Improvement of association between confidence and accuracy after integration of discrete evidence over time

Zahra Azizi[1], Sajjad Zabbah[2], Azra Jahanitabesh[3], Reza Ebrahimpour[2,4*]

[1] Department of Cognitive Modeling, Institute for Cognitive Science Studies, Tehran, Iran.

[2] Institute for Research in Fundamental Sciences, School of Cognitive Sciences, Tehran, Iran.

[3] Department of Psychology, University of California, Davis, California, United States.

[4] Department of Artificial Intelligence, Faculty of Computer Engineering, Shahid Rejaee Teacher Training University, Tehran, Iran.

* Correspondence:

Reza Ebrahimpour

ebrahimpour@ipm.ir

Keywords: Confidence; metacognition; discrete pieces of evidence; perceptual decision-making; pupillometry; ERP.

Abstract

When making decisions in real-life, we may receive discrete pieces of evidence during a time period. Although subjects are able to integrate information from separate cues to improve their accuracy, confidence formation is controversial. Due to a strong positive relation between accuracy and confidence, we predicted that confidence followed the same characteristics as accuracy and would improve following the integration of information collected from separate cues. We applied a Random-dot-motion discrimination task in which participants had to indicate the predominant direction of dot motions by saccadic eye movement after receiving one or two brief stimuli (i.e., pulse(s)). The interval of two pulses (up to 1s) was selected randomly. Color-coded targets facilitated indicating confidence simultaneously. Using behavioral data, computational models, pupillometry and EEG methodology we show that in double-pulse trials: (i) participants improve their confidence resolution rather than reporting higher confidence comparing with single-pulse trials, (ii) the observed confidence follow neural and pupillometry markers of confidence, unlike in weak and brief single-pulse trials. Overall, our study showed improvement of associations between confidence and accuracy in decision results from the integration of stimulus separated by different temporal gaps.

1 Introduction

Humans and animals can both make choices based on multiple discrete pieces of information. Imagine that a large bus is passing between you and a faraway car as you cross the street. In this situation, simply by collecting discrete pieces of information about the car's position through the windows of the bus, you can decide whether the car is moving toward or away from you. In this scenario, as the number of pieces of information increased, the interpretation of the car's direction would be improved. Indeed, research has shown that the accuracy of decisions can be significantly improved by integrating information from separate cues (Kiani, Churchland, & Shadlen, 2013; Kira, Yang, & Shadlen, 2015; tickle, Tsetsos, Speekenbrink, & Summerfield, 2020; Tohidi-Moghaddam, Zabbah, Olianezhad, & Ebrahimpour, 2019; Waskom & Kiani, 2018). Typically, our decisions are accompanied by feelings that reflect the likelihood that the decision is correct; such a feeling is called confidence (Kiani, Corthell, & Shadlen, 2014). For example, imagine that the scene in the previous scenario is also included a foggy weather. In this case, low visibility may reduce the confidence of your judgments. This diminished confidence per se may lead to change your mind (Fleming, Putten, & Daw, 2018; Resulaj, Kiani, Wolpert, & Shadlen, 2009), impact your behavioral adjustments, and affect how quickly and accurately you make your consecutive decisions (Meyniel, Sigman, & Mainen, 2015; van den Berg, Zylberberg, Kiani, Shadlen, & Wolpert, 2016). Due to the potential effects of confidence on decision-making, in the last few years, considerable progresses had been made in the understanding of the behavioral (Kiani et al., 2014; Zylberberg, Barttfeld, & Sigman, 2012) and the neuronal (Baranski et al., 2017; Gherman & Philiastides, 2015; Kiani & Shadlen, 2009) properties of confidence and its association with perceptual decision-making. However, how confidence is established within a discrete environment is still unclear.

According to the leading computational approach in perceptual decision-making (Gold & Shadlen, 2007; Shadlen & Kiani, 2013), when the accumulated evidence for one option, called a decision variable (DV), crosses a threshold or a boundary, a decision would be made. In addition, confidence is briefed by the probability that a decision relying on the DV is correct (Kiani et al., 2014; Kiani & Shadlen, 2009; van den Berg et al., 2016; Zylberberg, Fetsch, & Shadlen, 2016). Research has confirmed a strong positive relation between accuracy and confidence (Kiani et al., 2014; Vafaei Shooshtari, Esmaily Sadrabadi, Azizi, & Ebrahimpour, 2019). Moreover, it has been shown that, when we need to decide based on the discrete pieces of evidence, the decision is determined by integrating the DV of all those pieces (Kiani et al., 2013; Waskom & Kiani, 2018) and the accuracy even exceeded expectations predicted by evidence integration models (Kiani et al., 2013). Accordingly, one may suggest that confidence would follow the same characteristics as accuracy and would increase considerably after receiving separate pieces of information.

Nevertheless, a large body of evidence (e.g. (Herce Castañón et al., 2019; Zylberberg et al., 2016)) determines that human observers do not report their confidence in consistent with their accuracy. From this standpoint, noise can be considered as the key parameter to clarify variations in confidence (Kiani et al., 2014; Zylberberg et al., 2012). For instance, an underestimation of sensory noise in decisions would lead to over and/or under-confidence (De Gardelle & Mamassian, 2015; Herce Castañón et al., 2019; Zylberberg, Roelfsema, & Sigman, 2014) such that observers may ignore evidence in favor of other alternatives (Zylberberg et al., 2012). Moreover, confidence ratings may not only originate from the available sensory evidence (Rahnev & Denison, 2018; Zylberberg et al., 2016). So, the observers may integrate additional evidence into their confidence rating, which was not used for making their decision, allowing them to change their mind after the initiation of a response (Atiya et al., 2020; Resulaj et al., 2009). This suggests that computational description of confidence would be controlled by the attendance of both decision and confidence performance (Balsdon, Wyart, & Mamassian, 2020; Maniscalco & Lau, 2014).

To test the hypothetical relation between the accuracy and confidence, in binary decisions, signal detection theory (SDT) can provide a method to characterize how well the observers reporting the confidence ratings by introducing metacognitive sensitivity and efficiency (**Figure 1B**; (Fleming, 2017; Maniscalco & Lau, 2012, 2014)). In fact, for years, SDT has provided a simple yet powerful methodology to distinguish between an observer's ability to categorize the stimulus and the behavioral response (Green & Swets, 1966), and to determine confidence resolution.

Moreover, levels of confidence can be tracked by behavioral, neural and pupillometry signatures. Higher confidence are accompanied by faster and more accurate decisions (Kiani et al., 2014; van den Berg et al., 2016; Zylberberg et al., 2016). In addition, research on perceptual decision-making has established an EEG potential characterized by a centro-parietal positivity (CPP) as a neural correlate of sensory evidence accumulation (Kelly & O'Connell, 2013; O'connell, Dockree, & Kelly, 2012) and confidence (Boldt, Schiffer, Waszak, & Yeung, 2019; Herding, Ludwig, von Lautz, Spitzer, & Blankenburg, 2019; Tagliabue et al., 2019; Vafaei Shooshtari et al., 2019; Zizlsperger, Sauvigny, Händel, & Haarmeier, 2014). In particular, it has been shown that the CPP, despite the difference, is present for both correct and incorrect decisions (O'connell et al., 2012; Steinemann, O'Connell, & Kelly, 2018) and can reflect not only external evidence but also an internal decision quantity such as decision confidence. In addition, levels of confidence can be tracked by monitoring the pupil. The literature has suggested strong links between pupil dilation and both the decision (Murphy, Boonstra, & Nieuwenhuis, 2016) and confidence (Allen et al., 2016; Lempert, Chen, & Fleming, 2015; Urai, Braun, & Donner, 2017) via pupil-linked dynamics of the noradrenergic system (Laeng, Sirois, & Gredebäck, 2012). For example, pupillometry has provided some evidence that shows a partial dissociation between choice and confidence in decision-making (Balsdon et al., 2020). Considering the potential of response-time, CPP and pupillometry signatures to capture the distinction between choice and confidence in decision-making, they can be considered as informative paradigms to explore the confidence-accuracy association.

Accordingly, to bridge the existing gap in the confidence and perceptual decision-making literature, we implemented two separate experiments to explore three questions: First, how participants accumulate discrete evidence to establish confidence judgments. Second, whether the confidence ratings are in accordance with accuracy after integration of discrete evidence. Finally, how implicit markers of confidence —response-time, CPP and pupillometry— change after receiving separated pieces of information. Here, to clarify confidence, we required observers to make a two-alternative decision after viewing either one (single-pulse) or two (double-pulse) motion pulses separated by four various temporal gaps (similar to (Kiani et al., 2013; Tohidi-Moghaddam et al., 2019)). We performed several logistic regression models to measure the impact of stimulus characteristics on confidence. Also, we applied a set of computational models based on SDT to assess how accuracy and confidence varied throughout the experiments. Then, in the second experiment, we used EEG methodology to examine the relation between participants' brain activity and their confidence. We expected a neural indicator of perceptual decision making (CPP) would show amplitude changes between the two levels of confidence. In addition to behavioral data and EEG methodology, participants' pupil response was monitored across both experiments to examine the relation between participants' pupil dilation and their confidence. The findings expose that participant integrated information from pulses, invariant to the temporal gap, to improve the confidence resolution instead of reporting higher confidence. Likewise, in double-pulse trials, behavioral, neural and pupillometry markers of confidence would be distinguishable, entirely unlike in brief and weak single-pulse trials.

