which of the available options is more likely to yield the desired effect? The associated decision confidence then ought to follow the same statistical principles: how probable is it that the chosen option was the correct one? In statistical decision theory (Berger, 1993), this would be a very natural definition of confidence (Pouget et al., 2016).

The psychological study of the human sense of confidence, however, follows another tradition. There, confidence is interpreted as arising from a homunculus-like monitoring of the decision process, such that confidence judgments become a form of metacognition (e.g., Lau and Rosenthal, 2011). This is also reflected in heuristic models of decision confidence, where the process generating confidence judgments is frequently added onto the decision-making process (e.g., Pleskac and Busemeyer, 2010), rather than being an integral component of it. While for some parameter regimes these models might be able to qualitatively mimic the predictions of statistical decision theory, little work has attempted to test if human confidence judgments can be directly predicted by statistical decision theory.

In this issue of *Neuron*, Sanders, Hagya, and Kepecs (Sanders et al., 2016) do exactly that: they ask if human reports of confidence can be quantitatively explained by confidence computed according to statistical principles. In order to test this hypothesis, they first derive four predictions that need to hold if confidence indeed arises from statistical computations, and then compare these predictions against confidence judgments of humans performing two types of decision-making tasks. What they find is that, indeed, human confidence judgments in both tasks seem to arise from such statistical computations. This is good news, as it not only puts the study of confidence on a firm statistical basis, but also means that our sense of confidence is statistically consistent: it is the optimal quantity to use for many of the computational roles of confidence, such as learning, postdecision wagering, etc. (Meyniel et al., 2015; Pouget et al., 2016).

Let us consider the four predictions for confidence judgments that arise from statistical decision theory. The first is that the level of confidence ought to predict the accuracy of the associated choices (Figure 1A). That is, the higher the confidence, the more accurate the choices. If confidence is expressed directly as a posterior probability from Bayesian statistics, then the prediction becomes even stronger. In that case, the level of confidence should perfectly match the probability of making correct choices.

While the first prediction is fairly intuitive, the second and third are less so. The second states that the level of confidence should be larger for easy than for hard choices if these choices are correct, while the opposite should be the case for incorrect choices (Figure 1B). To understand this prediction, we need to consider that we are operating with uncertain evidence. This implies that what Sanders and colleagues call the ‘‘percept’’ of the evidence (that is, the brain's estimate of the evidence) does not perfectly match the actual evidence (associated with the correct choice), but instead is a noisy sample thereof. For hard choices, in which the evidence is weak, the percept might in some cases due to stochastic fluctuations strongly support one choice even if the actual evidence is in fact for the other choice (Figure 1D). For easy choices, in contrast, when the evidence is strong, such mistakes are less likely to happen. If they do, however, the percept will only weakly support the incorrect choice. For this reason, the confidence for incorrect choices should be higher for hard than for easy choices.

The same principle leads to the third prediction, which states that confidence should reach an average level once choices become impossibly hard, that is, once the decision maker is not provided any evidence toward which choice would be correct (Figure 1B). This prediction arises again because the percept is a noisy version of the actual evidence. This implies that even though there might not be any actual evidence, there might be some perceived evidence based upon which the decision maker commits to a choice. This perceived evidence causes the decision maker's confidence to reach an average level (Figure 1D, bottom).

The fourth prediction relates choice difficulty, choice accuracy, and confidence,

