Common computations in automatic cue combination and metacognitive confidence reports

Yi Gao[1], Kai Xue[1], Brian Odegaard[2], Dobromir Rahnev[1]

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

[2] Department of Psychology, University of Florida, Gainesville, FL

Keywords: cue combination; multisensory perception; confidence; visual metacognition; perceptual decision making

Author contributions: Yi Gao and Dobromir Rahnev conceived and designed the experiment; Yi Gao and Brian Odegaard programmed the experiment; Yi Gao conducted the experiment; Yi Gao, Kai Xue, and Dobromir Rahnev performed model fitting; Yi Gao and Dobromir Rahnev wrote the manuscript.

Acknowledgments: We thank Minzhi Wang for his help with data collection. This work was supported by the National Institute of Health (award: R01MH119189) and the Office of Naval Research (award: N00014-20-1-2622).

Corresponding author

Yi Gao; yi.gao0525@outlook.com

Abstract

Appropriate perceptual decision making necessitates the accurate estimation and use of sensory uncertainty. Such estimation has been studied in the context of both low-level multisensory cue combination and metacognitive estimation of confidence, but it remains unclear whether the same computations underlie both sets of uncertainty estimation. We created visual stimuli with low vs. high overall motion energy, such that the high-energy stimuli led to higher confidence but lower accuracy in a visual-only task. Importantly, we tested the impact of the low- and high-energy visual stimuli on auditory motion perception in a separate task. Despite being irrelevant to the auditory task, both visual stimuli impacted auditory judgments presumably via automatic low-level mechanisms. Critically, we found that the high-energy visual stimuli influenced the auditory judgments more strongly than the low-energy visual stimuli. This effect was in line with the confidence but contrary to the accuracy differences between the high- and low-energy stimuli in the visual-only task. These effects were captured by a simple computational model that assumes common computational principles underlying both confidence reports and multisensory cue combination. Our results reveal a deep link between automatic sensory processing and metacognitive confidence reports, and suggest that vastly different stages of perceptual decision making rely on common computational principles.

Word count: 205

Introduction

Sensory stimuli come with different levels of uncertainty (Ma & Jazayeri, 2014). For example, we can more precisely estimate the distance from the car ahead while driving in sunny conditions compared to heavy fog. Similarly, we can hear what another person is saying better in a quiet room compared to at a party. To make appropriate perceptual decisions, the brain must take such sensory uncertainty into account (Juslin & Olsson, 1997; Pouget et al., 2016; Rahnev & Denison, 2018; van Bergen et al., 2015). How people judge sensory uncertainty has been extensively investigated in two vastly different contexts.

First, a substantial literature has examined how information from different sensory cues is automatically combined within sensory areas to form a single perceptual judgment(Alais & Burr, 2004; Ernst & Banks, 2002; Fetsch et al., 2012; Kim et al., 2012; Landy et al., 2011; Morgan et al., 2008) This literature has shown that different cues are typically weighed by their uncertainty, thus enabling perceptual decisions that are close to optimal. Nevertheless, numerous studies have reported suboptimalities in cue combination (Arnold et al., 2019; Bruns, 2019; Fetsch et al., 2012; Knill & Saunders, 2003; Maiworm & Röder, 2011; Plaisier et al., 2014; Rosas et al., 2005), suggesting that uncertainty is not always estimated or used correctly.

Second, a separate literature has examined people's ability to metacognitively estimate how sensory uncertainty impacts accuracy via confidence ratings (Fleming et al., 2012; Mamassian, 2016; Rahnev, 2021a; Yeung & Summerfield, 2012). This literature has

demonstrated that confidence ratings typically reflect the uncertainty in the stimulus (Geurts et al., 2022; Sanders et al., 2016). However, many studies have also found that confidence computations can be noisy (Maniscalco & Lau, 2012; Mueller & Weidemann, 2008; Shekhar & Rahnev, 2021a) and that confidence and accuracy can be dissociated (Odegaard et al., 2018; Rahnev, 2021a; Samaha et al., 2016; Vlassova et al., 2014; Zylberberg et al., 2012).

Despite the fact that cue combination and decision confidence both involve the estimation and use of sensory uncertainty, it remains unknown whether the same computations underlie both processes. Indeed, as the brief overview above shows, both cue combination and decision confidence typically reflect the stimulus uncertainty but are often suboptimal. Therefore, a critical open question is whether the same sources of inefficiency corrupt both types of judgments.

Here, we investigated this question by examining whether stimuli that induce higher confidence but lower accuracy are given higher weight in multisensory cue combination. We employed an established method (Odegaard et al., 2018) to induce a confidence-accuracy dissociation by manipulating the energy levels in a random-dot kinematogram. We found that the high-energy visual stimuli led to lower accuracy but higher confidence than low-energy visual stimuli in a visual-only task. Critically, we also examined the automatic influence of high- and low-energy visual stimuli on auditory motion perception. Consistent with their associated higher confidence, the high-energy stimuli had a larger impact on

auditory perception. A simple computational model that assumes unified computations underlying both types of judgments reproduced this pattern of results. These findings suggest that sensory cue combination and metacognitive confidence judgments involve common computations.

Results

We tested whether the same or different uncertainty computations underlie automatic multisensory cue combination and metacognitive confidence judgments. We employed three different conditions: visual-only, auditory-only, and multisensory. In the visual-only condition, participants judged the direction of motion (left vs. right) of random-dot kinematograms (Figure 1A). We interleaved two different types of visual stimuli, low- and high-energy stimuli, designed to produce confidence-accuracy dissociations. In the auditory-only condition, participants experienced auditory stimulation of varying loudness separately in each ear designed to give the impression of leftward or rightward movement and judged the direction of motion (Figure 1B). Finally, in the multisensory condition, the visual and auditory stimuli were presented simultaneously (the two were congruent 50% of the time), and participants had to indicate the direction of motion of the auditory stimuli (Figure 1C). Nevertheless, based on previous research (Alink et al., 2012; Kitagawa & Ichihara, 2002; Soto-Faraco et al., 2002, 2005), we expected that visual stimuli would affect auditory judgments via automatic cue combination mechanisms. In the multisensory condition, the auditory stimuli were randomly paired with either the low- or the high-energy visual stimuli and we tested whether the weight given to the visual stimuli would reflect their corresponding confidence.

