Models of confidence to facilitate engaging task designs

Vanessa Ceja* (vanesssc@uci.edu)

Yussuf Ezzeldine* (yezzeldi@uci.edu)

Megan A. K. Peters (megan.peters@uci.edu)
Department of Cognitive Sciences, University of California Irvine
Irvine, California 92697 USA

Abstract

Decision confidence models classically depict decision-making circuitry as: 1) accumulating relative evidence for each choice alternative and 2) computing confidence estimates from the difference in evidence magnitude favoring each choice. Recently, however, new evidence suggests a dissociation between metacognitive (confidence) computations and those supporting low-level perceptual decisions, positing instead that confidence is predominantly influenced by evidence favoring the selected choice while simultaneously ignoring evidence for the non-selected choice. Low-level perceptual tasks completed by neurotypical subjects and/or within controlled experiments, coupled with computational modeling, have helped reveal the computations and brain areas involved, but we do not yet know to what degree these dissociations generalize to other types of perceptual or cognitive tasks or to clinical, developmental, or aging populations. Here, we begin to tackle this issue by proposing a task and computational modeling comparison framework aimed at understanding whether perceptual confidence computations are stable across varying levels of perceptual judgements, in service of creating more engaging tasks for use in wider and more diverse populations.

Keywords: metacognition; perceptual decision-making; Bayesian computational modeling; facial attractiveness

Introduction

Traditional models of decision confidence posit that decision-making circuitry is also responsible for encoding certainty about decisions, and confidence is the evidential difference in favor of each choice (Kiani & Shadlen, 2009; Moreno-Bote, 2010; Vickers, 1979). However, recent work suggests that confidence appears primarily influenced by evidence for the selected choice but seems to ignore evidence for the non-selected choice (Maniscalco, Peters, & Lau, 2016; Peters et al., 2017; Zylberberg, Barttfeld, & Sigman, 2012): a 'bias' towards decision-congruent evidence magnitude. This theory has been recently supported by electrocorticography and modeling work (Maniscalco et al., 2021; Peters et al., 2017).

Currently, we know that: 1) low-level perceptual tasks can reveal this dissociation between confidence and decision-making capacity, 2) various models exist which hypothesize possible computations taking place in responsible brain areas, and 3) much of this work has been done in neurotypical subjects and/or within highly controlled experimental settings.

Thus, many open questions remain, namely regarding the generality of our current knowledge. For example, do these dissociations hold in children, normal aging processes, or clinical populations? It is known that gray matter volume (Fleming, Weil, Nagy, Dolan, & Rees, 2010; Fleming, Huijgen, & Dolan, 2012; McCurdy et al., 2013) and activity (Morales, Lau, & Fleming, 2018) in the prefrontal cortex correlates with visual metacognition capacity or differences between memory and perception metacognition, but we also know that this region changes significantly across the lifespan and in disease and neurodegeneration (Fereshtehnejad et al., 2019; Yang et al., 2012; Ramanoël et al., 2018; Strikwerda-Brown, Ramanan, & Irish, 2019). We also do not know to what degree these dissociations between metacognitive and first-order processes differ between low-level perceptual decisions (e.g. dot-motion direction) and higher level perceptual decisions (e.g. attractiveness). Attractiveness is a particularly interesting target, as it has been shown to be susceptible to a 'wingman' effect wherein the attractiveness of other nearby faces can affect the perceived attractiveness of a target face (Furl, 2016); this 'divisive-normalization' pattern suggests that attractiveness may also exhibit the same performance-confidence dissociations that have been similarly explained using dot motion or Gabor patches, given the potential connection to tuned inhibition (Maniscalco et al., 2021).

To validate this possibility and expand into other perceptual or cognitive domains, we therefore must develop tasks that ask about the stability of observed metacognitive computations across stimulus types and tasks (dot motion, Gabor patches, facial attractiveness, etc.), and in doing so open doors for studying how metacognitive computations change across "non-college-student" populations through more engaging task designs. Therefore, we are developing a task and computational modeling comparison framework that will allow us to determine whether perceptual confidence computations are stable across perceptual judgments of various varieties (dots, stripes, facial attractiveness, etc.). Here we present the first of these tasks—facial attractiveness rating—and the family of models in development, and discuss how they will be used to ask whether the same computations can explain confidence in perceptual decisions across multiple types of judgments and populations.

Methods

Behavioral Methods

42 human participants were recruited via the University of California, Irvine online subject pool system and gave informed consent to participate in this online experiment.

