View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repository of the Academy's Library

Signatures of a statistical computation in the human sense of confidence

Authors: Joshua I. Sanders[1], Balázs Hangya[1,2] & Adam Kepecs[1,*]

[1] Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724 USA

[2] Lendület Laboratory of Systems Neuroscience, Institute of Experimental Medicine, Hungarian Academy of Sciences, Budapest, H-1083, Hungary

* Correspondance: kepecs@cshl.edu

Summary

Human confidence judgments are thought to originate from metacognitive processes that provide a subjective assessment about one's beliefs. Alternatively, confidence is framed in mathematics as an objective statistical quantity: the estimated probability that a chosen hypothesis is correct. Despite similar terminology, it remains unclear whether the subjective feeling of confidence is related to the objective, statistical computation of confidence. To address this, we collected confidence reports from humans performing perceptual and knowledge-based psychometric decision tasks. We observed two counterintuitive patterns relating confidence to choice and evidence: apparent overconfidence in choices based on uninformative evidence, and for erroneous choices, that confidence decreased with increasing evidence strength. We show that these patterns lawfully arise when statistical confidence qualifies a decision. Furthermore, statistical confidence quantitatively accounted for human confidence in our tasks without necessitating heuristic operations. Accordingly, we suggest that the human feeling of confidence originates from a mental computation of statistical confidence.

Introduction

The scientific study of confidence has emerged from different traditions, reflecting its dual manifestation as a subjective feeling and an objective forecast. In the psychological tradition, confidence is thought to arise from the monitoring of mental content; it is sometimes framed as a form of metacognition associated with other subjective human qualities, such as introspection, awareness and even self-reflective consciousness (Charles et al., 2013; Flavell, 1979; Kunimoto et al., 2001; Lau and Rosenthal, 2011; Metcalfe and Shimamura, 1994). A wealth of studies has confirmed that humans possess this ability, and have identified conditions under which confidence appears to be miscalibrated, predicting outcomes sub-optimally (Bar-Tal et al., 2001; Baranski and Petrusic, 1994; Björkman et al., 1993; Camerer and Lovallo, 1999; Griffin and Tversky, 1992; Juslin et al., 2000; Kvidera and Koutstaal, 2008; Moore and Healy, 2008; Olsson and Winman, 1996; Shea et al., 2014; Stankov, 1998). In fact, human confidence often does not appear to reflect the underlying performance, suggesting that it is generated by an error-prone heuristic computation (Gigerenzer and Goldstein, 1996; Koriat, 2012; Tversky and Kahneman, 1974).

A separate construct termed "confidence" has also been studied in many disciplines as a wholly objective mathematical quantity. Formally defined as the Bayesian posterior probability that a decision maker is correct, confidence refers to a computational tool used in statistical analysis to assess hypotheses based on noisy or unreliable evidence. This confidence formulation is central to statistical decision theory and can be exploited to improve machine learning algorithms (Schapire and Singer, 1999; Sollich, 2002). Statistical models have also been used successfully to account for the perceptual and motor systems in decision-making, which obey Bayesian principles when faced with uncertainty (Ernst and Banks, 2002; Fetsch et al., 2013; Fiser et al., 2010; Körding and Wolpert, 2004; Pouget et al., 2013; Stocker and Simoncelli, 2006; Trommershäuser et al., 2008). However, less is known about the degree to which these same principles can account for central cognitive processes such as confidence (Kepecs et al., 2008; Kiani and Shadlen, 2009; Komura et al., 2013; Tenenbaum et al., 2011).

The idea that the subjective sense of confidence avails a statistical likelihood readout to the decision maker has been suggested only sparsely as a conjecture (Griffin and Tversky, 1992). Indeed, the Bayesian

confidence computation is often the de facto working assumption in economic studies when comparing human confidence to an ideal accuracy predictor. However, numerous attempts to model human confidence algorithmically have only considered indirect correlates such as reaction time (Audley, 1960; Kiani et al., 2014), decision variable balance (De Martino et al., 2013; Drugowitsch et al., 2014; Insabato et al., 2010; Kepecs et al., 2008; Vickers, 1979; Wei and Wang, 2015), decision variable variance (Yeung and Summerfield, 2012) and post-decisional deliberation (Pleskac and Busemeyer, 2010). These models have successfully accounted for a range of psychometric and chronometric aspects of human confidence. Importantly, these algorithmic models can make qualitatively different predictions depending on parameter choices, and no unifying predictions have emerged that directly relate these models with statistical confidence (Drugowitsch et al., 2014).

Therefore we sought to better understand how the sense of confidence as a psychological construct relates to the statistical confidence formulation. First, we used the statistical formulation to generate empirically testable predictions relating confidence to choice correctness and evidence discriminability. We show that in two different decision tasks, human confidence reports satisfy these predictions. We further show that human confidence can be quantitatively accounted for with a single noise parameter added to model the confidence reporting stage of the statistical formulation. This striking similarity between human and statistical confidence suggests that a mental computation functionally equivalent to statistical decision confidence is performed by the brain, and manifests to humans as a subjective feeling.