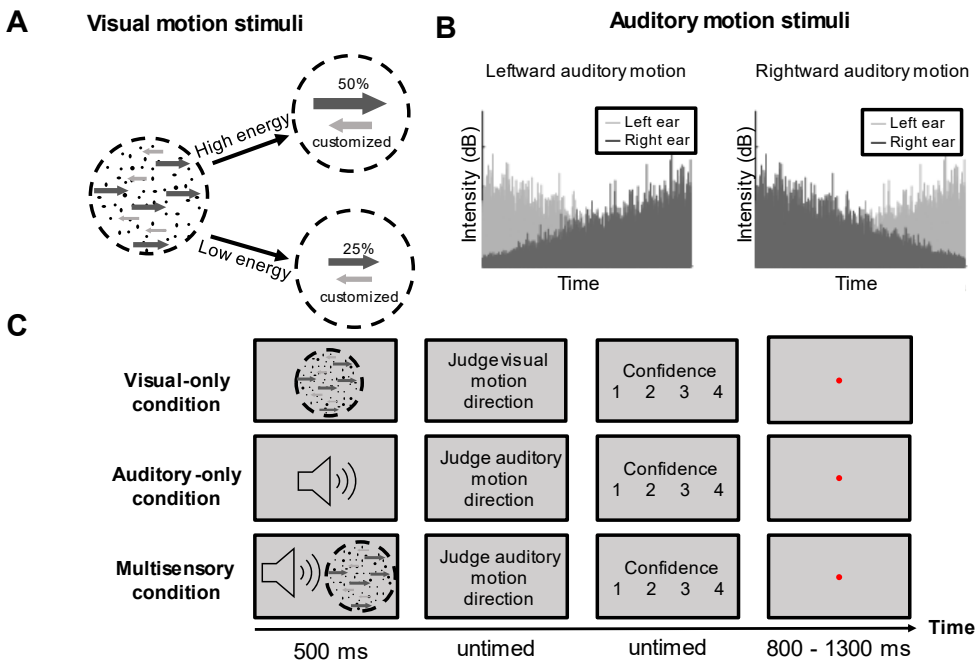

Figure 1. Experimental paradigm and stimuli. (A) Visual stimuli used in the experiments. The visual stimuli were random-dot kinematograms that consisted of dots moving in a dominant direction (either leftward or rightward for each trial), a non-dominant direction (always opposite to the dominant direction), and random directions. The number of dots moving in the dominant direction was fixed to 50% of the total number of dots for the high-energy stimuli and 25% for the low-energy stimuli. The number of dots moving in the non-dominant direction was customized in different ways for different participants (see Methods). The total number of dots was constant for stimuli with different energy levels. (B) Auditory stimuli used in the experiments. We used cross-faded white noise as auditory motion stimuli. For leftward motion, the sound played to the left ear faded in (i.e., the sound intensity increased over time), while the sound in the right ear faded out (i.e., the sound intensity decreased over time). The opposite was true for rightward motion. (C) Trial structure. Each trial started with motion stimuli (visual-only, auditory-only, or a combination of visual and auditory). Participants then judged the direction of motion (left vs. right) and provided confidence on a 4-point scale. In the multisensory condition, participants judged the auditory motion, but their judgments were typically influenced by the visual motion. The next trial started after a fixation interval of 800-1300 ms.

We first compared performance and confidence for the low- and high-energy stimuli in the visual-only condition. We found that high-energy visual stimuli led to lower performance (d') ($t(98) = 5.7$, $p = 1.0*10^{-7}$, Cohen's d = .58, $BF_{10} = 1.2*10^{5}$, two-sided paired t-test; Figure

2A left), but higher confidence ($t(98) = 9.1$, $p = 9.9*10^{-15}$, Cohen's $d = .92$, $BF_{10} = 6.9*10^{11}$; Figure 2A right) compared to the low-energy visual stimuli. Thus, we were able to achieve a strong confidence-accuracy dissociation such participants were more confident in judging the motion direction for high-energy visual stimuli despite their lower performance for those stimuli.