2 Materials and Methods

2.1 Participants

Consistent with methodological considerations in previous studies, a total of 19 observers participated in the two experiments. Six participants (three male; $M_{age} = 32.25$; $SD_{age} = 4.5$) attended in our behavioral experiment —Experiment 1— and 13 participants (three males; $M_{age} = 31.41$; $SD_{age} = 5.56$) took part in our EEG experiment —Experiment 2. All participants had normal or corrected-to-normal vision, and none of them had any history of psychiatric and neurological disorders. Previous studies with the same paradigm in which a large number of trials were presented to a small number of participants (e.g., five participants in(Kiani et al., 2013); six participants in (Kiani et al., 2014); four participants in (van den Berg et al., 2016) and six participants in (Stine, Zylberberg, Ditterich, & Shadlen, 2020)), assume that with extensive training, all participants would reach an acceptable level of performance. As such, a small number of trained participants would perform similar to the performance of a large number of participants. Accordingly, to make participants' performance reach the same criteria and reduce the between-participant variability, all participants received extensive training sessions on the Random-dot-motion discrimination task prior to data collection. Moreover, participants' understanding of the confidence reporting procedure was double-checked prior to the experiments. In Experiment 1, one participant was excluded due to the difficulty in reporting decision and confidence simultaneously, and another participant decided to leave the experiment shortly after participation. In addition, one participant was excluded from Experiment 2 because of the excessive noise in EEG electrodes crucial to the analysis.

2.2 Stimuli

We explored the confidence formation in discrete environment with a random-dot-motion (RDM) discrimination task. Participants had to indicate the predominant motion direction of a cloud of moving dots (left or right) presented within a 5° circular aperture at the center of the screen. The dot density was 16.7 dots/degree^2/s and the displacement of the coherently moving dots produced an apparent speed of 6 deg/s. The RDM movies were generated by three interleaved sets of dots presented on consecutive video frames. Three video frames later, each dot was redrawn at a location consistent with the direction of motion or at a random location within the stimulus space. More details can be found in previous studies (e.g. (Roitman & Shadlen, 2002)). The experiment code was programmed in MATLAB 2016a (Mathworks Inc., USA) using PsychToolbox (Brainard & Vision, 1997; Kleiner, Brainard, & Pelli, 2007)

2.3 Experimental Tasks

Participants performed the RDM task in blocks of 200 trials. Each trial started with participants fixating a small red point (diameter 0.3°) at the screen center. After 500 ms, two choice-targets appeared to the left and right of the fixation point (10° eccentricity; **Figure 1A**). Each target was shaped as a gradient rectangle (9° length and 0.5° width). After a variable duration of 200 - 500 ms (truncated exponential distribution), the RDM was presented. Participants had to indicate their choice after receiving one or two pulses of 120ms of motion pulses. The gap interval of double-pulse trials was selected randomly from 0, 120, 360, and 1080ms. On single-pulse trials, motion coherence was randomly selected from these six values: 0%, 3.2%, 6.4%, 12.8%, 25.6%, and 51.2%, whereas, on double-pulse trials, motion coherence of each pulse was randomly chosen from three values: 3.2%, 6.4%, and 12.8%. Both pulses had the same net direction of motion and participants were aware of it. In total, there were 6 single-pulse and 9 × 4 double-pulse trial types. After the offset of one or two motion pulses, a 400 to 1000 ms delay period (truncated exponential) was imposed before the Go signal appeared on the screen. In each trial, participants were required to indicate their response by directing the gaze to one of the targets, the upper extreme of targets representing full decision confidence and the lower extreme representing guessing (**Figure 1A**). To provide the approximate balance within the trials, we constructed a list of all possible conditions of motion coherences and gaps. Then, we shuffled the listed conditions and assigned them randomly to the trials in each block. Participants were instructed to achieve high performance. Distinctive auditory feedback (Beep Tones) was provided for correct and incorrect responses. The type of feedback of 0% coherence trials was selected randomly by a uniform distribution. In Experiment 1, each participant performed the task across multiple blocks on different days (12-20 blocks). Experiment 2 contained the same paradigm as Experiment 1. All variables of stimulus remained constant except, in Experiment 2, the EEG data were also recorded. In Experiment 2, each participant completed a session of 4-5 blocks.

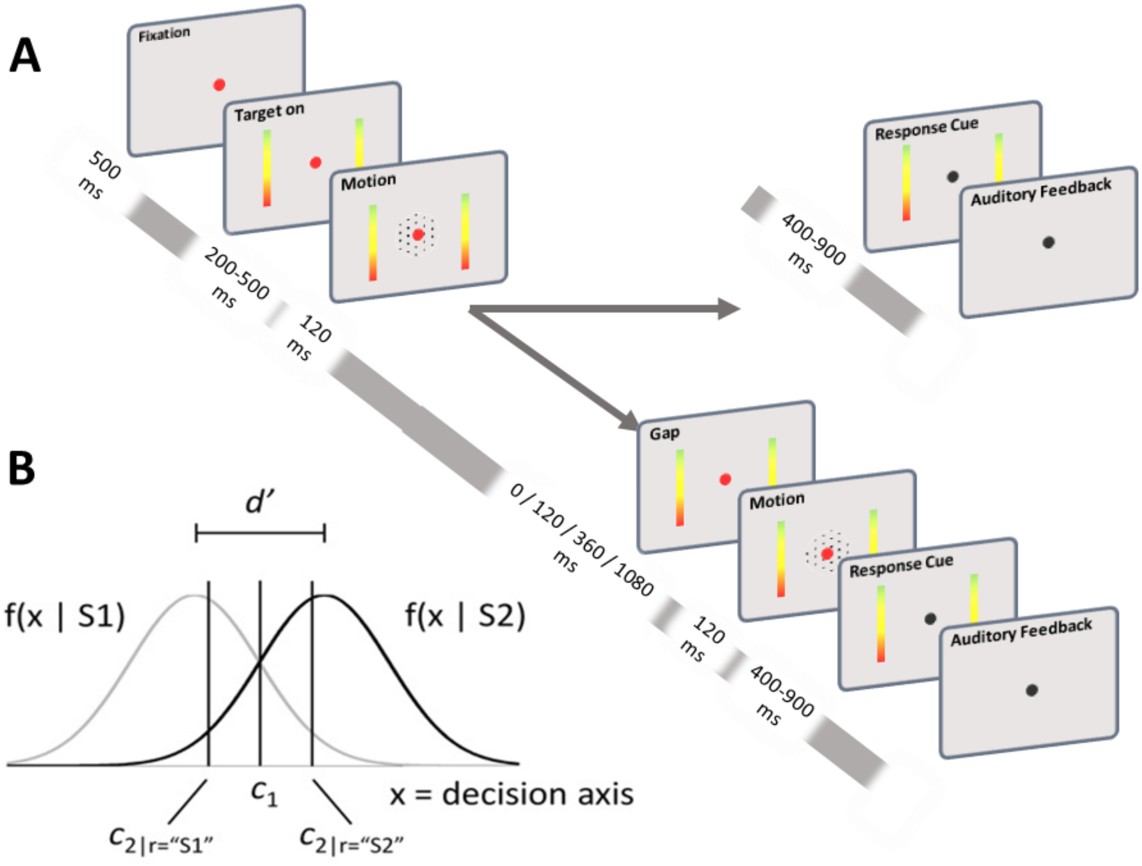


Figure 1. Task paradigm and Signal Detection Theory. **(A)** Participants had to indicate the predominant direction of motion of moving dots (left or right) by saccadic eye movement to one of the targets after receiving one or two pulse(s) of 120ms stimulus. The intervals between two pulses were selected randomly from 0 to 1080 ms and the direction of both pulses were the same. Color-coded targets enabled participants indicating their confidence simultaneously. **(B)** On each trial, a stimulus generates an internal response x within an observer, who must use x to decide whether the stimulus is S_1 or S_2, x is drawn from a normal distribution. The distance between these distributions is d', which measures the observer's ability to discriminate S_1 from S_2. The observer also rates decision confidence on a scale of high and low by comparing x to the additional response specific confidence criteria (cr_2 for each option). For details, see Supplementary Appendix 2 and refs (Fleming, 2017; Maniscalco & Lau, 2012, 2014).