Results

Decisions are commitments to a specific option among a set of possible alternatives. In statistical terms, this commitment can be viewed as a selection of a hypothesis (H_1, the alternative hypothesis) against all possible alternative choices (H_0, the null hypothesis). Thus the choice can be evaluated in terms of a hypothesis testing problem: the null-hypothesis, H_0, is that the choice is incorrect, while the alternative hypothesis, H_1, is that the choice is correct. This formalization immediately suggests a statistical definition for decision confidence as the Bayesian posterior probability, which quantifies the degree of belief in the chosen hypothesis. Thus we defined confidence, c, as the probability of the alternative hypothesis being true given the perceived evidence, referred to as the percept, $\hat{d}$, and the choice, ϑ:

$$c = P(H_1|\hat{d}, \vartheta)$$

According to this definition, decisions are based on the internal percept, $\hat{d}$, which is the decision maker's estimate of the corresponding external evidence, d. Hence, both choice and confidence depend on the quality of evidence informing the particular decision. However, since the choice can be a stochastic function (θ) of the percept: $\vartheta = \theta(\hat{d})$, confidence may depend on the combination of the percept and the choice. This way the definition generalizes over any theory of perception, relating the external evidence to the percept $P(\hat{d}, |d)$, and decision making, relating the percept to the choice $\vartheta = \theta(\hat{d})$. This definition of decision confidence is natural from a statistical perspective (Drugowitsch et al., 2014; Kahneman and Tversky, 1972) but difficult to apply to experimentally observed confidence because it is based on the percept, a variable internal to the decision maker (as illustrated in Figure 1A). Nevertheless, we show that this definition yields several strong, qualitative predictions about statistical decision confidence that can be empirically tested.

We first sought to generate testable, qualitative predictions about statistical decision confidence in terms of empirically measureable quantities. Therefore, we created a Monte Carlo simulation of the statistical confidence formulation (see methods). We used a deterministic decision rule such that choices were correct if the sign of the external evidence and percept matched. For the simulations we assumed Gaussian noise for perception, $P(\hat{d}|d)$ to generate an internal percept from an external stimulus on each simulation trial. First, we examined how accuracy was related to statistical confidence. We found that all levels of statistical confidence predicted the mean choice accuracy (Figure 1B).

Next, we examined how confidence varies with evidence of differing discriminability. We found that the mean statistical confidence for a given level of discriminability increases for correct and decreases for incorrect choices (Figure 1C). Interestingly, we found that the average confidence for trials with zero evidence discriminability is much higher than 0.5 (chance accuracy) - in this case, our model predicts confidence to be precisely mid-level at 0.75. The prediction of mid-range confidence holds when the noise model $P(\hat{d}|d)$ produces percepts symmetric about d, and when two other conditions are true, common in psychometric testing: first, the range of d yields decision accuracies spanning 0.5 (chance) to ~1 (perfect), and second, each trial's d is drawn from the range of d with equal probability.

We next asked whether confidence levels can be used to infer accuracy beyond the psychometric function, the proportion of correct choices as a function of evidence discriminability We divided choices into low and high confidence (based on a mean split) and found that for each level of evidence discriminability, accuracy for high confidence choices is greater than for low confidence choices (Figure 1D). Taken together these Monte-Carlo simulations illustrate four signatures of decision confidence in terms of externally quantifiable variables that can be experimentally examined.

Confidence reports about auditory perceptual decisions

Starting from these predictions we sought to examine whether the subjective feeling of confidence experienced by humans can be accounted for by the statistical definition of confidence. Under many circumstances, self-reported confidence can be modulated by personality traits and contextual factors that are difficult to account for. Therefore we used a perceptual decision task that provided precise experimental control over the discriminability of evidence sampled by our subjects on each trial, allowing us to compare the relationships among observable variables to the predictions shown in Figure 1. We designed a two-alternative forced choice perceptual decision task in which we varied auditory sensory evidence in a graded manner (Brunton et al., 2013; Sanders and Kepecs, 2012). Subjects listened to separate Poisson click streams delivered independently to each ear, and indicated the faster clicking stream within 3 seconds with a button press (Figure 2A). To construct trials with graded evidence discriminability, we varied the balance of left and right click rates from neutral evidence (50Hz/50Hz) to strong (65Hz/35Hz), resulting in subject accuracy ranging from chance to nearly perfect (Figure S1). After entering each choice, the click stream shut off and the subject was prompted to indicate their feeling of confidence in their choice on a 5-division scale between a random guess (1) and high confidence (5). Following each trial we indicated choice correctness. To achieve ~100μs click train offset with respect to the choice, we delivered clicks and captured responses with a modified (see supplemental methods) Pulse Pal device (Sanders and Kepecs, 2014). In total, we acquired responses from five human subjects (n=22,427 trials). Subjects were trained for two 1-hour sessions prior to testing, and performance was

used to adjust stimulus difficulty for the remaining sessions such that mean performance of each subject was ~80% (Figure S1; see methods). Trained subjects did not significantly improve or fatigue over time in the study; subject accuracy and confidence were consistent for the duration of the experiment, and within sessions (Figure S1).

Using these psychometric data we tested our key predictions. First, we observed that confidence strongly predicted choice accuracy (Figure 2B,E, S3), as had been shown previously (Jastrow and Peirce, 1884; Lichtenstein et al., 1981). Second, we found that self-reported confidence increased with evidence discriminability for correct trials, but *decreased* with discriminability on errors, in accordance with a counterintuitive prediction of statistical confidence (Figure 2C,F). Third, trials with zero evidence discriminability, where subjects performed at chance level, reliably elicited average confidence around 3 on the 5-division scale (Figure 2C,F), in agreement with the model prediction of mid-range confidence (0.75). Finally, low or high confidence reports predicted low or high choice accuracy respectively, at all levels of discriminability (Fig 2D,G). These confidence patterns were robust in each of our subjects (Figure S3), demonstrating a strong and consistent qualitative agreement between the human feeling of confidence and the properties of statistical decision confidence.