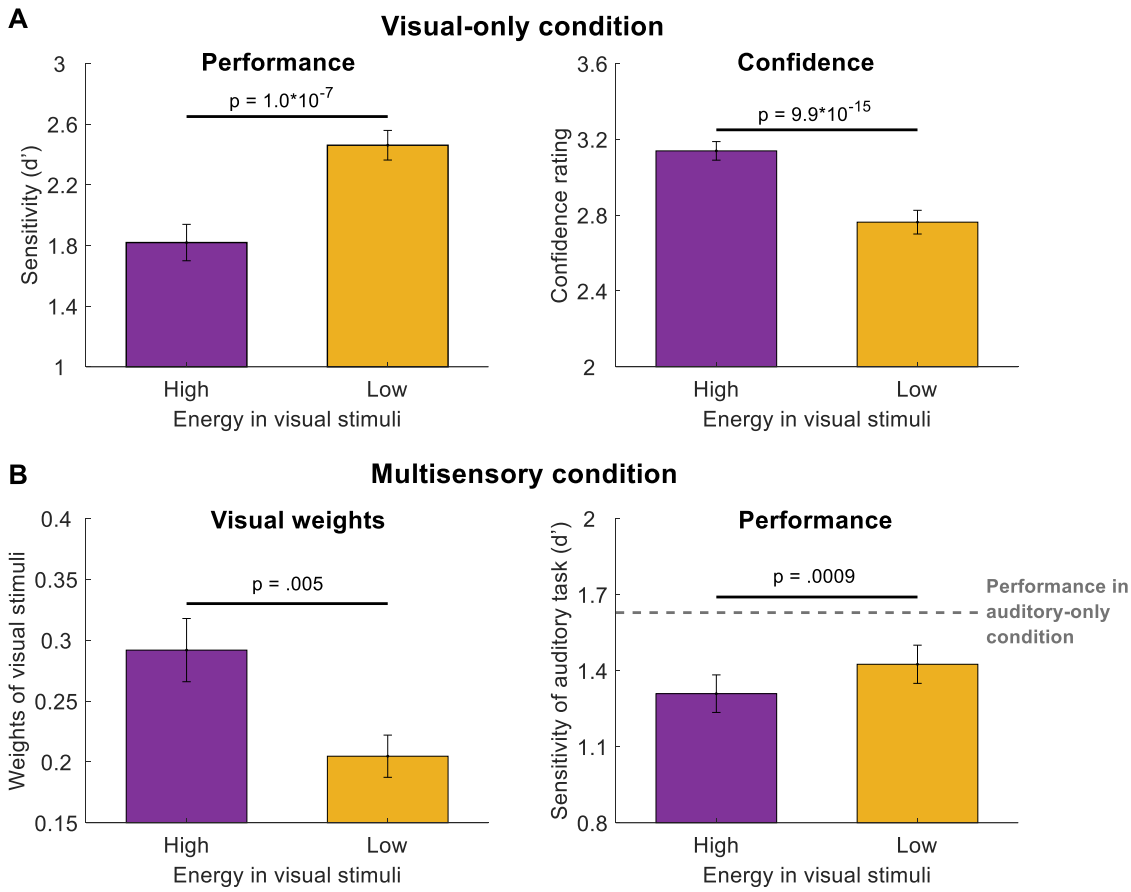

Figure 2. Empirical results. (A) Performance (d') and confidence in the visual-only condition. High-energy visual stimuli led to lower performance (left panel) but higher confidence (right panel). (B) Performance and weights for high- and low-energy visual stimuli in the multisensory condition. The left panel shows that visual stimuli were weighted more heavily in multisensory judgments for the high- compared to the low-energy visual stimuli. The right panel shows that both multisensory conditions had lower d' than the auditory-only condition (dashed line), with a larger decrease for the high- compared to the low-energy stimuli. Error bars indicate SEM.

Having established the existence of a confidence-accuracy dissociation in the visual-only condition between the low-and high-energy stimuli, we tested how each stimulus type affected cue combination in the multisensory condition. Using the observed performance in the visual-only and auditory-only conditions, we computed the weight of the visual information on the auditory judgments in the multisensory condition following standard procedures in multisensory cue combination studies (de Winkel et al., 2010; Drugowitsch et al., 2014, see Methods). When combining trials for the low- and high-energy visual stimuli, we found that the visual stimuli had a substantial influence (average visual weight = .25; average auditory weight = .75) even though participants were asked(de Winkel et al., 2010; Drugowitsch et al., 2014; Fetsch et al., 2009) to only judge the auditory direction in the multisensory condition. Critically, the weight was substantially higher for the high-energy (average weight = .29) compared to the low-energy stimuli (average weight = .20; t (98) = 3.43, p = .0009, Cohen's d = .35, BF_{10} = 25.27; Figure 2B left). Consistent with these results, we also observed that both multisensory conditions exhibited lower d' than the auditory-only condition (high-energy: t(98) = 5.48, p = $3.4*10^{-7}$, Cohen's d = .55, BF_{10} = $3.9*10^{4}$; low-energy: t(98) = 3.41, p = .001, Cohen's d = .34, BF_{10} = 23.28, Figure 2B right), but that the decrease was larger for the high- compared to the low-energy stimuli (t(98) = 2.90, p = .005, Cohen's d = .29, BF_{10} = 5.60). These results demonstrate that automatic multisensory cue combination was more strongly influenced by the high-energy visual stimuli in line with their higher confidence but despite their lower associated performance.

Having established that multisensory cue combination and metacognitive confidence judgments share similar behavioral signatures, we developed a model to reveal the computations underlying both types of processes. The model assumes that high-energy visual stimuli give rise to Gaussian distributions of internal evidence that have larger standard deviations than the distributions produced by low-energy visual stimuli (Figure 3A). As in previous work, we assume that participants give confidence ratings by applying the same confidence criteria for both high- and low-energy stimuli (Rahnev et al., 2011, 2012, 2013). Since the distributions for high-energy stimuli have higher variability, they occupy more of the high confidence regions and therefore lead to higher confidence even if they have lower d' levels, thus explaining the empirically observed confidence-accuracy dissociation in the visual-only condition. To make the multisensory judgments, the model then assumes that the same distributions with different variability for the low- and high-energy visual stimuli are directly combined with the auditory information without any normalization (Figure 3B). Because the high-energy distributions extend further to both extremes, they have a stronger influence on the judgments in the multisensory condition, an effect that mirrors their higher confidence in the visual-only condition.