2.4 EEG Recording and pre-processing

We used a 32-channel amplifier for the EEG signal recording (eWave, produced by ScienceBeam, http://www.sciencebeam.com/) which provided 1K sample/s of time resolution. EEG was recorded at

31 scalp sites (Fp1, Fp2, AF3, AF4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, FPz, Fz, Cz, Pz, Oz, POz, FC1, FC2, CP1, CP2, FC5, FC6, CP5, CP6). The EEG signals were referenced to the right mastoid. The recorded data were taken to Matlab (Mathworks Inc., USA) and pre-processed as follows. The signals were filtered using a band-pass filter from 0.1 Hz to 40 Hz (Zizlsperger et al., 2014) for removing high frequency and independent cognitive noises. Then, all trials were inspected, and those containing Electromyography (EMG) or other artifacts were identified and manually removed. The second artifact rejection step included independent components analysis (ICA) using the EEGLAB toolbox (Delorme & Makeig, 2004). To select the removable ICA component, the ADJUST plugin (Mognon, Jovicich, Bruzzone, & Buiatti, 2011) was used.

2.5 Pupillometry Recording and pre-processing

The eye data were collected using an EyeLink 1000 infrared eye-tracker system (SR Research Ltd. Ontaro, Canada). This device allowed a 1000-Hz sampling rate and was controlled by a dedicated host PC. The system was calibrated and validated before each block by presenting nine targets at the center, edges, and corners of the display monitor. The left eye's data was recorded and passed to the host PC via an Ethernet link during data collection.

Missing data and blinks, as detected by the EyeLink software, were padded and interpolated. Additional blinks were spotted using peak detection on the pupil signal's velocity and then linearly interpolated (Mathôt, 2013).

2.6 Experimental procedure

In this study we employed behavioral, neural, and pupillometry signatures. Participants were given a consent form in which the experiment was described in general terms. After providing written informed consent, in both experiments, participants completed the tasks in a semidark, sound-attenuating room to minimize distraction. All instructions were presented and stimuli were displayed on a CRT monitor (17 inches; PF790; refresh rate, 75 Hz; screen resolution, 800 × 600). A head and chin rest confirmed that the distance between the participants' eyes and the monitor's screen was 57 cm throughout the experiment. Participants were presented demographic questions followed by training sessions and main sessions, respectively. The experimental protocol was approved by the ethics committee of the Iran University of Medical Sciences.

2.7 Data Analysis

Data analysis was performed using Matlab 2019a (The MathWorks Inc., United States).

2.7.1 Quantifying confidence

Reported confidence was categorized as high and low. Since the participants were told to choose the upper part of the bar as high confidence and lower part as low confidence, we considered reported confidence higher than midline as high confidence and lower than midline as low confidence respectively. This categorization allowed us to take each confidence report as a binary variable comparable to the choice. Using categorical variables also provided the possibility of comparing the current data with our previous work (Vafaei Shooshtari et al., 2019). However, in addition to the midline, we tested various binary level set methods for categorizing participants' high and low confidence ratings. First, the highest 55% and 45% of each participant's confidence reports were considered high confidence (similar to (Zylberberg, Wolpert, & Shadlen, 2018)). Then, the mean of each participant's confidence was calculated separately, and the confidence ratings above the mean were considered as high ratings. Using these methods did not significantly alter reported confidence categorization (see **Supplementary Figure 6**).

2.7.2 Behavioral analyses

Except where otherwise specified, we reported behavioral data of the first experiment but all the analyses were repeated for the EEG experiment and if the results were inconsistent, it has been admitted (EEG experiment results were reported in **Supplementary Figures 1**, **2**, **3**, **4** and, **5**).

We performed several logistic regression models to measure the impact of stimulus characteristics on binary outcomes after confirming the assumptions of the linear regression were met. For logistic regression models, we used maximum likelihood under a binomial error model (i.e., a GLM) to evaluate the null hypothesis that one or more of the regression coefficients were equal to zero. P_{high} was the probability of high confidence, $Logit[P_{high}]$ indicated log $\frac{P_{high}}{1-P_{high}}$ and β_i denoted fitted coefficients. Also, $P_{correct}$ was the probability of correct response and $Logit[P_{correct}]$ indicated log $\frac{P_{correct}}{1-P_{correct}}$.

For single-pulse trials, the probability of a high confidence choice was given by the following:

$$Logit[P_{high}] = \beta_0 + \beta_1 C, \tag{1}$$

where C was motion strengths of the pulse. Likewise, the probability of a correct choice was stated by the logistic regression:

$$Logit[P_{correct}] = \beta_0 + \beta_1 C, \tag{2}$$

To examine whether confidence judgments were associated with more accurate choices, we fitted a logistic regression model to accuracy where the probability of high confidence is given by:

$$Logit[P_{high}] = \beta_0 + \beta_1 A, \tag{3}$$

where A was the accuracy of the response (0 or 1 for incorrect and correct) and our null hypothesis was that the accuracy would not affect reported confidence ($H_0: \beta_1 = 0$). We also used logistic regression to evaluate the effect of interpulse interval on confidence in double-pulse trials:

$$Logit[P_{high}] = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 T + \beta_4 C_1 T + \beta_5 C_2 T, \tag{4}$$

where C_1 and C_2 were motion strengths of each pulse, and T was the interpulse time interval. For double-pulse trials with equal pulse strength ($C_1 = C_2$), the redundant regression terms (β_2, β_4) were omitted. The null hypothesis was that the interpulse interval would not affect reported confidence ($H_0: \beta_{3-5} = 0$). The similar equation was used to assess relation of accuracy and time interval:

$$Logit[P_{correct}] = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 T + \beta_4 C_1 T + \beta_5 C_2 T, \tag{5}$$

The null hypothesis was that the interpulse interval would not affect performance ($H_0: \beta_{3-5} = 0$). To evaluate the impact of pulse sequence on confidence, the following regression model was fitted:

$$\mathrm{Logit}[P_{high}] = \beta_0 + \beta_1[C_1 + C_2] + \beta_2[C_2 - C_1], \quad (6)$$

where C_1 and C_2 were corresponding motion strengths of each pulse. β_2 indicated how the confidence varied from trials in which $C_1 > C_2$ to trials with a reversed sequence of motion pulses $C_1 < C_2$. The null hypothesis was that the sequence of motion pulses did not influence the confidence ($H_0: \beta_2 = 0$).

To examine the interaction between the two pulses (e.g., a stronger pulse 1 reduced the effect of pulse 2), we fitted the following regression model to all double-pulse trials:

$$\mathrm{Logit}[P_{high}] = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 C_1 C_2, \quad (7)$$

The null hypothesis was that there was not an interaction between motion strengths of pulses ($H_0: \beta_3 = 0$). In other words, higher influence of second pulse on confidence was due to higher sensitivity rather than an interaction of motion pulses and $\beta_2 > \beta_1$ confirmed greater sensitivity to the second pulse on the decision.

In addition to logistic regression models, to investigate the variation of confidence in double-pulse trials compared to single-pulse trials, we subtracted participants' confidence of double-pulse trials from corresponding confidence in single-pulse trials. For example, the confidence of a sequence of 3.2%, 6.4% motion strength trial, subtract separately once from 3.2% and once from 6.4% corresponding confidence in single-pulse trials. The process repeated for the data of each gap too. Moreover, the same method was used to compare accuracy of double-pulse trials and single-pulse trials. To assess the effect of choice accuracy on variation of confidence in double-pulse and single-pulse trials, we fitted the following:

$$S_{Conf} = \beta_0 + \beta_1 A, \quad (8)$$

where the S_{Conf} was the subtraction of confidence in double-pulse trials from corresponding single-pulse trials and A was the accuracy of the response (0 or 1 for incorrect and correct). The null hypothesis was the choice accuracy did not affect the variation of S_{Conf} ($H_0: \beta_1 = 0$).

2.7.2.1 Response-time analysis

In the current study, response-time was referred to the time between the cue onset and a participant's response. To evaluate the significance of the effect of response-time on confidence, we fitted the following linear regression model separately in double-pulse and single-pulse trials:

$$\mathrm{Logit}[P_{high}] = \beta_0 + \beta_1 R, \quad (9)$$

where R was the response-time of each trial and the null hypothesis was that confidence did not depend on the response-time ($H_0: \beta_1 = 0$). Moreover, to evaluate the relation of delay-time imposed before the cue onset and response-time, we fit a linear regression model as follows:

$$\mathrm{RT} = \beta_0 + \beta_1 D, \tag{10}$$

where D was the delay-time. The null hypothesis was that response-time did not rest on the delay-time ($H_0\!: \beta_1 = 0$).