We also considered whether a heuristic model that derived confidence reports directly from reaction times could account for our data. Reaction time was inversely correlated with confidence reports (Figure S2). Confidence also varied with discriminability, when conditioned on reaction time, consistent with previous findings (Kiani et al., 2014). In contrast to confidence reports, which always decreased with discriminability on error trials, error trial reaction time showed an inconsistent pattern between subjects (Figure S2). These observations show that a reaction time heuristic cannot fully account for confidence reports (Audley, 1960).

Quantitative prediction of human confidence reports

Next we asked whether our framework could be used to quantitatively account for the confidence reporting data. First, we assumed that the noise corrupting the evidence presented to human decision makers was Gaussian. Thus, fitting the psychometric function with a cumulative normal distribution function yielded an estimate for each subject's average noise level (Figure S1). The best-fit noise parameter was then used to generate percepts for Monte-Carlo simulations (Figure 1 B-C), yielding simulated choices and confidence reports (see Experimental Procedures). The resulting confidence value distribution was mapped to the 5-division scale by dividing the cumulative distribution by percentile, to match the subjects' own cumulative distribution of confidence reports. This parameter-free model predicted the qualitative trends in human data remarkably well (Figures 2 and S3, thick lines). Next, to capture imperfect reporting of the internal sense, we introduced one free parameter (α), confidence reporting efficacy (see Experimental Procedures), which modeled noise corrupting the statistical confidence value. Adding this parameter to the statistical model and fitting to human confidence reports with the maximum likelihood method further improved the model's approximation of human confidence (Figures 2, and S3 thin lines). The optimal parameter for human confidence reporting efficacy ranged from $\alpha = 0.49$ to 0.72 with an average of 0.57 (Figure S3). Thus the normative statistical model not only captures the qualitative patterns in human confidence reports but provides a quantitative account that is mostly within behavioral variability using a single free parameter. This free parameter provides a measure of the degree to which a subject's confidence reports reflect the noise-free statistical computation of

confidence. This new metric is a measure of confidence fitness; however we caution that its interpretation is not a correctness percentage, because resampling exploits the subject's distribution of reports.

Confidence reports about general knowledge in untrained subjects

Thus far, statistical confidence described subjects who were asked to perform a well-learned sensory decision task with feedback. Therefore we wondered whether the model could also describe confidence in more typical human decisions. To assess this we tested 27 additional subjects in a single-session decision making task. Unlike the sensory task, decisions and confidence here were informed by prior knowledge of generally known quantities – national populations (Pleskac and Busemeyer, 2010). On each trial, subjects were shown the names of two countries, and indicated with a key press which one they believed to have a larger population (Figure 3A). After responding, subjects were prompted to enter decision confidence on a 5 division scale. To avoid any influence of learning, we omitted feedback about choice correctness. An important feature of this task was that choice accuracy varied with the log ratio of country population pairs, providing a quantitative measure of evidence discriminability (Figure 3A, inset). Again as predicted by the model, (i) accuracy monotonically increased as a function of subject confidence (Figure 3B,E); (ii) confidence increased with evidence discriminability on correct trials but decreased on errors, (iii) mid-range confidence characterized neutral-evidence trials (Figure 3C,F) and (iv) confidence provided information about trial outcome at fixed levels of evidence discriminability (Figure 3D,G). We found that qualitative patterns in general knowledge confidence were well captured by the parameter-free model (thick lines in Figure 3B-G), while adding a noise parameter only marginally improved the quality of the fit (thin lines, Fig 3).

Discussion

Here we compared self-reported confidence in humans with the properties of a statistical confidence computation. Our main result is that the relationship between human confidence, evidence discriminability and choice correctness reveals robust patterns, which can be quantitatively predicted by statistical decision confidence. Intriguingly, our statistical framework predicts that in a regime with uniformly varying discriminability eliciting performance levels from chance to near-perfect, the average confidence in chance-accuracy decisions driven by *uninformative* evidence is much higher than chance at 0.75. This property of statistical confidence was confirmed in our human data for two behavioral tasks, and carries implications for interpreting studies that demonstrate over-confidence in low discriminability and under-confidence in high discriminability conditions - a controversial phenomenon termed the "hard-easy effect" (Merkle, 2009). Previous interpretations of this apparent miscalibration had pointed to differences in experimental design and interpretation of data (Juslin et al., 2000; Moore and Healy, 2008) and more recently, to the effect of examining confidence reports with respect to evidence discriminability (Drugowitsch et al., 2014). By relating statistical confidence to human reports, we have shown that human mid-range confidence in completely uninformative evidence does not imply an imperfectly calibrated sense of confidence.

Our results identify an important link that had previously been absent from the literature on computational modeling of confidence. Different algorithms for determining confidence had been proposed as extensions to accumulator (Kepecs et al., 2008; Vickers, 1979), drift diffusion (Kiani et al., 2014; Pleskac and Busemeyer, 2010; Zylberberg et al., 2012), and attractor models (Insabato et al., 2010; Wei and Wang, 2015). The confidence metrics proposed ranged from the simple difference between decision

variables, to post-decision evidence sampling and reaction time combined with evidence. Each model has successfully accounted for numerous aspects of choice and confidence. In fact, when these models are tuned to make similar predictions, they may algorithmically approximate Bayes-optimal computations. However, with different parameter settings these algorithmic accounts can also make different predictions (Kiani et al., 2014; Pleskac and Busemeyer, 2010; Zylberberg et al., 2012) and the link between the proposed confidence metrics and the statistical definition of confidence has not been clear (but see (Drugowitsch et al., 2014). Here, we provide a framework for using signatures of decision confidence based on first principles to understand how these metrics relate to statistical confidence.