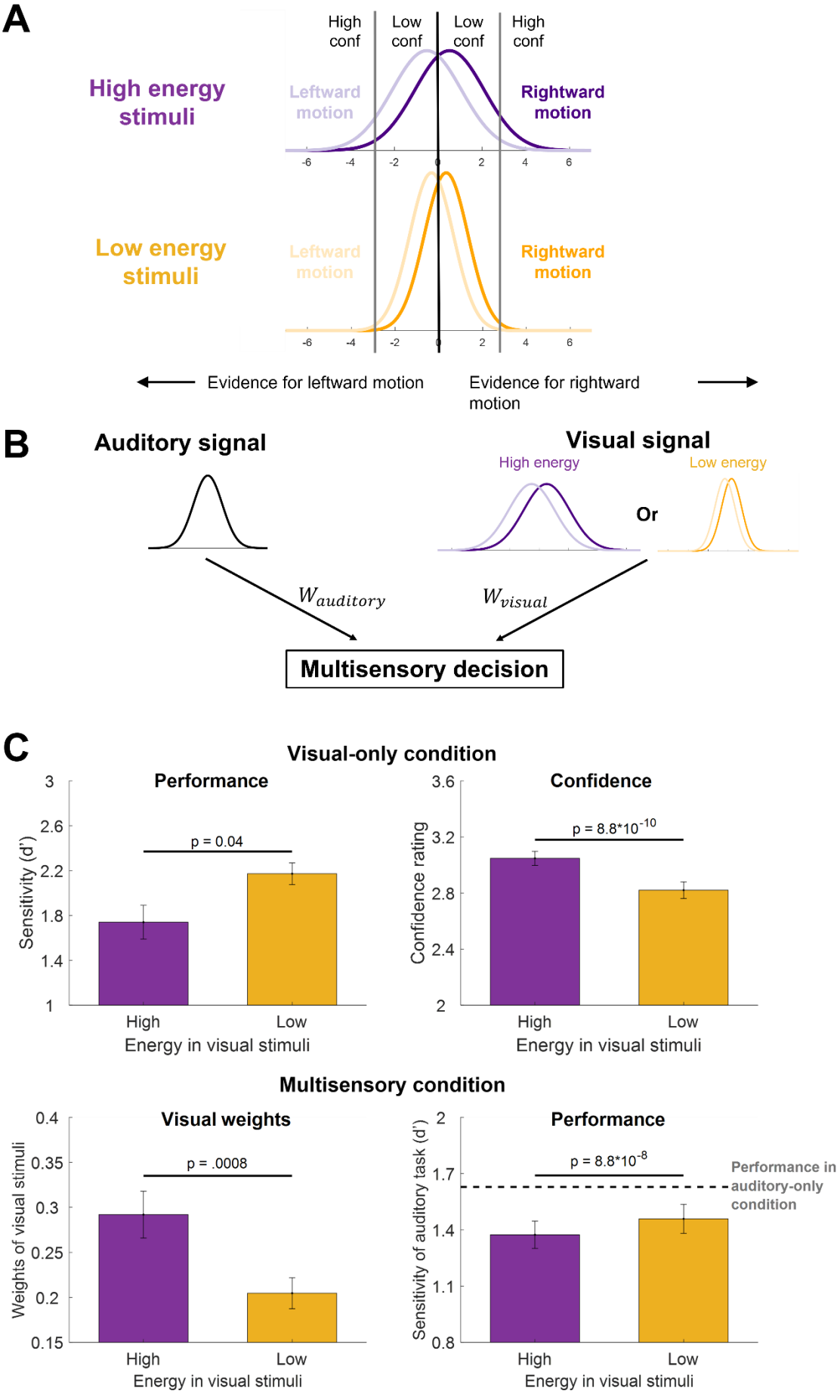
A
High conf
Low conf
Low conf
High conf
High energy stimuli
Leftward motion
Rightward motion
-6
-4
-2
0
2
4
6
Low energy stimuli
Leftward motion
Rightward motion
Evidence for leftward motion
Evidence for rightward motion
B
Auditory signal
Visual signal
High energy
Low energy
Or
$W_{auditory}$
W_{visual}
Multisensory decision
C
Visual-only condition
Performance
Sensitivity (d')
p = 0.04
High
Low
Energy in visual stimuli
Confidence
Confidence rating
p = 8.8*10^-10
Multisensory condition
Visual weights
Weights of visual stimuli
p = .0008
Performance
Sensitivity of auditory task (d')
p = 8.8*10^-8
Performance in auditory-only condition

Figure 3. Model description and fits. (A) Internal distributions of evidence for high- vs. low-energy stimuli. The model assumes that the distribution for high-energy stimuli has a larger variability compared to that of low-energy stimuli, resulting in more trials falling in the range in the tails with higher confidence. The distributions shown in the figure are the average distributions obtained after fitting the model to the data. (B) Multisensory-decision model. Visual signals are combined directly with auditory signals without any normalization. The perceptual decisions are made according to the weighted sum of the visual and auditory signals. (C) Model fits. The model successfully reproduced lower multisensory d' for the high-energy stimuli compared to the low-energy stimuli, consistent with the associated higher confidence but lower accuracy for the high-energy stimuli (top panels). The model well reproduced the higher estimated weight and the overall lower multisensory d' and for the high- compared to the low-energy visual stimuli (bottom panels).

We found that the model successfully reproduced the confidence-accuracy dissociation by producing lower performance ($t(98) = 2.98$, $p = .004$, Cohen's $d = .30$, $BF_{01} = 6.92$) but higher confidence ($t(98) = 6.79$, $p = 8.8*10^{-10}$, Cohen's $d = .68$, $BF_{10} = 1.1*10^{7}$) for the high- compared to the low-energy stimuli in the visual-only condition (Figure 3C, top panels). Critically, despite assuming a single true weight for the low-and high-energy visual stimuli in the multisensory condition, the model faithfully reproduced the higher estimated weight for the high- compared to the low-energy stimuli ($t(98) = 1.67$, $p = .10$, Cohen's $d = .17$, $BF_{01} = 2.34$), as well as the overall lower multisensory d' for the high- compared to the low-energy visual stimuli ($t(98) = 3.47$, $p = .0008$, Cohen's $d = .35$, $BF_{10} = 28.00$; Figure 3C, bottom panels). Therefore, despite its simplicity, the model was able to explain both the confidence-accuracy dissociation and the multisensory cue combination bias by postulating a unified computational principle.