In addition, confidence is tracked by both evidence and response-time (Kiani et al., 2014; van den Berg et al., 2016; Zylberberg et al., 2016), and indeed accuracy is relied on evidence. Furthermore, to study the profile of high and low confidence from behavioral data, an equal number of trials from each participant's trials was selected randomly from single/double-pulse trials. Same procedure repeated 100 times, then individual response-time were rank-ordered and binned into four quintiles. Then, the accuracy of high and low confidence trials in each bin was calculated. We expected to see a significant difference between accuracy of each bin grouped by levels of confidence. We only included motion strength of 3.2, 6.4, 12.8 of single-pulse trials (similar to coherence used in double-pulse trials) to control the impact of coherence on response-time.

2.7.3 Motion energy analysis

Random dot stimulus is stochastic, so the sensory evidence fluctuated within and across trials but around the nominal motion coherence level. To examine the fluctuations in motion during each trial, we filtered the sequence of random by using two pairs of quadrature spatiotemporal filters, as specified in previous studies (Adelson & Bergen, 1985; Kiani, Hanks, & Shadlen, 2008; Zylberberg et al., 2012). Since we aimed to understand the temporal course of choice and confidence, we summed the energies across trials for each pulse in single/double-pulse trials.

We used logistic regression to test whether the confidence was more influenced by the second pulse's motion energy than that of the first pulse in double-pulse trials. We tested double-pulse trials with equal motion strength using the following logistic regression model:

$$\mathrm{Logit}\left[P_{high}\right] = \beta_0 + \beta_1 \mathrm{C} + \beta_2 (M_1 + M_2) + \beta_3 M_2, \tag{11}$$

where M_1 and M_2 were the motion energy of each pulse. The null hypothesis was that the second pulse was not more functional ($H_0\!: \beta_3 = 0$). We tested double-pulse trials with unequal motion strength by modifying the regression model to:

$$\mathrm{Logit}\left[P_{high}\right] = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 (M_1 + M_2) + \beta_4 M_2\,, \tag{12}$$

and the null hypothesis was ($H_0\!: \beta_4 = 0$). To evaluate the relation of P_{high} and motion energy in single-pulse trials, we fitted a linear regression model as follows:

$$\mathrm{Logit}\left[P_{high}\right] = \beta_0 + \beta_1 \mathrm{C} + \beta_2 M, \tag{13}$$

where M was the motion energy of the presented motion stimulus and the null hypothesis was that confidence did not depend on the motion energy ($H_0\!: \beta_2 = 0$).

2.7.4 General computational modeling approach

We implemented a set of computational models based on signal detection theory to provide a mechanistic explanation of the experimental data. According to SDT, observers set a decision criterion (cr) to discriminate between two stimuli (e.g., labeled as S_1 and S_2). They also set criteria $cr_{2,``S1"}$ and $cr_{2,``S2"}$ to determine confidence ratings around the decision criterion cr (**Figure 1B**; for more details, see Supplementary Appendix 2). We computed stimulus sensitivity (d') and measures of metacognitive ability ($Meta\text{-}\ d'$, $Meta\text{-}\ d'/d'$). We used code provided by Maniscalco and Lau (Maniscalco & Lau, 2012) in which metacognitive sensitivity ($Meta\text{-}\ d'$) is computed by setting the d' value that would produce the observed confidence. In addition, $Meta\text{-}\ d'/d'$ were calculated by normalizing $Meta\text{-}\ d'$ by d' through division. Here, d', $Meta\text{-}\ d'$ and, $Meta\text{-}\ d'/d'$ of single-pulse and double-pulse trials were computed separately. In addition, we fitted SDT model with trials simulated by a perfect integrator model (the model is described later). We then addressed the trend of d', $Meta\text{-}\ d'$ and, $Meta\text{-}\ d'/d'$ of three models for each participant. To support the fact that our findings were not relevant to variation of coherence of single and double-pulse trials, we only included single-pulse trials with motion strength of 3.2, 6.4, 12.8. However, one difference between groups was that they might not be matched for the number of trials: the single-pulse included on average fewer trials for each coherence per participant compared to double-pulse trials. Previous research has suggested that the number of trials could bias measures of metacognitive ability (Fleming, 2017). Therefore, in a control analysis, we created 100 sets of trials randomly from the single/double-pulse trials and from trials simulated by the perfect integrator model. Each set contained the same number of trials for each participant. We then averaged the metacognitive scores obtained from these 100 sets and repeated the comparison procedure (see **Supplementary Figure 6**).

2.7.5 Perfect integrator Model

To estimate the expected high confidence ($P_{e(high)}$) in double-pulses trials, we assumed that each trial's confidence was achieved based on evidence integrating from both pulses by using a perfect integrator model. In the perfect integrator model, the expected accuracy ($P_{e(correct)}$) for double-pulse trials computed as the following (Kiani et al., 2013):

$$P_{e(correct)} = 1 - \phi(0, e_1 + e_2, \sqrt{2}), \quad (14)$$

where e_1 and e_2 were the pieces of evidence that underlie by P_1 and P_2 (the probabilities of the correct answer in corresponding single-pulse trials) and were computed as:

$$e_i = \phi^{-1}(P_i, 0, 1), \mathrm{I} = 1{,}2, \quad (15)$$

Where ϕ^{-1} was inverse ϕ, which represented the cumulative Gaussian distribution (Kiani et al., 2013).

To predict the confidence of double-pulse trials by this model, after calculating cr and d' (see Supplementary Appendix 2), cr was shifted to zero and d' was normalized. Then, confidence Hit Rate and False Alarm Rate were calculated based on confidence performance from corresponding single-pulse trials (similar to Eq.14, 15). Accordingly, high confidence probability (for both correct response or incorrect response) would be predicted by the perfect integrator model. Besides, the model parameters, including confidence criteria along with $Meta\text{-}\ d'$ were computed.

2.7.6 Confidence optimized model

In the confidence optimized model, we optimized the confidence criteria in the perfect integrator model by providing each participant's confidence performance computed of double-pulse trials. The purpose of this simulation was to understand why the perfect integrator model was not able to predict confidence well.

2.7.7 Model evaluation

We evaluated the models qualitatively (i.e., parameter recovery exercises) and quantitatively (i.e., maximum likelihood estimation).

In the qualitative method, based on the calculated parameters of the model, the probability of choosing high confidence for all combinations of motion strength for each participant were calculated (see Supplementary Appendix 2). We compared the expected high confidence predicted by models to the observed confidence in double-pulse trials using regression, as follows:

$$P_{high} = \beta_1 P_{e(high)} + \beta_0, \qquad (16)$$

where $P_{e(high)}$ was the expected probability of high confidence. We regressed predicted vs. observed P_{high} and compare slope (β_1) against the 1:1 line in each model. In this linear regression, we expected the predicted values to be close to the actual values.

In addition, to compare models quantitatively, an equal number of trials from each subject's trials selected randomly and then each model fitted to the selected data. This procedure repeated for 100 times, then the computed MLEs of each model was averaged.

2.7.8 Confidence suboptimality

The optimal decision-making is disrupted by several sources of suboptimality (Balsdon et al., 2020). In SDT, an added noise, ξ_n, represents a potential loss of information between sensory decision information and metacognitive information, such as confidence rating. This noise has a Gaussian distribution with zero mean, and standard deviation σ (Maniscalco & Lau, 2014). The parameter σ determines how much noisier the metacognitive variable is than the decision variable (Maniscalco & Lau, 2014).

$$\xi_n = N\ (0, \sigma)) \qquad (17)$$

This noise is correlated to metacognitive efficiency ($Meta\text{-}d'/d'$) (Maniscalco & Lau, 2014). To consider this suboptimality, we simulated trials using the same parameter values resulted from the perfect integrator model except this noise was increased.

2.7.9 EEG analysis

The EEG analysis focused on a neural marker of perceptual decision-making linked with stimulus preparation and stimulus processing. The component we focused on was the centro-parietal positivity (CPP) which possibly identical to the classic P300 component (Herding et al., 2019; Twomey, Murphy, Kelly, & O'connell, 2015). The CPP is associated with the sampling of available evidence in perceptual decisions and confidence rating at time period of 200-500 ms after stimulus onset (Herding et al., 2019; Rausch, Zehetleitner, Steinhauser, & Maier, 2020; Vafaei Shooshtari et al., 2019; Zizlsperger et al.,

2014) or at the time of the response (Boldt et al., 2019). Here, CPP amplitude was measured as the mean amplitude in a time-window ranging from 200 ms to 500 ms after stimulus onset in an electrode cluster containing the electrodes CP1, CP2, Cz, and Pz (Boldt et al., 2019; Herding et al., 2019; Rausch et al., 2020; Tagliabue et al., 2019; Twomey, Kelly, & O'Connell, 2016; Vafaei Shooshtari et al., 2019; Zizlsperger et al., 2014). We epoched the EEG responses were aligned with respect to the stimulus onset, from 200 ms pre-stimulus to 500 ms post-stimulus of each pulse. Then, these epochs were baselined to a window -100 ms to stimulus-locked to prevent differences in the visual response to the stimulus affecting the baseline. The ERP signals were examined for levels of confidence separately in double-pulse trials and single-pulse trials. We analyzed correct trials of each coherence level distinctly to support the fact that our findings were not relevant to participants' performance and motion pulse strength. Also, in double-pulse trials, we tested double-pulse trials with non-zero gaps and equal motion strength pulses.