To establish how objective and subjective confidence are related, we sought to isolate how confident subjects felt as directly as possible by instructing subjects to select their confidence on a 1 to 5 scale with a dedicated motor response, instead of asking them to explicitly estimate a probability or to cast a wager (Persaud et al., 2007). This way we sought to sidestep complex calibration issues related to explicit estimation of a probability (Juslin et al., 2000; Lichtenstein et al., 1981) and risk-sensitivity (Fleming and Dolan, 2010). Our confidence report method contrasts with other assays where choice and confidence are reported with the same motor response (Bahrami et al., 2012; Kiani et al., 2014; Zylberberg et al., 2012). For instance, a recent study considered the relationship between confidence and discriminability, where the direction of a ballistic eye movement reported choice, terminating the decision evidence stream, and its magnitude indicated confidence (Kiani et al., 2014). In their results, the relationship between confidence and discriminability for error trials varied among subjects from negative slopes to neutral and in some cases, positive This variability may be due to subjects constructing their confidence reports while new decision evidence was being delivered, rather than while reflecting on a terminated stream of evidence available to the experimenter when computing the trial's discriminability.

Often people seem to provide systematically miscalibrated confidence reports, leading to the view that human confidence is generated by an error-prone mental heuristic (Björkman et al., 1993; Griffin and Tversky, 1992; Kvidera and Koutstaal, 2008; Olsson and Winman, 1996; Tversky and Kahneman, 1974). By using a psychophysical approach, we could examine confidence reports quantitatively as a function of the degree of evidence provided before the choice on each trial. This enabled us to show for two different tasks, that evidence processing is consistent with normative statistical principles and does not require error-prone heuristic computations (Gigerenzer and Goldstein, 1996; Koriat, 2012; Tversky and Kahneman, 1974). Although we established a quantitative match between human confidence and statistical decision confidence, we suspect that the statistical computation provides only an initial confidence estimate for human decision makers. In some circumstances, this internal estimate may be further modified by context (Jönsson et al., 2005), social factors (Bahrami et al., 2012) or other conditions, accounting for a range of reported assays where confidence can be divorced from choice accuracy (De Lange et al., 2011; Metcalfe and Shimamura, 1994). To account for these effects, the Bayesian confidence formula now provides an experimentally validated starting point from which to construct computational models of distortions in ideal human confidence.

In summary, the striking similarity of patterns in confidence feelings to the predictions of statistical decision confidence suggests that these two constructs can share functional equivalence for the decision maker. In decision making with imperfect evidence, the brain faces the same computational challenge as a statistician evaluating noisy data. The fact that this important value is experienced as a feeling supports

the idea that subjective feelings provide an interface to ethologically beneficial mental computations – in our case, a statistical estimation tool for judgments.

Experimental Procedures

Model simulations

We created a Monte Carlo simulation of the statistical definition of decision confidence (main text figure 1 b-d). For each of 1 billion simulation trials, we chose a uniform distribution for evidence discriminability, from the range -1 to 1. Next, we computed *percepts* by adding Gaussian noise to each trial's discriminability (μ=0, σ=0.18). To compute outcomes, we scored each trial "correct" if both the trial's discriminability and percept had the same sign, and "incorrect" if not. Next, we computed the probability of a correct response for each percept. We grouped percepts into 200 equally sized bins spanning the range of percepts, and computed confidence in each percept as the fraction of trials in each bin that were scored "correct". We then assigned a confidence to each simulation trial, by matching the trial's percept to the percept's corresponding confidence value. The confidence values produced by the model thus represent explicit probabilities, and range from 0.5 – 1.0.

Model fitting

The statistical model simulation has one free parameter modeling the decision maker: the variance of Gaussian noise, and two parameters of the evidence stream: the range and frequency of evidence discriminability. We sought to produce a parameter-free model fit, by determining each of these parameters from the subjects' trial and behavior data. Perceptual noise level σ was estimated as the variance of a cumulative Gaussian function fitted to each subject's psychometric function using maximum likelihood method. Next, we generated a ~10 million trial dataset starting with replicates of discriminability for all completed trials in each subject's dataset. We used the best-fit variance (Figure S1) as the model's noise parameter, and computed percepts, choice and confidence for each simulation trial. We partitioned the model's confidence into five categories based on the subject's frequency of scale division use.

We added a single free noise parameter to our model (α), which was implemented as a mixing parameter between the confidence value of the noise-free model and a randomly sampled confidence value from the same distribution:

$$c_n = (\alpha * c) + (1 - \alpha) * \eta \qquad (1)$$

Here c_n is confidence with added noise, c is the raw confidence value produced by the model, η is a confidence value randomly resampled from the 10 million trial simulation, and α is the noise parameter. Note that the noise parameter provides a convenient index of the degree to which confidence reporting matches statistical confidence: at 1, confidence is the value defined by the model and at 0, confidence is random for each trial. We fitted this one-parameter model to human data using a maximum likelihood procedure (see supplemental methods).

Measure for evidence discriminability of presented click train stimulus

We determined a post-hoc measure of discriminability based on the portion of the click train experienced before responding. This measure is Δ, the ratio of differential clicks to total clicks that occurred prior to choice reaction time (Equation 2) where nL is the number of left clicks experienced, and nR is the number of right clicks.

$$\Delta = \left|\frac{nL - nR}{nL + nR}\right| \quad (2)$$

Subjects

For the sensory task, we recruited five adult human subjects (2 male, 3 female). For the general knowledge task, we recruited 20 males and 7 females. Subjects ranged from ages 20 to 50. All subjects reported normal hearing and normal or corrected-to-normal vision. All experimental procedures were approved by the Cold Spring Harbor Laboratory institutional review board.