Discussion

Perceptual decision making can span a wide range of processing from unconscious and automatic sensory inference in early sensory cortices (Alais & Burr, 2004; Ernst & Banks, 2002; Fetsch et al., 2012) to conscious and deliberate metacognitive reports mediated by the prefrontal cortex (Fleming et al., 2012; Mamassian, 2016; Rahnev, 2021b). Here we show that these vastly different stages of the perceptual decision-making process rely on common computational mechanisms for uncertainty estimation. Using an established manipulation for producing confidence-accuracy dissociations (Odegaard et al., 2018), we created visual stimuli with high or low energy and replicated previous findings that the high-energy stimuli lead to lower accuracy but higher confidence. Critically, we found that the high-energy visual stimuli had a stronger impact on auditory motion judgment in a multisensory cue combination task. This effect was consistent with the higher confidence of the high-energy visual stimuli but in contrast to their lower accuracy. A model assuming common computations underlying multisensory cue combination and confidence reports explained the data, suggesting the existence of shared mechanisms for uncertainty estimation between low-level automatic and high-level metacognitive stages of perceptual decision making.

Our results represent a critical extension to a growing literature that has sought to uncover underlying computations that apply across different domains of cognition, perception, and action. Previous studies have demonstrated that the metacognitive ability to provide confidence ratings predictive of one's accuracy is at least partly domain-general (Faivre et al.,

2018; Mazancieux et al., 2020). Similarly, recent work has demonstrated similar computations across cognition, metacognition, and even meta-metacognition (Zheng et al., 2023). However, none of this previous work has examined whether automatic sensory inference performed within sensory areas of the brain may also share mechanisms with the deliberate computations associated with higher-order cognition.

Our model implies that stimuli with higher variability are judged with higher confidence and are given higher weight in multisensory judgments. This implication may at first appear counterintuitive. Indeed, in estimation tasks, higher variability occurs for more uncertain stimuli, which consequently leads to both lower confidence (Bertana et al., 2021; Geurts et al., 2022) and lower weight in multisensory judgments (Alais & Burr, 2004; Ernst & Banks, 2002; Morgan et al., 2008). The critical distinction here is that while stimulus strength in estimation tasks is reflected in the variability around the true stimulus feature (e.g., length, orientation, etc.), the stimulus strength in 2-choice tasks is reflected in the mean of the distributions of internal evidence (Macmillan & Creelman, 2004). As such, the effects of larger variability in estimation tasks mimic the effects of lower distance between the two distributions in 2-choice tasks. On the other hand, the variability in the internal distributions in 2-choice tasks has unique properties such that larger variance can decrease sensitivity but increase the proportion of very strong signals, thus leading to both higher confidence and higher weights for cue combination.

A critical component of our model is that both metacognition and multisensory cue

combination feature incomplete normalization of sensory signals that differ in their level of uncertainty. Such incomplete normalization has been previously proposed in the context of confidence judgments (Morales et al., 2015; Rahnev et al., 2011, 2012, 2013) but has not been examined in the context of multisensory cue combination. In fact, incomplete normalization may explain many puzzling deviations from optimality observed in cue combination studies (Battaglia et al., 2003; Fetsch et al., 2012; Rosas et al., 2005). It should be noted that our model assumes a complete lack of normalization, which has been criticized before (Denison et al., 2018; Lee et al., 2023). Our choice was driven by the fact that our data do not allow us to distinguish models with no normalization vs. only partial normalization, and models with partial normalization require extra parameters that make them less parsimonious. Future work should explore the degree of normalization achieved in metacognitive and multisensory cue combination judgments. The critical insight of the current work is thus not that no normalization exists, but that normalization in multisensory cue combination and confidence judgments must necessarily be incomplete.

In conclusion, our work demonstrates the existence of common computational mechanisms across vastly different stages of perceptual decision making, and may point to the existence of unified inference mechanisms throughout cortex.

Methods

Participants

We collected data from 99 participants (30 females and 69 males, aged from 17 to 33, mean age = 20.5, SD = 2.6). Participants were compensated with either 1 SONA credit or monetary reward ($10/hr). All participants had normal or corrected-to-normal vision and normal hearing abilities. Before the experiment, all participants signed the consent form approved by the Institutional Review Board of Georgia Institute of Technology.

Stimuli

The experiment featured both visual and auditory motion stimuli. The visual stimuli were random-dot kinematograms that included three types of dot motion: (1) dots moving in the dominant direction, (2) dots moving in the non-dominant direction (opposite to the dominant direction), and (3) dots moving in random directions (Figure 1A). The dominant direction was either leftward or rightward, randomized in each trial. The total number of dots in the motion stimuli was fixed to 4,241. Visual stimuli were presented at the center of the screen. Each dot had a diameter of 0.05 degrees and moved at 5 degrees per second. All dots were black and were moving within an invisible circle with a diameter of 6 degrees. A red fixation dot was presented at the center of the screen throughout the experiment.

We created two types of motion stimuli: high-energy stimuli where 50% of all dots moved in the dominant direction, and low-energy stimuli where 25% of all dots moved in the dominant direction. We customized the percent of dots moving in the non-dominant

direction. In one group of participants (N=24) we attempted to match the performance for high- vs. low-energy stimuli. To do so, we ran 3-down-1-up staircases on the proportion of dots moving in the non-dominant direction separately for the high- vs. low-energy stimuli. The initial proportion of dots moving in the non-dominant direction was 6% and 3.5% for high- and low-energy stimuli. The initial step sizes of the staircases were 1.5% and 0.9%, and were reduced to 1.0% and 0.6% after three reversals, and to 0.5% and 0.3% after another three reversals. The staircase ended after 150 trials or 14 reversals, whichever came first. The threshold was determined as the average value at the time of the last eight reversals. In a second group of participants (N=25), we fixed the percent of dots moving in the non-dominant direction to 45% for the high-energy stimuli and 0% for the low-energy stimuli. Finally, in a third group of participants (N=50), we fixed the percent of dots moving in the non-dominant direction to 31% for the high-energy stimuli and 0% for the low-energy stimuli. To maximize power, we analyzed the data from all participants together. However, analyzing each group separately led to similar results.