2.7.10 Pupillometry analysis

Previous work showed that pupil dilation after choice and before feedback reflected decision uncertainty (Colizoli, De Gee, Urai, & Donner, 2018; Urai et al., 2017). Accordingly, as the confidence is uncertainty complement (Hebart, Schriever, Donner, & Haynes, 2014; Kepecs & Mainen, 2012), to study the confidence profile, the method was implemented here. The mean baseline-corrected pupil signal throughout 200 ms before feedback was calculated as our single-trial measure of pupil response. We epoched trials and baselined each trial by subtracting the mean pupil diameter 50 ms before the response. We included all trials of both experiments in the analyses reported in this paper.

According to the temporal low-pass characteristics of the slow peripheral pupil apparatus (Hoeks & Ellenbroek, 1993), trial-to-trial variations in response-time can impact trial-to-trial pupil responses, even in the absence of amplitude variations in the underlying neural responses (Urai et al., 2017). To isolate trial-to-trial variations in the amplitude (not duration) of the underlying neural responses, we removed components explained by response-time via linear regression:

$$y' = y - (y^T R)R, \tag{18}$$

where y was the original vector of pupil responses, R was the vector of the corresponding response-time (log-transformed and normalized to a unit vector), and T indicated matrix transpose. Consequently, after removing the variance explained by trial-by-trial response-time, the residual y' reflected pupil responses. This residual pupil response was used for analyses reported in this study. To evaluate the relation of confidence and pupil response, we fit a linear regression model as follows:

$$\text{Logit}[P_{high}] = \beta_0 + \beta_1 P, \tag{19}$$

where the P was pupil response in each trial. The null hypothesis was that confidence did not change with the pupil response ($H_0\colon \beta_1 = 0$). To control the impact of coherence on pupil response, we only included motion strength of 3.2, 6.4, 12.8 of single-pulse trials.

2.7.11 General statistical analysis

We used repeated-measures two-tailed *t*-tests. As suggested, we considered small (d = .2), medium (d = .5), and large (d = .8) effect sizes for this assessment (see (Cohen, 1970)) and the statistical significance for *t*-tests was set to a probability from data ≥ .90.

Moreover, to test our hypotheses, a series of regression analyses were run after confirming the assumptions of the linear regression are met. Effect sizes were reported and as suggested, here, we considered small (f^2 = .02), medium (f^2 = .15), and large (f^2 = .35) effect sizes (see (Cohen, 1970)) at the alpha level of 5%.

For tests of pupil response signals and ERPs between two levels of confidence, statistical inferences were performed using *t*-tests at each time-point (at a statistical threshold of $p < .05$).

3 Results

We tested our predictions in two studies that applied the same paradigm (**Figure 1A**). The first study used behavioral measures and pupillometry analyses, whereas for the second experiment, we recorded EEG signals as well. Participants decided about the direction of the RDM motion based on brief motion pulses. The task design contained different conditions which allowed us to compare participants' behavior in (i) double-pulse vs single-pulse, (ii) different coherence of motion stimulus, and (iii) four distinct gaps intervals.

3.1 Behavioral results

We used the single-pulse trials to benchmark the effect of coherence on choice accuracy and confidence. As shown in **Figure 2A**, for single-pulse trials, participants were more confident for high coherence stimuli (**Figure 2A**; Eq.1; β_1 = .06, $p < .001$, 95% CI = [.04, .08], f^2 = .23), ranged from .96 for 51.2 coherence to .43 for 3.2 coherence. Also, accuracy improved with motion strength reached from .56 for 3.2% to .99 for 51.2% (**Figure 2B**, black line; Eq.2; β_1 = .10, $p < .001$, 95% CI = [.08, .12], f^2 = .36). They also had better performance whenever they had reported higher confidence comparing to lower confidence (**Figure 2B**, red and green; Eq.3; β_1 = 1.1, $p < .001$, 95% CI = [.88, 1.31], f^2 = .09). Moreover, in double-pulse trials, the accuracy improved with motion strength (**Figure 2C**, black dots) and participants were more accurate while reporting higher confidence (**Figure 2C**, green dots). Along with accuracy (**Figure 2D**; Eq.5; $p > .1$, (Kiani et al., 2013; Tohidi-Moghaddam et al., 2019), see **Supplementary Figure 1A** for Experiment 2 data), the confidence was largely unaffected by interpulse interval in both double-pulse trials with equal pulse strength (**Figure 2E**; Eq.4; $p > .1$; **Supplementary Table 2**, results of individual participants) and those with unequal pulse strength (**Figure 2E**; Eq.4; $p > .1$; **Supplementary Table 2**; see **Supplementary Figure 1B** for Experiment 2 data). The two pulses separated by up to 1 s supported a level of confidence that was indistinguishable from a pair of pulses separated by no gap.

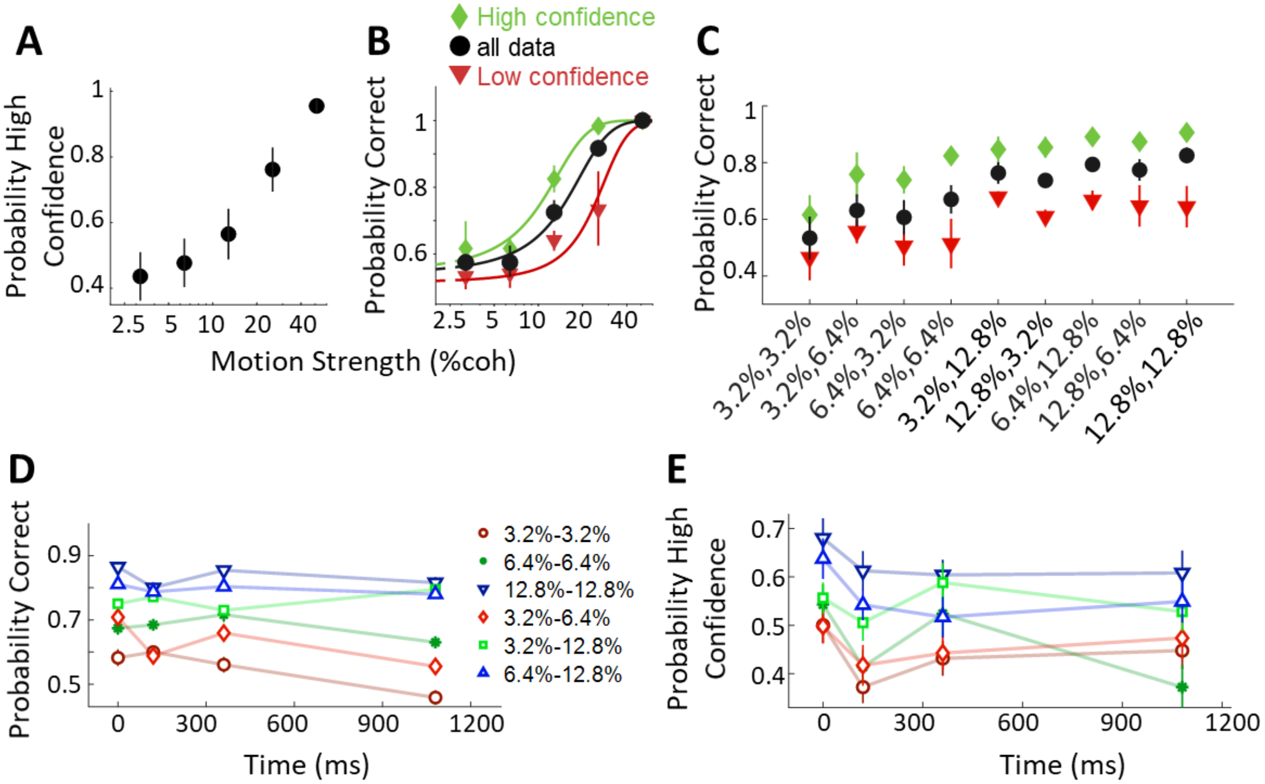


Figure 2. Interplay between confidence, accuracy, and coherence in single/double-pulse trials, and interpulse interval in double-pulse trials. (A) Probability of high confidence as a function of motion coherence. **(B) (C)** Accuracy in single-pulse trials and double-pulse trials in all trials (black), split by high (green) and low (red) confidence decisions. In (B) curves are model fits. **(D)** Choice accuracy for double-pulse trials grouping in all possible interval conditions. **(E)** Confidence of double-pulse trials was calculated by pooling data across all time intervals. In (D) and (E) each data point reports pooled data from indicated sequence pulse and its reverse order (e.g., 12.8– 3.2% and 3.2 –12.8%).