Our behavioral task closely resembles (Furl, 2016)'s (Fig. 1). During phase 1, participants rated the attractiveness of the person shown on the screen on a scale from extremely attractive to extremely unattractive. Participants rated 30 faces, 15 male and 15 female, three times each (a departure from (Furl, 2016), who asked participants to only rate attractiveness twice; we made this change for stability of ratings). The average attractiveness from the 2nd and 3rd rating of each face were averaged and used as estimates of each participant's attractiveness judgment for that face for phase 2. The ten most consistently-rated images were then ranked based on these average ratings, with 10 being the 'most attractive' face and 1 being the 'least attractive' face. These 10 ranked faces were then presented three at a time to the participant in pseudo-random configuration, and the participant was asked to select which face was the most attractive and then rate confidence.

We pre-selected the combination of attractiveness ranks

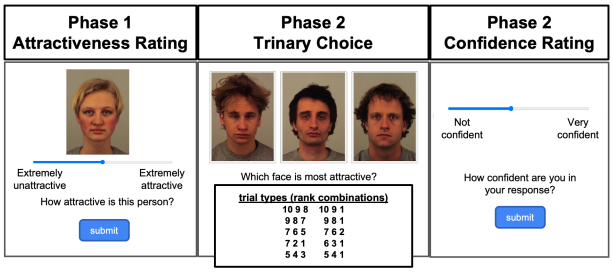

Figure 1: Behavioral task.

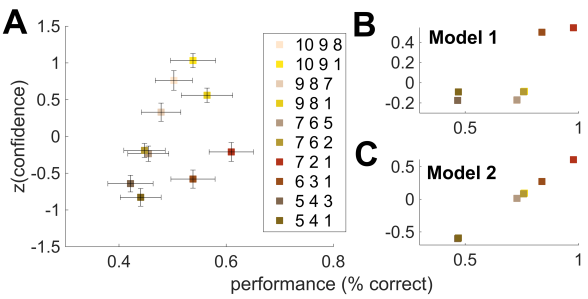

Figure 2: (A) Behavioral results and (B,C) two sample models. Neither model captures the behavioral data, even qualitatively.

(Fig. 1, 'trial types'), allowing us to manipulate the perceptual difficulty of every trial by creating easier or harder trial type combinations based on the closeness of ranks—e.g., a 'hard' trial was one in which the most attractive and second most attractive face were very close in attractiveness rank—as well as to independently manipulate the overall level of attractiveness of the most attractive (correct choice) face. This orthogonal manipulation of difficulty and overall attractiveness is critical to testing the models; see below.

Models

We developed 7 models, all of which are variants on the same 'Bayesian' theme. All models make type 1 ("which face is most attractive?") ratings based on the *maximimum a posteriori estimate*, $p(F|x) = \frac{p(x|F)p(F)}{p(x)}$, where $p(x) = \sum_i^3 p(x|F=i)(p(F=i))$ with all $p(F) = \frac{1}{3}$. The models differ in how they make confidence judgments. With F_c defined as the face the observer chose as the most attractive face, each defines confidence C as:

Model 1: $C \propto \frac{p(x|F_c)p(F_c)}{\sum_i^3 p(x|F_i)(p(F_i))}$

Model 2: $C \propto \frac{p(x|F_c)p(F_c)}{\sum_i^3 p(x|F_i)(p(F_i))+p(x|N)p(N)}$, where N is the 'pure noise' distribution (least attractive face possible) centered at [0,0,0] and $p(N) = 0.01$, while the a priori probabilities of the remaining 3 options are set to $\frac{1-0.01}{3}$

Model 3: $C \propto \frac{p(x|F_c)p(F_c)-p(x|F_2)p(F_2)}{\sum_i^3 p(x|F_i)(p(F_i))}$, where F_2 is the face the observer believed was the middle attractive face

Model 4: $C \propto \frac{p(x|F_c)p(F_c)-p(x|F_2)p(F_2)}{\sum_i^3 p(x|F_i)(p(F_i))+p(x|N)p(N)}$

Model 5: $C \propto p(x|F_c)p(F_c) - p(x|F_2)p(F_2)$

Model 6: $C \propto \frac{p(x|F_c)p(F_c)-p(x|F_2)p(F_2)}{p(x|F_3)p(F_3)}$, where F_3 is the face the observer believed was the least attractive face

Model 7: $C \propto p(x|F_c)$

We simulated choices and confidence for all models using Monte Carlo simulations in Matlab (1e5 trials per condition). We assume all generating distributions are trivariate Gaussian with covariance matrix $\Sigma = I$, the 3d identity matrix. Means to generate x in each condition were set according to trial types (Fig. 1). Likelihoods $p(x|F)$ were calculated against a 'perfect 10', i.e. comparing x to a canonical case in which $F_1 = [10,0,0]$, $F_1 = [0,10,0]$, and $F_3 = [0,0,10]$.