References

Audley, R.J. (1960). A stochastic model for individual choice behavior. Psychol Rev *67*, 1-15.
Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., and Frith, C. (2012). What failure in collective decision-making tells us about metacognition. Philosophical Transactions of the Royal Society B: Biological Sciences *367*, 1350-1365.
Bar-Tal, Y., Sarid, A., and Kishon-Rabin, L. (2001). A test of the overconfidence phenomenon using audio signals. The Journal of General Psychology *128*, 76-80.
Baranski, J.V., and Petrusic, W.M. (1994). The calibration and resolution of confidence in perceptual judgments. Attention, Perception, & Psychophysics *55*, 412-428.
Björkman, M., Juslin, P., and Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. Attention, Perception, & Psychophysics *54*, 75-81.
Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. Science *340*, 95-98.
Camerer, C., and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. American economic review, 306-318.
Charles, L., Van Opstal, F., Marti, S., and Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. NeuroImage *73*, 80-94.
De Lange, F.P., Van Gaal, S., Lamme, V.A., and Dehaene, S. (2011). How awareness changes the relative weights of evidence during human decision-making. PLoS biology *9*, e1001203.
De Martino, B., Fleming, S.M., Garrett, N., and Dolan, R.J. (2013). Confidence in value-based choice. Nature Neuroscience *16*, 105-110.
Drugowitsch, J., Moreno-Bote, R., and Pouget, A. (2014). Relation between Belief and Performance in Perceptual Decision Making. PloS one *9*, e96511.
Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature *415*, 429-433.
Fetsch, C.R., DeAngelis, G.C., and Angelaki, D.E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. Nature Reviews Neuroscience *14*, 429-442.
Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. Trends in cognitive sciences *14*, 119-130.
Flavell, J.H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American psychologist *34*, 906.
Fleming, S.M., and Dolan, R.J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. Consciousness and cognition *19*, 352-363.
Gigerenzer, G., and Goldstein, D.G. (1996). Reasoning the fast and frugal way: models of bounded rationality. Psychological Review *103*, 650.
Griffin, D., and Tversky, A. (1992). The weighing of evidence and the determinants of confidence. Cognitive psychology *24*, 411-435.
Insabato, A., Pannunzi, M., Rolls, E.T., and Deco, G. (2010). Confidence-related decision making. Journal of neurophysiology *104*, 539-547.
Jastrow, J., and Peirce, C. (1884). On small differences in sensation. Memoirs of the National Academy of Science *3*, 1884.
Jönsson, F.U., Olsson, H., and Olsson, M.J. (2005). Odor emotionality affects the confidence in odor naming. Chem Senses *30*, 29-35.
Juslin, P., Winman, A., and Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. Psychological Review *107*, 384.
Kahneman, D., and Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive psychology *3*, 430-454.
Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature *455*, 227-231.

Kiani, R., Corthell, L., and Shadlen, M.N. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. Neuron *84*, 1329-1342.
Kiani, R., and Shadlen, M.N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. Science *324*, 759-764.
Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., and Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. Nature neuroscience.
Körding, K.P., and Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. Nature *427*, 244-247.
Koriat, A. (2012). The self-consistency model of subjective confidence. Psychological Review *119*, 80.
Kunimoto, C., Miller, J., and Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. Consciousness and cognition *10*, 294-340.
Kvidera, S., and Koutstaal, W. (2008). Confidence and decision type under matched stimulus conditions: overconfidence in perceptual but not conceptual decisions. Journal of Behavioral Decision Making *21*, 253-281.
Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. Trends in cognitive sciences *15*, 365-373.
Lichtenstein, S., Fischhoff, B., and Phillips, L.D. (1981). Calibration of probabilities: The state of the art to 1980 (DTIC Document).
Merkle, E.C. (2009). The disutility of the hard-easy effect in choice confidence. Psychonomic bulletin & review *16*, 204-213.
Metcalfe, J.E., and Shimamura, A.P. (1994). Metacognition: Knowing about knowing (The MIT Press).
Moore, D.A., and Healy, P.J. (2008). The trouble with overconfidence. Psychological Review *115*, 502.
Olsson, H., and Winman, A. (1996). Underconfidence in sensory discrimination: The interaction between experimental setting and response strategies. Attention, Perception, & Psychophysics *58*, 374-382.
Persaud, N., McLeod, P., and Cowey, A. (2007). Post-decision wagering objectively measures awareness. Nat Neurosci *10*, 257-261.
Pleskac, T.J., and Busemeyer, J.R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychological Review *117*, 864.
Pouget, A., Beck, J.M., Ma, W.J., and Latham, P.E. (2013). Probabilistic brains: knowns and unknowns. Nature Neuroscience *16*, 1170-1178.
Sanders, J., and Kepecs, A. (2012). Choice Ball: a response interface for psychometric discrimination in head-fixed mice. Journal of neurophysiology.
Sanders, J.I., and Kepecs, A. (2014). A low-cost programmable pulse generator for physiology and behavior. Frontiers in neuroengineering *7*.
Schapire, R.E., and Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. Machine learning *37*, 297-336.
Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., and Frith, C.D. (2014). Supra-personal cognitive control and metacognition. Trends in cognitive sciences *18*, 186-193.
Sollich, P. (2002). Bayesian methods for support vector machines: Evidence and predictive class probabilities. Machine learning *46*, 21-52.
Stankov, L. (1998). Calibration curves, scatterplots and the distinction between general knowledge and perceptual tasks. Learning and Individual Differences *10*, 29-50.
Stocker, A.A., and Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. Nature Neuroscience *9*, 578-585.
Tenenbaum, J.B., Kemp, C., Griffiths, T.L., and Goodman, N.D. (2011). How to grow a mind: Statistics, structure, and abstraction. Science *331*, 1279-1285.
Trommershäuser, J., Maloney, L.T., and Landy, M.S. (2008). Decision making, movement planning and statistical decision theory. Trends in cognitive sciences *12*, 291-297.
Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science *185*, 1124-1131.
Vickers, D. (1979). Decision processes in visual perception (New York, London: Academic Press).