Direct comparison between single-pulse and double-pulse trials, along with previous studies (Kiani et al., 2013; Tohidi-Moghaddam et al., 2019), showed that participants' accuracy significantly differed (t(11490) = -3.09, p < .05, 95% CI = [-.08, -.02], Cohen's d = .11). However, in double-pulse trials participants were not more confident comparing to single-pulse trials (t(11490) = 1.35, p = .18, 95% CI = [-.01, .06], Cohen's d = -.05).

Although, the order of the pulses affected accuracy (**Figure 3A**, (Kiani et al., 2013; Tohidi-Moghaddam et al., 2019)), participants were not more confident in double-pulse trials with unequal pulse strength where the stronger motion appeared in a second order (**Figure 3B**; Eq. 6; β_2 = .01, p = .08, 95% CI = [.00, .02]], f^2 = .02; see **Supplementary Figure 2B** for Experiment 2 data). Also, the increased confidence was not because of an interaction of motion pulses (Eq. 7; β_3 = -.01, p = .13, 95% CI = [-.02, .00], f^2 = .03).

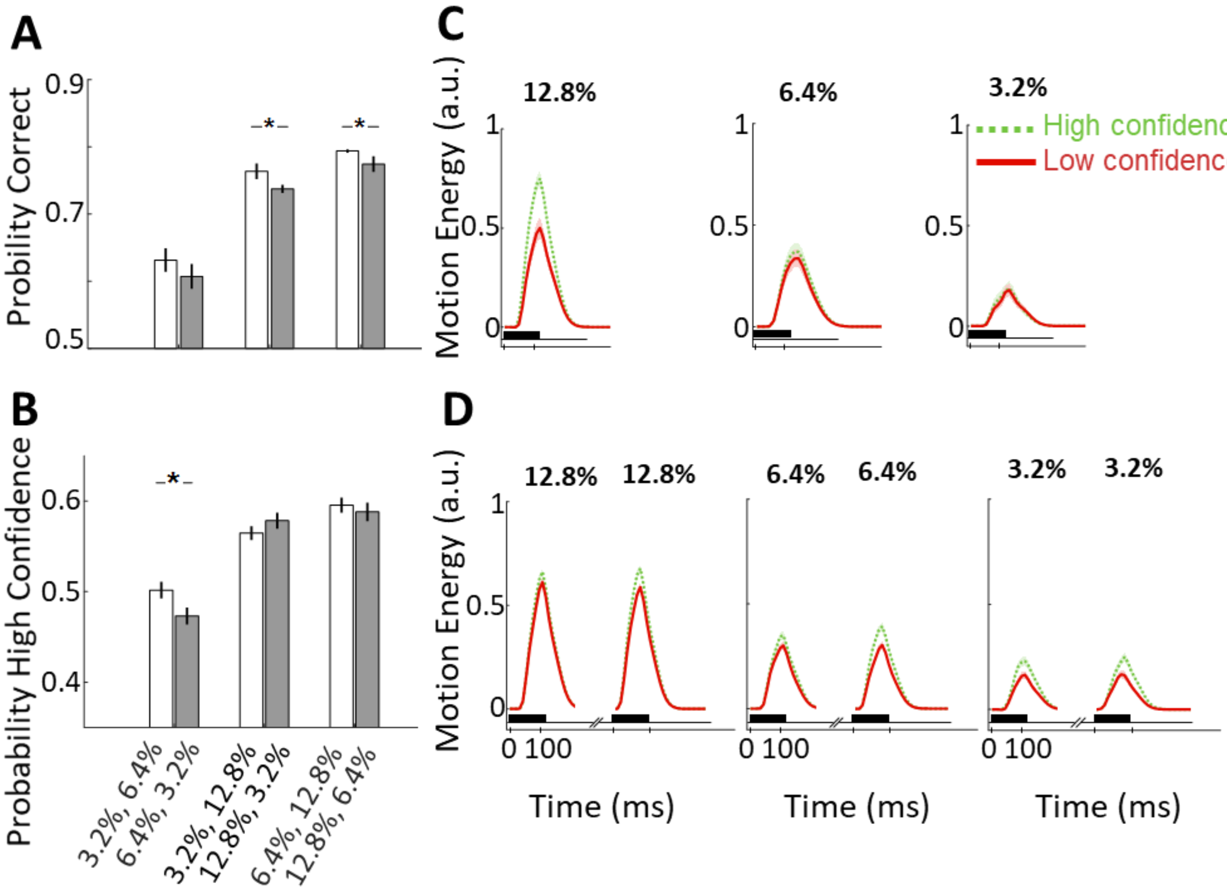


Figure 3. Choice confidence was not depended of the sequence of motion pulses (A) The weak–strong pulse sequence contributed higher accuracy than the strong–weak sequence. **(B)** The weak–strong pulse sequence did not contribute higher confidence than the strong–weak sequence. In all panels, data are represented as group mean ± SEM. (*p<0.05) **(C)** In single-pulse trials, low and high confidence cannot be determined by motion energy profiles in weaker pulses **(D)** The second pulse had slightly more impact on confidence. Data were pooled for all nonzero interpulse intervals. Only correct trials with equal pulse strength are included. In (C) and (D), the shaded region around the mean indicates SEM. The black horizontal bars show the duration of the stimulus display. The units of motion energy are arbitrary and the same for all motion strengths.

3.2 Motion energy results

To yield a precise estimate of the decision-relevant sensory evidence accommodated in the stochastic stimuli, we employed motion energy filtering to the random dot motion stimuli. **Figure 3D** displays the average motion energy in double-pulse trials when the strength of pulses was the same. Accordingly, the difference of the motion energy profiles for high and low confidence responses was slightly larger for the second pulse than the first pulse. A logistic regression confirmed the influence of trial-to-trial fluctuations of motion energy on confidence (Eq.11; β_2 = .10, p = .001, 95% CI = [.04, .16], f^2 = .19). Also, there was slightly larger impact of motion energy of the second pulse with equal pulse strength (Eq.11; β_3 = .11, p = .04, 95% CI = [.07, .15], f^2 = .13). On the contrary, the impact of motion energy of the second pulse was not significant (Eq.12, β_4 = .10, p = .06, 95% CI = [.06, .14], f^2

= .08). Consequently, motion energy analysis could not provide independent confirmation of asymmetric effect of both pulses for confidence.

As well, in single-pulse trials, the difference of the motion energy profiles for high and low confidence with stronger pulse strength (12.8%, 6.4%) was significant (**Figure 3C**; Eq.13; $\beta_2 = .41$, $p = 2.25 \times 10^{-5}$, 95% CI = [.23, .59], $f^2 = .24$). However, the difference in weak motion pulse was insignificant (**Figure 3D**; Eq.13; $\beta_2 = .17$, $p = .44$, 95% CI = [-.26, .60], $f^2 = .10$). Thus, motion energy analysis thus suggests that when the pulses' motion strengths are weak, the subjects decide about their confidence almost randomly.

3.3 The Interplay between confidence in single vs double-pulse trials

To address accuracy and confidence variation in double-pulse from single-pulse trials, we consider $P_{correct}$ or P_{high} of each coherence (3.2%, 6.4% and 12.8%) in single-pulse as baseline and measure the $P_{correct}$ or P_{high} variation of any corresponding sequence in double-pulse trials. As we expected in all combinations of three coherence as the baseline, $P_{correct}$ improved (**Figure 4A**). Additionally, when considering all the trials, in all combinations of three coherence as the baseline, P_{high} increased when the other pulse was a strong pulse (12.8%) (**Figure 4B** and **Figure 4C** for correct trials). On the contrary, P_{high} decreased or not changed considerably whenever the other pulse was a weak motion strength (3.2%, 6.4%). Interestingly, in incorrect trials, the confidence decreased comparing to single-pulse for all the coherence and conditions (**Figure 4D**). These data did not correlate with the interval duration (**Figure 4A**, **B**, **C**, **D**).

