Journal of Memory and Language 102 (2018) 142–154

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Comparing the confidence calculation rules for forced-choice recognition memory: A winner-takes-all rule wins

Kiyofumi Miyoshi[a,b,*], Ayumi Kuwahara[c], Jun Kawaguchi[a]

[a] Graduate School of Informatics, Nagoya University, Japan
[b] Japan Society for the Promotion of Science, Japan
[c] School of Informatics and Sciences, Nagoya University, Japan

ARTICLE INFO

Keywords:
Recognition memory
Decision making
Confidence
Metacognition
Signal-detection theory

ABSTRACT

Using a new signal-detection-theory-based approach, Experiments 1 and 2 of this study were found to reveal that the internal confidence for two-alternative forced-choice (2-AFC) recognition memory is calculated in a winner-takes-all manner. The signal strength for one of a pair of stimuli exclusively determines confidence and the other piece of useful information is discarded. Similar winner-takes-all confidence calculation has been reported in different kinds of visual perception tasks, which, together with the present findings, confirms that it is a domain-general metacognitive heuristic. Furthermore, by using a less common 2-AFC recognition memory task to select a new stimulus, Experiment 3 elucidated the underlying mechanism of this confidence calculation heuristic. Previously, it had been assumed that the chosen stimulus's signal strength selectively determines confidence and that the signal strength for the unchosen stimulus is disregarded (heuristic use of response-congruent evidence or post-decisional confirmation bias). However, the present study demonstrates that the strength of a stronger signal selectively determines confidence and that of a weaker signal is disregarded. That is, the winner in signal competition, not response selection, takes all in the construction of confidence.

Computational rules of metacognition

Humans and other animals make decisions based on ambiguous sensory and/or mnemonic information. Given this fundamental uncertainty, a metacognitive ability to construct internal confidence for one's own decision is critical for guiding adaptive behavior. For example, one can take steps to reconsider choices when low confidence is perceived. Furthermore, confidence plays a key role in collective decision making as those opinions expressed with higher confidence tend to carry more weight (e.g., Sniezek & Henry, 1989; Mahmoodi, Bang, Ahmadabadi, & Bahrami, 2013). Considering its importance in various decision-making situations, it is desirable for our internal confidence to be a statistically optimal quantity that precisely reflects the probability that the decision made is correct. Although internal confidence shows such optimality in some experimental situations (e.g., Sanders, Hangya, & Kepecs, 2016), a good number of studies have reported that it miscalibrates the probability of a decision being correct (e.g., Bar-Tal, Sarid, & Kishon-Rabin, 2001; Baranski & Petrusic, 1994; Griffin & Tversky, 1992; Olsson & Winman, 1996).

Interestingly, psychophysical model-based evidence on how internal sensory signals are used to construct internal confidence in two-alternative forced-choice (2-AFC) visual perception tasks has emerged (Aitchison, Bang, Bahrami, & Latham, 2015; Maniscalco, Peters, & Law, 2016; Samaha, Iemi, & Postle, 2017; Zylberberg, Barttfeld, & Sigman, 2012). Beyond the investigation of metacognitive accuracy mentioned above, these model-based approaches enable inferences on internal computational rules underlying the optimality/suboptimality of metacognition. Classical psychophysical models posit that the difference in two signals' strength determines confidence (e.g., King & Dehaene, 2014). This "difference rule" leads to optimal metacognitive accuracy; confidence precisely tracks the probability of a decision being correct (Fig. 2 represents a signal-detection theory [SDT] model with this confidence calculation rule). However, intriguingly, Zylberberg et al. (2012) reported that confidence in a 2-AFC random-dot motion-direction discrimination task and that in a 2-AFC luminance discrimination task were selectively determined by the selected option's signal, but unaffected by the unselected option's signal (this is a suboptimal heuristic, which leads to miscalibration of the probability of being correct). Subsequent studies have demonstrated this "winner-takes-all" confidence calculation in different kinds of visual perceptual tasks, grating detection tasks (Maniscalco et al., 2016), grating orientation discrimination tasks (Samaha et al., 2017), and contrast discrimination

* Corresponding author at: Graduate School of Informatics, Nagoya University, 4648601, Chikusa, Nagoya, Japan.
E-mail addresses: miyoshi80@gmail.com (K. Miyoshi), kuwahara.ayumi@e.mbox.nagoya-u.ac.jp (A. Kuwahara), kawaguchijun@nagoya-u.jp (J. Kawaguchi).

https://doi.org/10.1016/j.jml.2018.06.001
Received 8 December 2017; Received in revised form 12 June 2018
Available online 26 June 2018

tasks (in only one experimental condition; Aitchison et al., 2015). These consistent observations of the winner-takes-all confidence calculation may corroborate its feasibility as a general metacognitive heuristic for visual perception, supported by some universal mechanism (Fig. 3 represents a SDT model with this confidence calculation rule).

Domain-generality of confidence calculation rules

The previous section focused exclusively on metacognition in visual perception tasks; however, the perceived probability of being correct can be considered a dimensionless quantity that is universal across different cognitive domains (e.g., vision, audition, and memory; Navajas et al., 2017; Song et al., 2011). The possible use of confidence as a domain-general scale for the probability of being correct could facilitate the combination of information across different dimensions, optimizing choice behavior (de Gardelle, Le Corre, & Mamassian, 2016). For more information on the roles of confidence in guiding adaptive choice behavior, see Bahrami et al. (2010) and Meyniel, Schlunegger, and Dehaene (2015).