Wei, Z., and Wang, X.J. (2015). Confidence estimation as a stochastic process in a neurodynamical system of decision making. J Neurophysiol *114*, 99-113.
Yeung, N., and Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. Philosophical Transactions of the Royal Society B: Biological Sciences *367*, 1310-1321.
Zylberberg, A., Barttfeld, P., and Sigman, M. (2012). The construction of confidence in a perceptual decision. Frontiers in Integrative Neuroscience *6*.

Acknowledgments

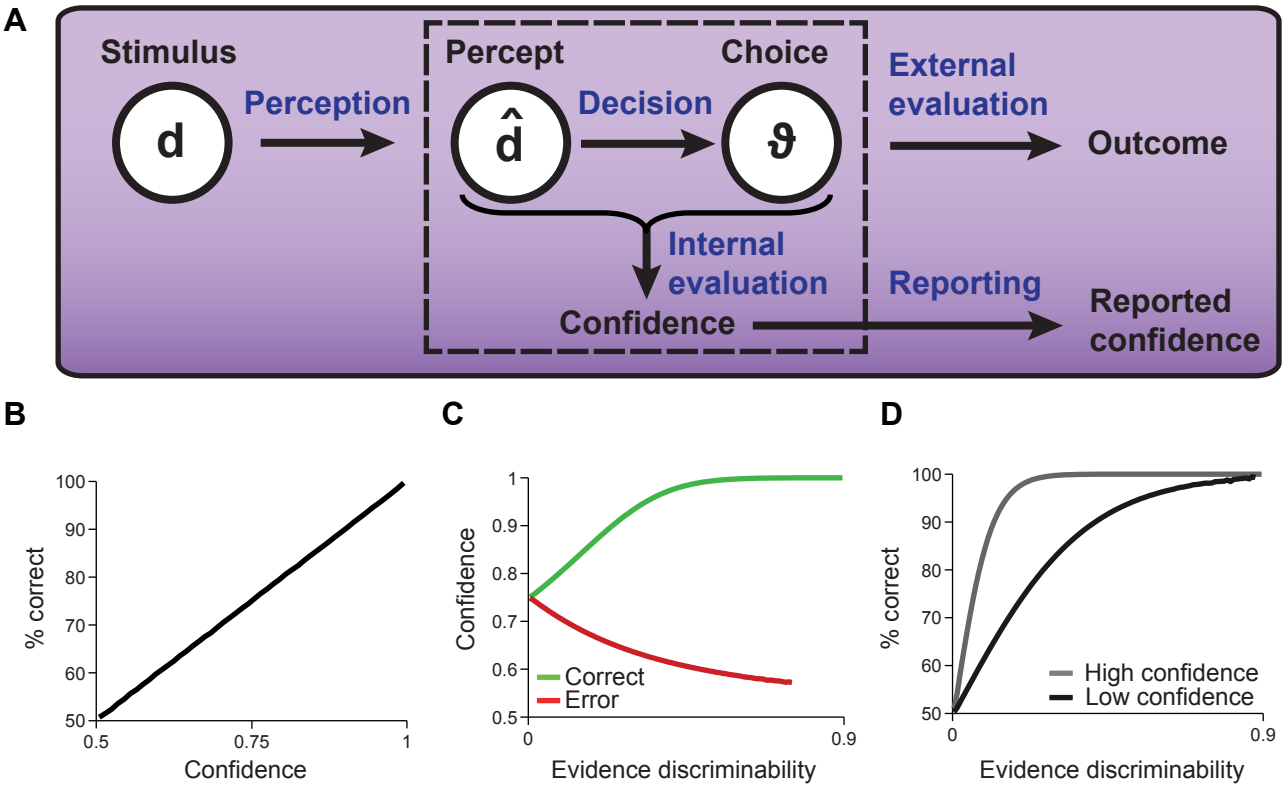
We are grateful to Drs. Steve Fleming, Alex Koulakov and Brett Mensh for insightful discussions and comments.