The auditory motion stimuli consisted of signal sounds of either leftward or rightward direction and white noise sounds with no direction. To create rightward auditory motion, the intensity of the signal sound in the right ear increased from zero to maximum loudness while the intensity of the signal sound in the left ear decreased from maximum loudness to zero (Figure 1B). The opposite was true for leftward motion. To determine the relative loudness intensity of the signal and noise sounds, we ran a 2-down-1-up staircase on the ratio of the signal sounds to the total loudness (signal + noise sounds), and used the obtained ratio for

the target motion in the formal sessions. The starting point of the staircase was 70%; the initial step size was 3%, which was reduced to 2% after three reversals, and to 1% after another three reversals. The staircase ended after 120 trials or 12 reversals, whichever came first. The threshold was determined as the average value at the time of the last six reversals.

Procedure

The experiment included three different conditions: visual-only, auditory-only, and multisensory. In all cases, participants judged the direction of motion (leftward vs. rightward) and rated their confidence on a 4-point scale (lowest, low, high, highest). The visual-only and auditory-only conditions featured stimuli from a single modality. Critically, the multisensory condition included both visual and auditory stimuli, but participants were instructed to judge the auditory motion direction and ignore the visual motion direction. Motion stimuli were presented for 500 ms and all responses were untimed. After the end of a trial, a fixation circle was shown for a period randomly chosen between 800-1300 ms before the next trial started (Figure 1C).

The experiment began with a screening session intended to remove participants who could not properly perform the visual-only task. During this session, participants completed blocks that contained 25 trials with low-energy stimuli (where we set 0% of dots to move in the non-dominant direction) and 25 trials with high-energy stimuli (where we set 8.3% of dots to move in the non-dominant direction). We passed participants who could perform at 80% correct or better for both types of stimuli. Fourteen participants were unable to reach that

threshold for the low-energy stimuli even after up to eight screening blocks and therefore did not participate in the main experiment. These participants are not included in the count of recruited participants.

The main experiment consisted of two visual-only, one auditory-only, and two multisensory blocks. Each block consisted of 80 trials and the different types of blocks were randomly interleaved for each participant (400 trials total). The visual-only and multisensory blocks contained an equal number of trials with high- and low-energy stimuli. In the multisensory blocks the direction of auditory motion was pseudo-randomized to contain equal number of leftward and rightward motion, whereas the direction of visual motion was fully randomized such that the two stimuli were congruent on average in 50% of the trials. The third group of participants mentioned above completed four visual-only blocks and 10 blocks where auditory-only and multisensory trials were interleaved (45 trials per block, 630 trials total) but did not need to indicate confidence for the auditory-only and multisensory trials.

Apparatus

We presented all visual stimuli on a Dell monitor (47.5 cm x 26.5 cm; refresh rate = 70 Hz) positioned 50 cm away from the participants. We delivered the auditory stimuli through SENNHEISER HD 280 PRO headphones. All sounds had a sampling frequency of 44.1 kHz.

Analysis

For each condition, we calculated task performance (d') based on the signal detection theory (Green & Swets, 1966) using the formula:

$$d' = \varphi^{-1}(hit\ rate) - \varphi^{-1}(false\ alarm\ rate) \quad (1)$$

where φ^{-1} denotes to the inverse of the cumulative standard normal distribution converting the hit and false alarm rates to Z scores.

We performed two-sided paired sample t-tests to compare performance (d') and confidence levels between the high- and low-energy stimuli in the visual-only condition, as well as both the visual weights (see below for how those were estimated) and overall performance (d') in the multisensory condition. We report Cohen's d and the Bayes factors (Krekelberg, 2022) for all t-tests. All the statistical analyses were run with MATLAB (The Math Works, Version 2022a).

<u>Model description</u>

We developed a model of the computations involved in confidence ratings in the visual-only condition and multisensory decisions in the multisensory condition. The model assumes that both visual and auditory evidence is sampled from Gaussian distributions in accordance with signal detection theory (Green & Swets, 1966) such that leftward and rightward motion stimuli produce distributions of internal evidence coming from $N\left(-\frac{\mu}{2}, \sigma^2\right)$ and $N\left(\frac{\mu}{2}, \sigma^2\right)$, respectively. Critically, more energy in the visual signals for the high-energy stimuli increases

the variability of the internal signal (Figure 3A). Participants give confidence ratings by applying the same criteria for both high- and low-energy stimuli. Therefore, the high-energy distributions occupy more of the high-confidence regions compared to the low-energy distributions, leading to more high-confidence trials (Rahnev et al., 2011, 2012, 2013). To perform multisensory cue combination, the model assumes that the auditory signals are directly combined with the visual stimuli without any normalization, such that the decision variable for the multisensory decisions is:

$$x_{multisensory} = w * x_{visual} + (1 - w) * x_{auditory} \quad (2)$$

where w is between 0 and 1 (Figure 3B). Because the distributions for the high-energy visual stimuli extend further to both extremes compared to the distributions for the low-energy visual stimuli, such combination results in the high-energy visual stimuli having a stronger impact on auditory judgments.