Results

We observed that participants' capacity to select the face they previously rated as most attractive depended on the relative attractiveness of all three faces (Fig. 2A), as expected. Moreover, task performance followed results reported by (Furl, 2016): the most unattractive face of the triad further modulated performance (e.g. [10 9 8] led to lower % correct than [10 9 1]). However, confidence did not always monotonically follow task performance: Despite the highest performance occurring for [7 2 1], this condition was far from the highest confidence, which was occupied by [10 9 1]. This shows a clear 'decision-congruent evidence' bias as has been reported previously using lower level perceptual stimuli (Koizumi, Maniscalco, & Lau, 2015; Maniscalco et al., 2016; Odegaard et al., 2018; Samaha, Barrett, Sheldon, LaRocque, & Postle, 2016).

We also present a sample of the models (Fig. 2A & B, Models 1 and 2) to demonstrate that neither can even qualitatively capture the behavioral patterns shown by the human participants. Ongoing work is developing a full factorial family of models, and will use hierarchical fitting and formal model comparison practices to ask whether the same model 'wins' in explaining all tasks for a given subject.

Discussion & future directions

Here, we have developed a behavioral task that successfully replicates findings from lower level perceptual tasks, i.e. that decision-congruent evidence magnitude appears to overly drive confidence judgments (Maniscalco et al., 2016). Ongoing work pairs this behavioral task with 'lower-level' perceptual tasks and quantitative model comparisons to ask whether *exactly* the same computations drive decision-confidence dissociations regardless of whether the task uses random dot kinematograms, Gabor patches, or facial attractiveness. This suite of tools holds exceptional promise for characterizing metacognitive computations across developmental and normal aging trajectories and in clinical populations who are less tolerant of many hours of repetitive psychophysics tasks. As we expand our library of behavioral tasks, our goal is to have a single participant complete a whole series spanning lower to higher level perception and to use within-subject comparisons to provide novel and exciting, generalizable insights into the nature of confidence computations in perceptual decisions.

Acknowledgments

This project was supported in part by the Canadian Institute for Advanced Research Azrieli Global Scholars Fellowship in Brain, Mind, & Consciousness (to MAKP).

References

Fereshtehnejad, S.-M., Yao, C., Pelletier, A., Montplaisir, J. Y., Gagnon, J.-F., & Postuma, R. B. (2019, July). Evolution of prodromal parkinson's disease and dementia with lewy bodies: a prospective study. *Brain*, *142*(7), 2051–2067.

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012, May). Prefrontal contributions to metacognition in perceptual decision making. *J. Neurosci.*, *32*(18), 6117–6125.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010, September). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543.

Furl, N. (2016, October). Facial-Attractiveness choices are predicted by divisive normalization. *Psychol. Sci.*, *27*(10), 1379–1387.

Kiani, R., & Shadlen, M. N. (2009, May). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764.

Koizumi, A., Maniscalco, B., & Lau, H. (2015, May). Does perceptual confidence facilitate cognitive control? *Atten. Percept. Psychophys.*, *77*(4), 1295–1306.

Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S. H., Basso, M. A., Lau, H., & Peters, M. A. K. (2021, March). Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. *PLoS Comput. Biol.*, *17*(3), e1008779.

Maniscalco, B., Peters, M. A. K., & Lau, H. (2016, April). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Atten. Percept. Psychophys.*, *78*(3), 923–937.

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013, January). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *J. Neurosci.*, *33*(5), 1897–1906.

Morales, J., Lau, H., & Fleming, S. M. (2018, April). Domain-General and Domain-Specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.*, *38*(14), 3534–3546.

Moreno-Bote, R. (2010, July). Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators. *Neural Comput.*, *22*(7), 1786–1811.

Odegaard, B., Grimaldi, P., Cho, S. H., Peters, M. A. K., Lau, H., & Basso, M. A. (2018, February). Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proc. Natl. Acad. Sci. U. S. A.*, *115*(7), E1588–E1597.

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., ... Lau, H. (2017, July). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat Hum Behav*, *1*.

Ramanoël, S., Hoyau, E., Kauffmann, L., Renard, F., Pichat, C., Boudiaf, N., ... Baciu, M. (2018, August). Gray matter volume and cognitive performance during normal aging. a Voxel-Based morphometry study. *Front. Aging Neurosci.*, *10*, 235.

Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016, June). Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Front. Psychol.*, *7*, 851.

Strikwerda-Brown, C., Ramanan, S., & Irish, M. (2019, February). Neurocognitive mechanisms of theory of mind impairment in neurodegeneration: a transdiagnostic approach. *Neuropsychiatr. Dis. Treat.*, *15*, 557–573.

Vickers, D. (1979, December). Uncertainty, choice, and the marginal efficiencies. *J. Post Keynes. Econ.*, *2*(2), 240–254.

Yang, J., Pan, P., Song, W., Huang, R., Li, J., Chen, K., ... Shang, H. (2012, May). Voxelwise meta-analysis of gray matter anomalies in alzheimer's disease and mild cognitive impairment using anatomic likelihood estimation. *J. Neurol. Sci.*, *316*(1-2), 21–29.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012, September). The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.*, *6*, 79.