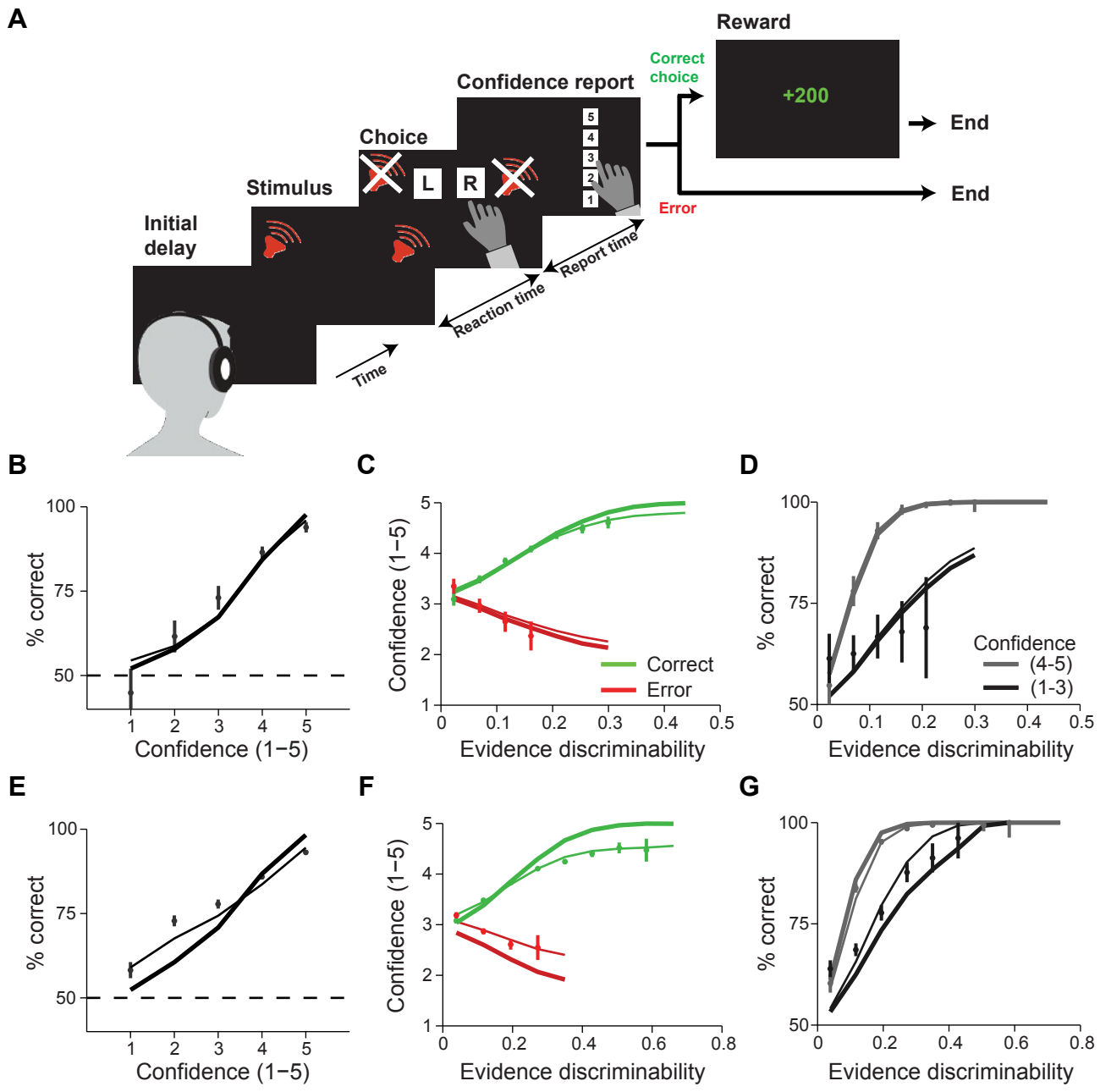
Figure Legends

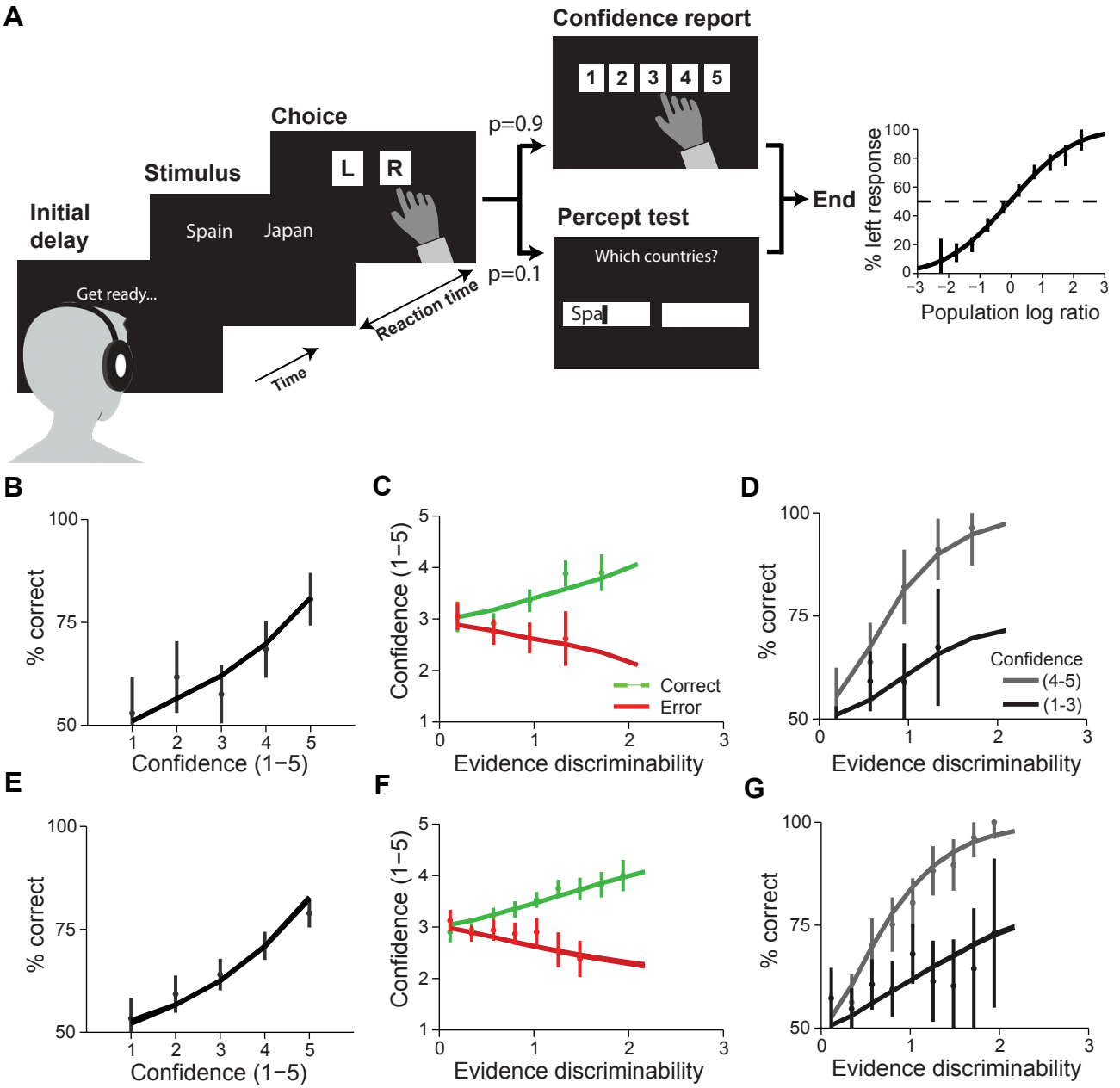
Figure 1. Statistical decision confidence predicts specific interrelationships between evidence discriminability, choice outcome and confidence. (A) Illustration of the statistical framework for decision confidence. The dashed box delineates variables internal to the decision maker. A presented stimulus d is corrupted in perception, producing percept $\hat{d}$, which informs the decision, ϑ. Confidence is computed based on the statistical definition. In humans, confidence is then explicitly mapped to a rating scale, producing the measured report. An external evaluation determines decision correctness. B-D: Monte Carlo simulation of the statistical definition of decision confidence. For all panels, evidence discriminability (see Methods) is the absolute distance of the stimulus from zero. (B) Confidence equals accuracy. (C) Average confidence increases with evidence discriminability from 0.75 for correct choices, and decreases for errors. (D) Conditioning psychometric performance on high or low confidence changes its slope.