Previous studies have tried to elucidate the domain-generality of metacognition by investigating the consistency of participants' metacognitive accuracies in visual perception and recognition memory tasks. Most of these studies focused mainly on individual differences in cross-domain metacognitive accuracy and their underlying neural mechanisms (Baird, Cieslak, Smallwood, Grafton, & Schooler, 2015; Baird, Smallwood, Gorgolewski, & Margulies, 2013; McCurdy et al., 2013; Fitzgerald, Arvaneh, & Dockree, 2017). However, to the best of our knowledge, no attempt has been made to contrast the difference rule and the winner-takes-all rule in 2-AFC recognition memory tasks, leaving the domain-generality of metacognitive computation an open question (regarding confidence calculation for Yes/No recognition memory, see Klauer & Kellen, 2015; Ratcliff & Starns, 2009, 2013; Rutishauser et al., 2015; Wixted, 2007). Supposedly, technical difficulties have precluded the investigation of these confidence calculation rules in 2-AFC recognition memory tasks. Previous studies on confidence calculation rules in 2-AFC visual perception tasks are grounded in SDT models or sequential sampling models to make psychophysical model-based inferences (Aitchison et al., 2015; Maniscalco et al., 2016; Samaha et al., 2017; Zylberberg et al., 2012). Most of these studies used quantitative model fittings based on many experimental trials in which sensory input strength was parametrically manipulated. However, in recognition memory tasks, stimulus strength (and memory strength derived from it) cannot be parametrically manipulated, and the number of trials are often restricted by the limited number of available stimuli, making it difficult to make robust model-based quantitative inferences on metacognitive rules.

To overcome these difficulties, we employed a SDT-based qualitative prediction approach (for related approaches, see Kellen & Klauer, 2014; King & Dehaene, 2014) in a basic 2-AFC recognition memory task to select an old (studied) stimulus. Just manipulating memory strength between two experimental conditions allows winner-takes-all and difference rules to provide mutually exclusive predictions of qualitative patterns of empirical data, without the need for quantitative model fittings. In short, as a non-optimal heuristic, the winner-takes-all rule predicts an unreasonable metacognitive pattern that the mean confidence in false alarm (FA) trials becomes higher when memory strength for target stimuli is strong than when memory strength for target stimuli is weak (i.e., participants more convincingly believe their wrong decision to be correct when the task is easier). This pattern can be intuitively understood if it is imagined that winners (lures) who have beaten strong opponents (targets) have higher absolute strength than winners who have beaten only weak opponents. On the contrary, the difference rule predicts a more reasonable metacognitive pattern that the mean confidence in FA trials becomes higher when memory strength for target stimuli is weak than when memory strength for target stimuli is strong (i.e., participants more convincingly believe their wrong decision to be correct when the task is more difficult) (for details, see the Model predictions section below). Experiment 1, which is based on this qualitative prediction approach provides psychophysical model-based evidence of confidence calculation rules for 2-AFC recognition memory.

Origin of winner-takes-all confidence calculation

Though the winner-takes-all confidence calculation is adopted in various visual perception tasks, its underlying mechanisms are still a matter of speculation. Zylberberg et al. (2012) inferred that the possible cause of this heuristic is the sensory metacognitive system's limited signal readout capacity. It is assumed that information used for confidence calculation is retrospectively retrieved in a post-decisional period (between forced-choice and confidence rating), and top-down processes selectively weight the signal for the selected option. That is, a binary choice is assumed to define the subset of sensory information used in a capacity-limited metacognitive system. More recent studies do not refer to this post-decisional retrieval process; however, they fundamentally adhere to the idea that information for the "selected option" is selectively used to construct confidence (Aitchison et al., 2015; Maniscalco et al., 2016; Samaha et al., 2017). Furthermore, Navajas, Bahrami, and Latham (2016) discussed this selective use of choice-congruent information in the context of post-decisional confirmation bias for the selected option (see the General discussion section for more information).

The term "winner" refers to the response selection winner, and the choice behavior itself is assumed to play an important role in this suboptimal confidence calculation heuristic. However, in previous studies (Aitchison et al., 2015; Maniscalco et al., 2016; Samaha et al., 2017; Zylberberg et al., 2012), the response selection winner was also the signal competition winner. That is, the stimulus generating a stronger sensory signal was always chosen. Therefore, the results in these studies can also be explained by a winner-takes-all heuristic stating that a stronger signal's strength exclusively determines confidence. The response-selection-induced post-decisional metacognitive bias sounds fascinating, but, without empirical support, such an elaborate process might be untenable in light of the principle of parsimony.

Given this existing knowledge, we designed additional experiments to further elucidate 2-AFC recognition memory confidence calculation mechanisms. In Experiment 2, we provided participants trial-by-trial correct/incorrect performance feedback designed to encourage optimal metacognitive behavior (Maniscalco et al., 2016). The performance feedback could remind participants of their undesirable high confidence for a wrong choice. If the use of the winner-takes-all rule derives from the metacognitive system's intrinsic incapacity, participants cannot discontinue using this suboptimal rule even when performance feedback is given. Otherwise, if the winner-takes-all rule is an optional metacognitive strategy to reduce the cognitive load, participants could switch from this suboptimal rule to the optimal difference rule with performance feedback. Furthermore, in Experiment 3 we presented participants with one old and one new (unstudied) stimulus and instructed them to select the new stimulus. In this situation, the signal competition winner runs counter to the response selection winner. That is, participants choose a stimulus with a weaker memory signal. Thus, the winner-in-signal-competition-takes-all rule, the winner-in-response-selection-takes-all rule, and the difference rule provide qualitatively exclusive predictions regarding experimental results, enabling a closer investigation of confidence calculation mechanisms.