We fit the model to the data in two steps: first, we fit the model to the visual-only and auditory-only data, and then using the parameters obtained in the first step we fit to the multisensory data using a single additional free parameter. For the auditory-only condition, the model had one free parameter for the distance between the peaks of the Gaussian distributions of internal evidence, $\mu_{auditory}$, with the standard deviation (SD) of the evidence distributions fixed to 1. The parameter $\mu_{auditory}$ was estimated directly from the data using equation 1 above. For the visual-only condition, the model had 10 free

parameters: the distance between the peaks of the high-energy distributions, μ_{HE}, the distance between the peaks of the low-energy distributions, μ_{LE}, the SD of the high-energy distributions, σ_{HE} and 7 response criteria c_i with $i = -3, -2, -1, 0, 1, 2, 3$. Because the model predictions depend only on the ratio between the SDs for the high- and low-energy stimuli, without loss of generality the SD of the low-energy distributions, σ_{LE}, was fixed to 1. The criterion c_0 was the decision criterion that separated the two stimulus categories, whereas the rest of the criteria determined the confidence ratings, such that the criteria c_i and c_{-i} separated the ratings i and $i + 1$. Negative criteria separate the confidence ratings for leftward motion decisions, whereas positive criteria separate the confidence ratings for rightward motion decisions. Thus, given the presentation of a stimulus that produces an internal activation $N(\mu, \sigma^2)$, the probability of confidence rating of i for a rightward decision is:

$$P(conf = i, dec = rightward) = \int_{c_{i-1}}^{c_i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3)$$

and the probability of confidence rating of i for a leftward decision is:

$$P(conf = i, dec = leftward) = \int_{c_{-i}}^{c_{-i+1}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (4)$$

where $c_{-4} = -\infty$ and $c_4 = \infty$.

Once we fit the auditory-only and visual-only data, we then fit the multisensory data using a

single free parameter for the weight with which visual stimuli contributed to the multisensory decision variable. We used the obtained parameters μ_{HE}, μ_{LE}, σ_{HE}, and $\mu_{auditory}$ to simulate the internal visual and auditory activations, and equation 2 to obtain the decision variable for the multisensory decisions, $x_{multisensory}$. Rightward decisions were made for $x_{multisensory} \geq 0$, and leftward decisions for $x_{multisensory} < 0$.

To fit the model to the data, we used maximum likelihood estimation as in previous studies (Rahnev, 2021; Shekhar & Rahnev, 2021; Yeon & Rahnev, 2020). Model fitting was performed using the Bayesian Adaptive Direct Search (BADS) toolbox, version 1.0.5 (Acerbi & Ma, 2017).

To estimate the weights for the high- and low-energy visual stimuli (Figure 2B left), we used a similar procedure as above except that the weights were separately estimated for high- and low-energy stimuli with the SDs of both distributions set to 1. This procedure is equivalent to what is usually done in multisensory cue combination studies (de Winkel et al., 2010; Drugowitsch et al., 2014) where the signals from different modalities are combined based on their respective sensitivities irrespective of the internal variability of each signal.

Data availability

Data and code for analysis and modeling are available at https://osf.io/crkjd/files/osfstorage.

References

Acerbi, L., & Ma, W. J. (2017). *Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search*. https://proceedings.neurips.cc/paper/2017/file/df0aab058ce179e4f7ab135ed4e641a9-Paper.pdf

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alink, A., Euler, F., Galeano, E., Krugliak, A., Singer, W., & Kohler, A. (2012). Auditory motion capturing ambiguous visual motion. *Frontiers in Psychology*, *3*(JAN). https://doi.org/10.3389/fpsyg.2011.00391

Arnold, D. H., Petrie, K., Murray, C., & Johnston, A. (2019). Suboptimal human multisensory cue combination. *Scientific Reports*, *9*(1), 5155. https://doi.org/10.1038/s41598-018-37888-7

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391. https://doi.org/10.1364/JOSAA.20.001391

Bertana, A., Chetverikov, A., van Bergen, R. S., Ling, S., & Jehee, J. F. M. (2021). Dual strategies in human confidence judgments. *Journal of Vision*, *21*(5), 21. https://doi.org/10.1167/jov.21.5.21

Bruns, P. (2019). The Ventriloquist Illusion as a Tool to Study Multisensory Processing: An Update. *Frontiers in Integrative Neuroscience*, *13*. https://doi.org/10.3389/fnint.2019.00051

de Winkel, K. N., Weesie, J., Werkhoven, P. J., & Groen, E. L. (2010). Integration of visual and inertial cues in perceived heading of self-motion. *Journal of Vision*, *10*(12), 1–1. https://doi.org/10.1167/10.12.1

Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proceedings of the National Academy of Sciences*, *115*(43), 11090–11095. https://doi.org/10.1073/pnas.1717720115

Dobromir Rahnev. (2021). A robust confidence–accuracy dissociation via criterion attraction. *Neuroscience of Consciousness*, *2021*(1). https://doi.org/10.1093/nc/niab039

Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *ELife*, *3*. https://doi.org/10.7554/eLife.03005

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* , *415*, 429–433.

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, Modeling, and Electrophysiological Evidence for Supramodality in Human Metacognition. *The Journal of Neuroscience*, *38*(2), 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017

Fetsch, C. R., Pouget, A., Deangelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–154. https://doi.org/10.1038/nn.2983

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic Reweighting of

Visual and Vestibular Cues during Self-Motion Perception. *The Journal of Neuroscience*, *29*(49), 15601–15612. https://doi.org/10.1523/JNEUROSCI.2574-09.2009

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1280–1286. https://doi.org/10.1098/rstb.2012.0021

Geurts, L. S., Cooke, J. R. H., van Bergen, R. S., & Jehee, J. F. M. (2022). Subjective confidence reflects representation of Bayesian probability in cortex. *Nature Human Behaviour*, *6*(2), 294–305. https://doi.org/10.1038/s41562-021-01247-w

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review*, *104*(2), 344–366. https://doi.org/10.1037/0033-295X.104.2.344

Kim, R., Peters, M. A. K., & Shams, L. (2012). 0 + 1 > 1: How Adding Noninformative Sound Improves Performance on a Visual Task. *Psychological Science*, *23*(1), 6–12. https://doi.org/10.1177/0956797611420662