Figure 2. The human feeling of confidence follows statistical predictions in a perceptual decision task. For all panels, evidence discriminability is the absolute difference to sum ratio of number of left and right clicks in the experienced click train |((L-R)/(L+R))|. (A) Schematic of task events. (B-D) Confidence patterns of a single subject. Thick lines show parameter-free normative statistical model simulations. Thin lines show one-parameter model fits with a confidence efficacy parameter. Each individual subject is shown in Figure S3. (E-G) Combined data of all subjects (n = 5). Error bars show 95% confidence interval of the mean.

Figure 3. The human feeling of confidence follows statistical predictions in on general knowledge decision task. For all panels, evidence discriminability is the log ratio of the national populations compared. (A) Schematic of the general knowledge task. After initiating each trial and following a random delay, subjects were shown the names of two countries and asked to indicate which had a larger population within 3 seconds by pressing a response key. On 90% of trials, subjects then entered their decision confidence. On sensory probe trials (10%), subjects typed the names of the countries they had just compared. Inset panel: general knowledge task psychometric function for 27 pooled subjects (3,450 trials) showing that choice varied as a function of population log ratio. Errors show binomial 95% confidence intervals. (B-D) Confidence patterns of a single subject who completed 1200 trials. Thick lines show the parameter-free model simulation. Thin lines show a single-parameter model fit with a confidence noise parameter (mostly obscured by the thick lines). (E-G) Combined data of 27 subjects, each completing 100-150 trials. Notably, subjects were only 78.6% correct for trials where confidence was 5/5, consistent with 82.3% accuracy on the strongest fifth of the range of presented evidence (panel E). Error bars show 95% confidence interval of the mean.


A
Stimulus
d
Perception
Percept
$\hat{d}$
Decision
Choice
ϑ
External evaluation
Outcome
Internal evaluation
Confidence
Reporting
Reported confidence
B
% correct
100
90
80
70
60
50
0.5
0.75
1
Confidence
C
Confidence
1
0.9
0.8
0.7
0.6
0.5
Correct
Error
0
0.9
Evidence discriminability
D
% correct
100
90
80
70
60
50
High confidence
Low confidence
0
0.9
Evidence discriminability


A
Reward
Correct choice
+200
End
Confidence report
Choice
Stimulus
Initial delay
L
R
5
4
3
2
1
Error
End
Reaction time
Report time
Time
B
% correct
100
75
50
1
2
3
4
5
Confidence (1−5)
C
Confidence (1−5)
Correct
Error
0
0.1
0.2
0.3
0.4
0.5
Evidence discriminability
D
% correct
Confidence
(4-5)
(1-3)
Evidence discriminability
E
% correct
Confidence (1−5)
F
Confidence (1−5)
0
0.2
0.4
0.6
0.8
Evidence discriminability
G
% correct
Evidence discriminability


A
Confidence report
1 2 3 4 5
p=0.9
Choice
L R
Stimulus
Spain Japan
Initial delay
Get ready...
Reaction time
Time
p=0.1
Percept test
Which countries?
Spa
End
% left response
Population log ratio
B
% correct
Confidence (1−5)
C
Confidence (1−5)
Correct
Error
Evidence discriminability
D
% correct
Confidence
(4-5)
(1-3)
Evidence discriminability
E
% correct
Confidence (1−5)
F
Confidence (1−5)
Evidence discriminability
G
% correct
Evidence discriminability