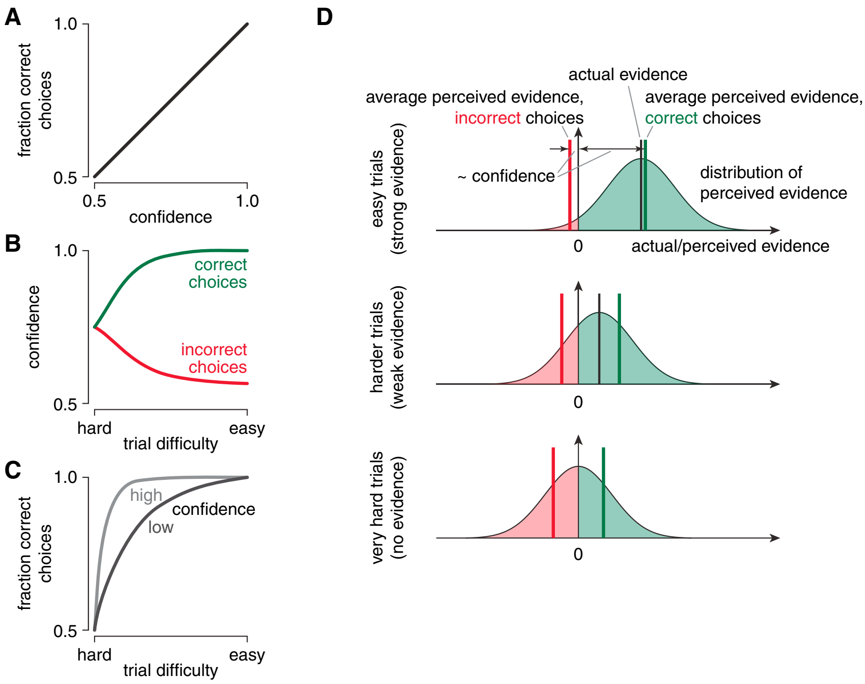

Figure 1. The Hallmarks of Confidence Judgments of Statistical Decision Theory

(A) If confidence judgments follow statistical decision theory, then for a fixed level of confidence, the fraction of correct choices should equal the confidence. Confidence could be measured in other units than probabilities (e.g., integer scales, as in Sanders et al., 2016), but even then, an increase in the level of confidence should result in an increase in the fraction of correct choices.

(B) For correct choices, confidence should be higher for easier trials. The opposite should be the case for incorrect choices. Impossible trials (no choice-related evidence) should result in an average level of confidence.

(C) For a fixed trial difficulty, high-confidence choices should be more accurate than low-confidence choices.

(D) To illustrate the intuition behind (B), consider that the evidence perceived by the decision maker is a noisy version of the actual evidence, illustrated here by plotting the distributions over perceived evidences across trials (smooth Gaussian distribution) for a fixed actual evidence (black vertical bar). Positive (negative) perceived evidence here results in correct (incorrect) choices, such that the average perceived evidence associated with correct (incorrect) choices is the center of mass (green/red horizontal bar) of the green-shaded (red-shaded) part of the distribution. In this simple model, a large (small) distance of this average perceived evidence to the origin corresponds to high (low)-confidence choices. The plots from top to bottom show increasingly more difficult trials, corresponding to moving from right to left in (B).

and states that for the same choice difficulty, the choice accuracy should be higher for decisions with higher confidence (Figure 1C). In that sense, it is similar to the first prediction, only that it refines this prediction to additionally include choice difficulty.

In order to test these predictions, Sanders and colleagues test human subjects on two decision-making tasks. The first task is a purely perceptual task in which in each trial the subjects hear sequences of clicks of a different click rate (average number of clicks per second) in each ear and need to indicate which side is associated with the higher rate (Brunton et al., 2013). After each choice they were furthermore asked to indicate their confidence in this choice on a scale from one ("random guess") to five ("high confidence"). Analyzing these confidence judgments, Sanders and colleagues found that they qualitatively followed all the criteria outlined by the above predictions. Higher confidence was associated with higher choice accuracy, even when conditioned on the difficulty of the choices (that is, the difference in click rates across ears). Furthermore, confidence for easier trials was higher for correct choices, but lower when choices were incorrect. Impossible trials, in which the click rate was the same for both ears, led to an average confidence level.

While these tests already provided a good qualitative match to the theory, Sanders and colleagues went a step further and looked for a quantitative match. First, they provided what they called "parameter-free" predictions of the statistical model. While the model actually had parameters, these were fully constrained by only considering the subjects' choices while ignoring their confidence judgments. The resulting confidence judgment predictions were close to, but did not fully match, those made by the subjects. In particular, the model predicted more extreme judgments (that is, closer to one and five). Second, Sanders and colleagues thus introduced noise that caused the model to make a fraction of confidence judgments by chance rather than based on the perceived evidence. With this adjustment, they finally achieved a good quantitative match between model predictions and observed confidence judgments.

The same procedure was repeated for data from subjects performing a second task that involved general knowledge rather than perceptual decisions. In this task, subjects were in each trial asked to decide which of two countries had the larger population. As in the first task, they were additionally queried after each choice for their level of confidence. Performing the same analysis as for the first task, Sanders and colleagues found that the subjects' confidence reports again match the predictions of the statistical model both qualitatively and quantitatively. This time, they didn't even need to add additional noise to the model's quantitative predictions, but it is hard to tell if this was due to overall noisier confidence judgments or due to better model predictions than in the first task.