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, *416*(6877), 172–174. https://doi.org/10.1038/416172a

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. https://doi.org/10.1016/S0042-6989(03)00458-9

Krekelberg, B. (2022). *BayesFactor* (2.3.0). Zenodo. https://doi.org/https://doi.org/10.5281/zenodo.7006300

Landy, M. S., Banks, M. S., & Knill, D. C. (2011). Ideal-Observer Models of Cue Integration. In *Sensory Cue Integration* (pp. 5–29). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195387247.003.0001

Lee, J. L., Denison, R., & Ma, W. J. (2023). Challenging the fixed-criterion model of perceptual decision-making. *Neuroscience of Consciousness*, *2023*(1). https://doi.org/10.1093/nc/niad010

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, *37*, 205–220. https://doi.org/10.1146/annurev-neuro-071013-014017

Macmillan, N., & Creelman, D. (2004). *Detection theory: A user's guide*. Psychology Press.

Maiworm, M., & Röder, B. (2011). Suboptimal auditory dominance in audiovisual integration of temporalcues. *Tsinghua Science and Technology*, *16*(2), 121–132. https://doi.org/10.1016/S1007-0214(11)70019-0

Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, *2*(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Maniscalco, B., & Lau, H. (2014). Signal Detection Theory Analysis of Type 1 and Type 2 Data: Meta-d', Response-Specific Meta-d', and the Unequal Variance SDT Model. In *The Cognitive Neuroscience of Metacognition* (pp. 25–66). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-45190-4_3

Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2020). Is there a G factor for

metacognition? Correlations in retrospective metacognitive sensitivity across tasks. *Journal of Experimental Psychology: General*, *149*(9), 1788–1799. https://doi.org/10.1037/xge0000746

Morales, J., Solovey, G., Maniscalco, B., Rahnev, D., de Lange, F. P., & Lau, H. (2015). Low attention impairs optimal incorporation of prior knowledge in perceptual decisions. *Attention, Perception, & Psychophysics*, *77*(6), 2021–2036. https://doi.org/10.3758/s13414-015-0897-2

Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (2008). Multisensory Integration in Macaque Visual Cortex Depends on Cue Reliability. *Neuron*, *59*(4), 662–673. https://doi.org/10.1016/j.neuron.2008.06.024

Mueller, S. T., & Weidemann, Christoph. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, *15*(3), 465–494. https://doi.org/10.3758/PBR.15.3.465

Odegaard, B., Grimaldi, P., Cho, S. H., Peters, M. A. K., Lau, H., & Basso, M. A. (2018). Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), E1588–E1597. https://doi.org/10.1073/pnas.1711628115

Plaisier, M. A., van Dam, L. C. J., Glowania, C., & Ernst, M. O. (2014). Exploration mode affects visuohaptic integration of surface orientation. *Journal of Vision*, *14*(13), 22–22. https://doi.org/10.1167/14.13.22

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374. https://doi.org/10.1038/nn.4240

Rahnev, D. (2021a). A robust confidence–accuracy dissociation via criterion attraction. *Neuroscience of Consciousness*, *2021*(1). https://doi.org/10.1093/nc/niab039

Rahnev, D. (2021b). Visual metacognition: Measures, models, and neural correlates. *American Psychologist*, *76*(9), 1445–1453. https://doi.org/10.1037/amp0000937

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology*, *107*(6), 1556–1563. https://doi.org/10.1152/jn.00985.2011

Rahnev, D., Bahdo, L., de Lange, F. P., & Lau, H. (2012). Prestimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *J Neurophysiol*, *108*, 1529–1536. https://doi.org/10.1152/jn.00184.2012.-Attention

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/S0140525X18000936

Rahnev, D., Kok, P., Munneke, M., Bahdo, L., de Lange, F. P., & Lau, H. (2013). Continuous theta burst transcranial magnetic stimulation reduces resting state connectivity between visual areas. *Journal of Neurophysiology*, *110*(8), 1811–1821. https://doi.org/10.1152/jn.00209.2013

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, *14*(12), 1513–1515. https://doi.org/10.1038/nn.2948

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). *Texture and haptic cues in*

slant discrimination: reliability-based cue weighting without statistically optimal cue combination.

Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016). Dissociating Perceptual Confidence from Discrimination Accuracy Reveals No Influence of Metacognitive Awareness on Working Memory. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00851

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. *Neuron*, *90*(3), 499–506. https://doi.org/10.1016/j.neuron.2016.03.025

Shekhar, M., & Rahnev, D. (2021a). Sources of Metacognitive Inefficiency. *Trends in Cognitive Sciences*, *25*(1), 12–23. https://doi.org/10.1016/j.tics.2020.10.007

Shekhar, M., & Rahnev, D. (2021b). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review*, *128*(1), 45–70. https://doi.org/10.1037/rev0000249

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research*, *14*(1), 139–146. https://doi.org/10.1016/S0926-6410(02)00068-X

Soto-Faraco, S., Spence, C., & Kingstone, A. (2005). Assessing automaticity in the audiovisual integration of motion. *Acta Psychologica*, *118*(1–2), 71–92. https://doi.org/10.1016/j.actpsy.2004.10.008

van Bergen, R. S., Ji Ma, W., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*, *18*(12), 1728–1730. https://doi.org/10.1038/nn.4150

Vlassova, A., Donkin, C., & Pearson, J. (2014). Unconscious information changes decision accuracy but not confidence. *Proceedings of the National Academy of Sciences*, *111*(45), 16214–16218. https://doi.org/10.1073/pnas.1403619111

Yeon, J., & Rahnev, D. (2020). The suboptimality of perceptual decision making with multiple alternatives. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17661-z

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Zheng, Y., Recht, S., & Rahnev, D. (2023). Blurring the lines between cognition and metacognition: the case of meta-metacognition. *PsyArXiv*.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00079