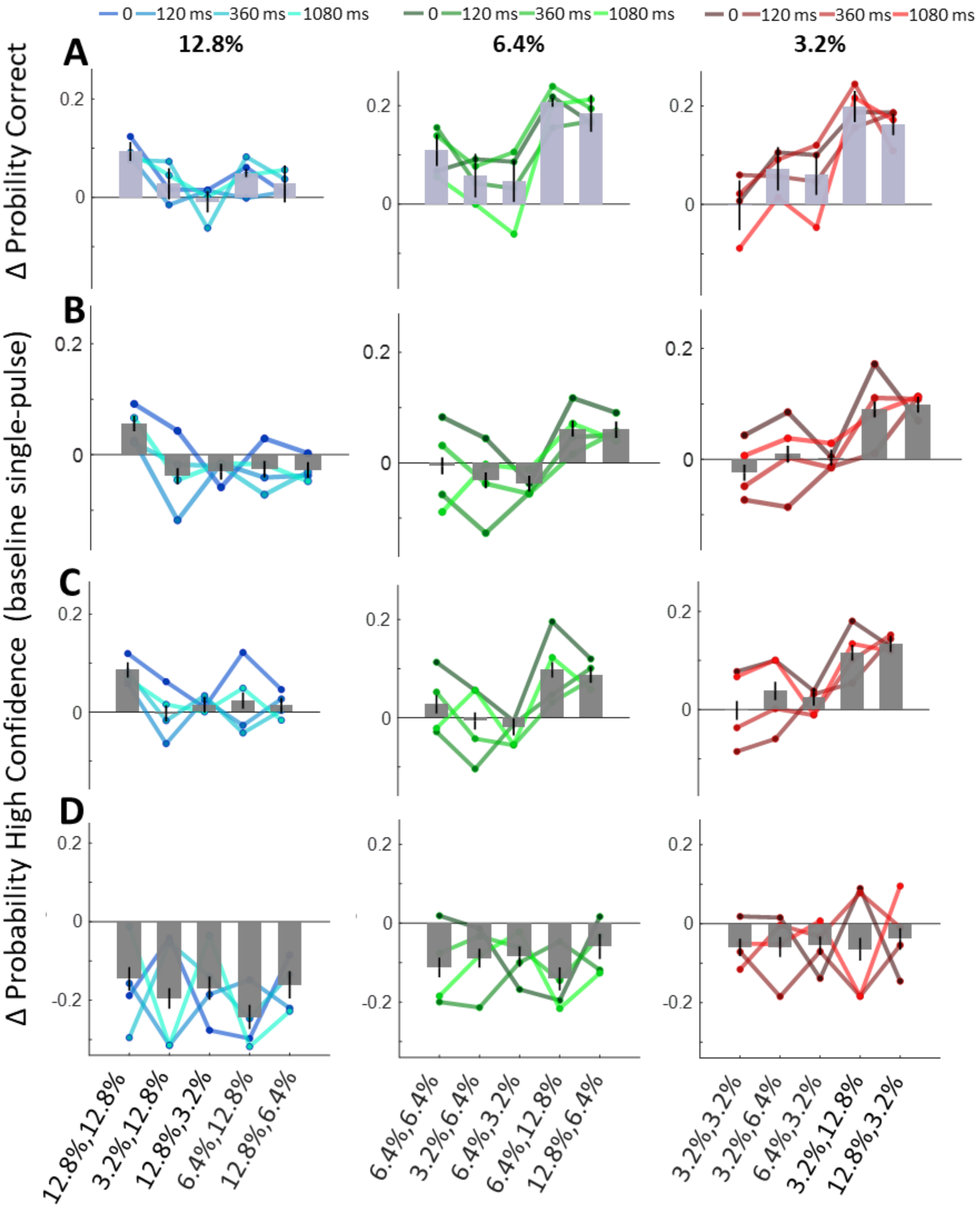


Figure 4. Variation of accuracy or confidence in double-pulse trials baselined by corresponding coherence (3.2%, 6.4% and 12.8% for each column). (A) Considering all the trials, the accuracy improved in almost all pulses combination. **(B)** Considering all the trials, the confidence improved in combination with stronger pulses while the confidence in sequence with a weaker pulse either decreased or remained constant. **(C)** In correct-choice trials, the increasing effect of stronger pulses is more significant and the confidence even slightly improved in combination with weaker pulses comparing to corresponding baseline. **(D)** Interestingly, in incorrect trials, the confidence decreased in every condition. The colored line representing matching data for each of four possible gaps. The data are represented as group mean ± SEM.

In other words, the participants reported lower confidence in double-pulse trials compared to single-pulse trials for incorrect choices but reported higher confidence for correct choices (**Figure 4** Eq.8, β_1 = .15, *p* < .001, 95% CI = [.13, .17], f^2 = .29; **Supplementary Figure 3** for Experiment 2; **Supplementary Table** 1, results of individual participants). This data is in line with the fact that the good metacognitive sensitivity will provide higher confidence for correct responses, and lower for incorrect ones.

3.4 Computational models

The accuracy in double-pulse trials surpasses the expectation measured by the perfect integrator (**Figure 5A**; (Kiani et al., 2013)). Considering the strong positive relation of accuracy and confidence (Kiani et al., 2014; Vafaei Shooshtari et al., 2019), we expected the observed confidence would exceed the predicted confidence (Eq.16) calculated by the perfect integrator model (Eq.14), but it did not (**Figure 5B**).

According to SDT models, d' is stimulus sensitivity and has relation to task performance. As the d' of the perfect integrator model was calculated based on single-pulse trials performance, if participants' performance in single-pulse trials failed, their performance prediction missed the double-pulse trials (**Figure 5C**).

$Meta\text{-}d'$ in all participants increased in double-pulse trials but perfect integrator model failed to imitate the increasing (**Figure 5C**). We also computed metacognitive efficiency ($Meta\text{-}d'/d'$), as another index of the ability to discriminate between correct and incorrect trials. Here, $Meta\text{-}d'/d'$ in all participants missed to track their $Meta\text{-}d'/d'$ in double-pulse trials. Altogether, the perfect integrator was incapable of employment observed metacognitive ability in double-pulse trials. The same modeling procedure of data from EEG experiment has provided similar results (**Supplementary Figure 4**).

As a control investigation, we examined whether the differences in estimated metacognitive ability between models could result from the different number of trials. We averaged the metacognitive scores obtained from equal numbers of samples, and found very similar results. Thus, the difference in the estimated metacognitive efficiency cannot be explained by the difference in the number of trials between the single-pulse, double-pulse, and perfect integrator models (**Supplementary Figure 7**).