Overall, their work provides strong support for the idea that human confidence judgments indeed originate from computations of statistical confidence. This is consistent with the stance that the computations underlying human decisions implement ideal or approximate Bayesian statistical inference (Beck et al., 2012; Doya et al., 2007). However, it is at odds with previous reports of a mismatch between choice accuracy and the reported level of confidence (e.g., Juslin et al., 2000). Sanders and colleagues argue that this can be accounted for by details of their experimental design. First, they chose to limit the confidence reports to integer values between one and five rather than asking for explicit probabilities, as has been done previously. With this, they might have side-stepping issues of probability calibration and risk sensitivity. Second, they claim that asking for confidence reports after the choice rather than at the same time avoids divided attention between confidence reporting and choice. However, it has previously been shown that in the presence of a stream of evidence, as is the case in their click-rate discrimination task, evidence presented close to the choice might be processed further after choices have been made (Resulaj et al., 2009). This could cause confidence reports to be based on different evidence than that used to make the choices (Zylberberg et al., 2012). In this light, it remains to be seen if it was indeed the asynchronous choice and confidence report that caused these reports to match the statistical model. It might in fact explain why, for this particular task, confidence judgments appeared noisier than predicted by the noise-free model.

An interesting observation made by Sanders and colleagues is that the human confidence judgments in the click-rate discrimination task are a function of both the decision time (that is, the duration between stimulus onset and choice) and the difficulty of the trial, as measured by the difference in click rates across ears. This is compatible with previous empirical findings (Kiani et al., 2014), but at odds with simple ideal-observer models that predict that confidence should only depend on decision time, irrespective of trial difficulty (Drugowitsch et al., 2012; Kiani and Shadlen, 2009). These models are based on the same statistical framework as that of Sanders and colleagues, such that it needs to be clarified how these seemingly contradictory findings can be brought in line.

This should not distract, however, from their important main findings that human confidence reports indeed feature the same hallmarks as confidence computed according to the principles of statistical decision theory. Hence, these reports might arise from the same computations that underlie our decisions under uncertainty, thus suggesting confidence to be a central and integral component of everyday decisions, rather than just an afterthought.

REFERENCES

Beck, J.M., Ma, W.J., Pitkow, X., Latham, P.E., and Pouget, A. (2012). Neuron *74*, 30–39.

Berger, J.O. (1993). Statistical Decision Theory and Bayesian Analysis, Second Edition (Springer-Verlag).

Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Science *340*, 95–98.

Doya, K., Ishii, S., and Pouget, A. (2007). Bayesian Brain: Probabilistic Approaches to Neural Coding (MIT Press).

Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N., and Pouget, A. (2012). J. Neurosci. *32*, 3612–3628.

Juslin, P., Winman, A., and Olsson, H. (2000). Psychol. Rev. *107*, 384–396.

Kiani, R., and Shadlen, M.N. (2009). Science *324*, 759–764.

Kiani, R., Corthell, L., and Shadlen, M.N. (2014). Neuron *84*, 1329–1342.

Lau, H., and Rosenthal, D. (2011). Trends Cogn. Sci. *15*, 365–373.

Meyniel, F., Sigman, M., and Mainen, Z.F. (2015). Neuron *88*, 78–92.

Pleskac, T.J., and Busemeyer, J.R. (2010). Psychol. Rev. *117*, 864–901.

Pouget, A., Drugowitsch, J., and Kepecs, A. (2016). Nat. Neurosci. *19*, 366–374.

Resulaj, A., Kiani, R., Wolpert, D.M., and Shadlen, M.N. (2009). Nature *461*, 263–266.

Sanders, J.I., Hangya, B., and Kepecs, A. (2016). Neuron *90*, this issue, 499–506.

Zylberberg, A., Barttfeld, P., and Sigman, M. (2012). Front. Integr. Nueurosci. *6*, 79.

C9ORF72-ALS/FTD: Transgenic Mice Make a Come-BAC

Lindsey R. Hayes[1] and Jeffrey D. Rothstein[1,*]
[1]Brain Science Institute, Department of Neurology, Johns Hopkins University, Baltimore, MD 21205, USA
*Correspondence: jrothstein@jhmi.edu
http://dx.doi.org/10.1016/j.neuron.2016.04.026

For five years, since the landmark discovery of the *C9ORF72* hexanucleotide repeat expansion in ALS/FTD, a transgenic mouse model has remained elusive. Now, two laboratories (Liu et al., 2016; Jiang et al., 2016) report the development of BAC transgenic mice that recapitulate features of the human disease.

Discovered in 2011, the GGGGCC hexanucleotide repeat expansion (HRE) in *C9ORF72* is now regarded as the most common genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (DeJesus-Hernandez et al., 2011; Renton et al., 2011). The HRE, located in the first intron, consists of 2–30 repeats in the general population and can range from hundreds to thousands of repeats in affected patients. Difficult to clone and prone to germline and somatic instability, these large expansions have presented a technical hurdle for the development of transgenic