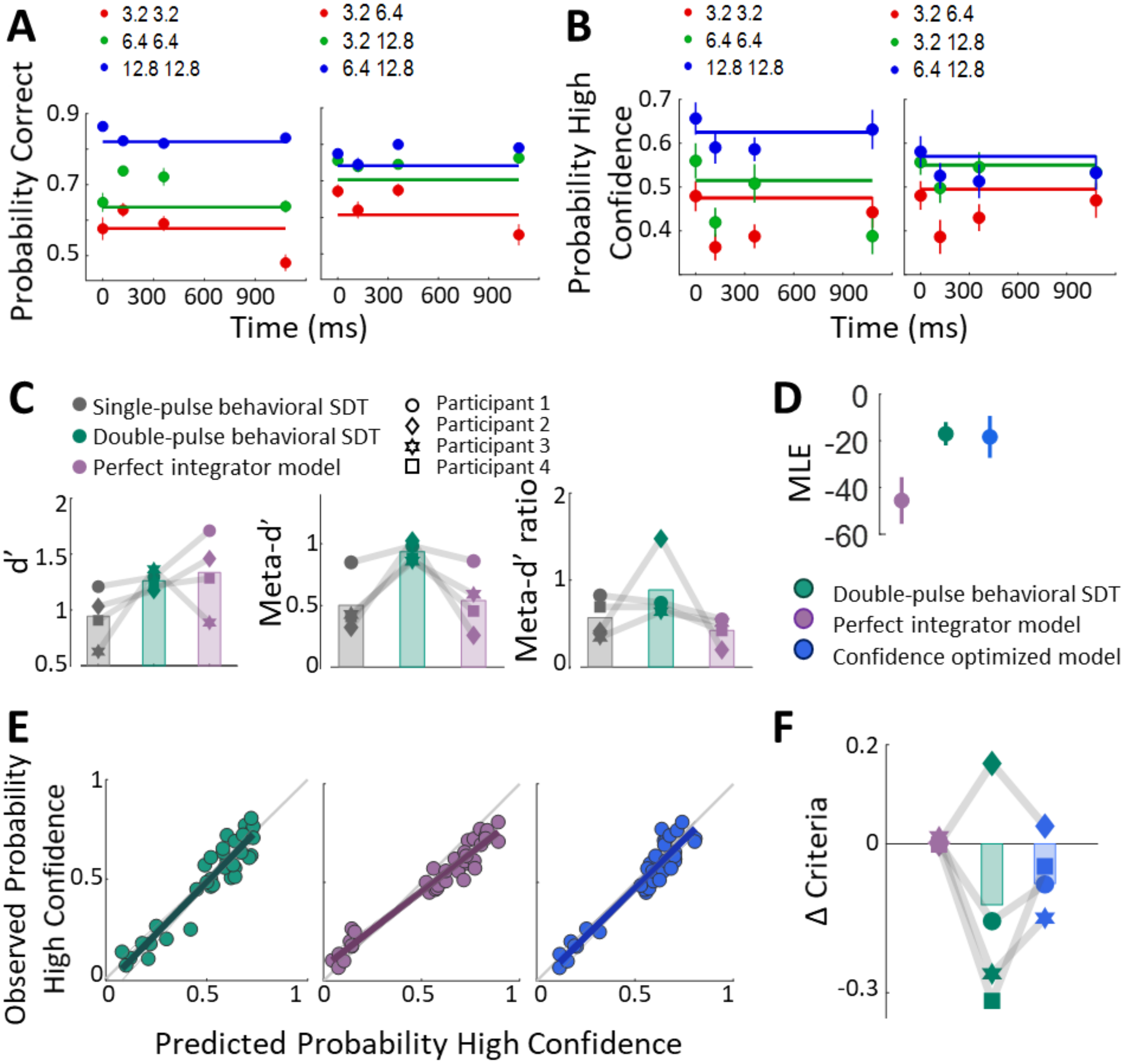


Figure 5. Comparison of the models and human behavior. (A) Accuracy in double-pulse trials. Horizontal lines show accuracy prediction by the perfect integrator model. **(B)** Confidence in double-pulse trials. Horizontal lines show confidence prediction by the perfect integrator model. In (A) and (B) Each data point represents pooled data from the pulse sequence indicated by the legend and its reverse order. **(C)** Stimulus sensitivity (d'), metacognitive sensitivity ($Meta\text{-}d'$), metacognitive efficiency ($Meta\text{-}d'/d'$) estimated for single-pulse trials, double-pulse trials and perfect integrator models for each participant. **(D)** Model comparison suggests strong evidence in favor of the confidence optimized model over the perfect integrator **(E)** Relation of predicted confidence and observed data. SDT model fitted to double-pulse trials (green), the perfect integrator model (purple), and optimized model (blue). Colored lines indicate best-fitting slope of a linear regression analysis. Each data point represents pooled data from different sequence of pulses of each participant. **(F)** Variation of confidence criteria comparing to single-pulse trials in perfect integrator vs optimized model. For panels A, B and, D, data are represented as group mean ± SEM.

512

As in the perfect integrator model, the β_1 (slope in Eq.16) differed from 1:1 line and confidence prediction failed to account for behavioral data (**Figure 5E**; Eq.16, β_1 = .77, p < .001, 95% CI = [.71, .83], f^2 = 15.66), we introduced a model in which the metacognitive sensitivity ($Meta\text{-}d'$) calculated in the perfect integrator model was optimized. The stimulus parameter (d') remained constant whereas the placement of confidence criteria was optimized to fit best to observed data. So, the predicted P_{high} improved intensely (**Figure 5E**, Eq.16, β_1 = .98, p < .001, 95% CI = [.88, 1.08], f^2 = 10.11). Additionally, we take the confidence criteria of the single-pulse model as the baseline and measure the variation of criteria of the perfect integrator and the optimized model. This variation in the optimized model has changed comparing to the perfect integrator (**Figure 5F**, and **Supplementary Figure 5** for EEG experiment). Failure to predict the proper change in confidence criteria in the perfect integrator model was the factor that made the model unable to estimate the confidence from single-pulse trials.

In addition, to consider the suboptimality in confidence reporting, we simulated data using the perfect integrator model's parameters while setting higher confidence noise (Eq.15). The predicted P_{high} from this simulation improved (Eq.16, β_1 = .97, p < .001, 95% CI = [.83, 1.07], f^2 = 9.00). Consequently, the perfect integrator model simply highlighted accumulating decision evidence and ignored the effect of confidence noise.

3.4.1 Models' evaluation

We conducted parameter recovery simulations to evaluate models fitted to single/double-pulse trials. We regressed predicted vs. observed P_{high} confidence for each coherence of each participant. In single-pulse trials, linear regression indicated that there was a significant effect between the predicted and observed P_{high}, (Eq.16, β_1 = 1.04, p < .001, 95% CI = [.90, 1.18], f^2 = 8.09). In double-pulse trials, regression coefficient was statistically significant and close to 1:1 line (**Figure 5E**; Eq.16, β_1 = 1.03, p < .001, 95% CI = [.91, 1.15], f^2 = 11.05) meaning predicted P_{high} by classic SDT also explained a significant proportion of variance in the observed P_{high}.

A quantitative model comparison unsurprisingly favored the optimized model (mean MLE = -18.31) and the SDT behavioral model (mean MLE = -16.98) over the perfect integrator model (mean MLE = -45.53) (**Figure 5D**).

In summary, comparing between the models, both quantitatively (**Figure 5E**) and qualitatively (**Figure 5D**) in doube-pulse trials, also showed that the confidence optimized model has a better prediction in estimating confidence. Accordingly, these investigations indicated: (i) participants integrated the decision evidence perfectly but to report their confidence, their confidence resolution improved rather than reporting higher confidence, (ii) the inability to predict the proper change in confidence criteria in the perfect integrator model was the factor that made the model unable to estimate the confidence from single-pulse trials, (iii) the confidence noise was changed after receiving the second pulse in double-pulse trials.

3.5 Response-time analysis

Response-time had a significant effect on confidence in double-pulse (Eq.9, β_1 = .09, p = .04, 95% CI = [0.01, 0.17], f^2 = .10) but not in single-pulse trials (Eq.9, β_1 = - .02, p = .90, 95% CI = [-1.78, 1.74], f^2 = .00). Moreover, the confidence profile as a function of response-time was significant in double-pulse trials (**Figure 6A; Table 1**) but not in single-pulse trials (**Figure 6B; Table 1**).

Table 1. Result from t-tests to compare confidence profile in single/double-pulse trials in each response-time bin.

Trial type	**Double-pulse**				**Single-pulse**			
RT bin	1	2	3	4	1	2	3	4
tstat	8.89	6.77	9.20	8.10	1.00	2.71	1.46	1.69
df	2652	2652	2652	2652	2150	2150	2150	2152
p	$.11 \times 10^{18}$	$.16 \times 10^{11}$	$.71 \times 10^{20}$	$.77 \times 10^{16}$	.31	.007	.14	.09
95% CI	[-.17, -.11]	[-.15, -.08]	[-.19, -.12]	[-.19, -.12]	[-.18, .06]	[-.30, -.05]	[-.22, .03]	[-.24, .02]
Cohen's *d*	.35	.26	.36	.32	.14	.37	.20	.23

Additionally, in our double-pulse trials, participants decided faster than single-pulse trials in all interval durations (**Figure 6C**). We regress the delay-time before cue onset (0.4 to 1 s truncated exponential) and response-time in both single-pulse and double-pulse trials to examine the effect of imposed delay time on response-time. The effect was small in both single-pulse (Eq.10, $\beta_1 = -.01 \times 10^{-5}$, $p = .004$, 95% CI = [0.00, 0.00], $f^2 = .00$) and double-pulse trials (Eq.10, $\beta_1 = -.0003 \times 10^{-6}$, $p < .001$, 95% CI = [0.00, 0.00], $f^2 = .01$).

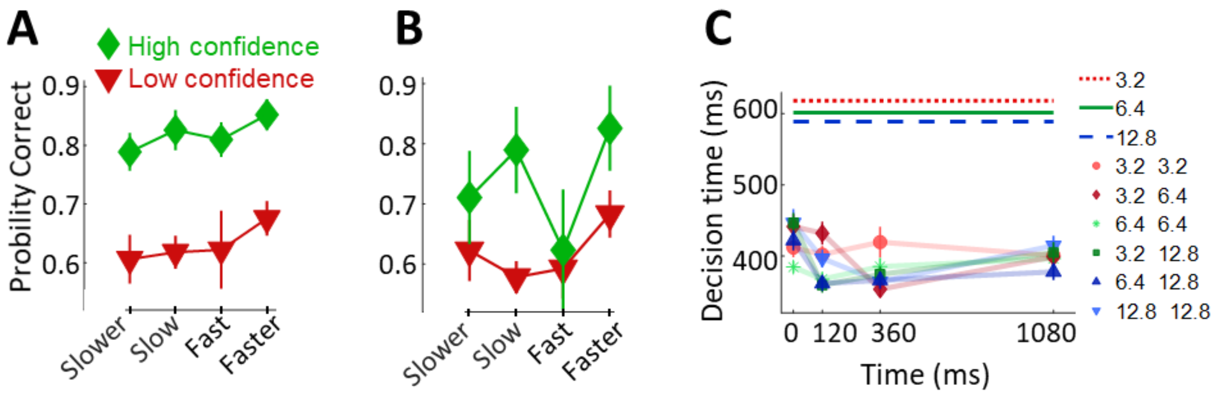


Figure 6. Response-time profiles in single and double-pulse trials. (A) (B) Accuracy as a function of response-time split by high (green) and low (red) confidence in double-pulse trials (A) and single-pulse trials (B). **(C)** Response-time of all coherence combination clustered by gap interval in double-pulse trials (dots) comparing to single-pulse trials (lines). Data are represented as group mean ± SEM.

3.6 EEG Analysis

We derived the ERPs of averaged signals for two levels of confidence to verify whether there was a significant difference in the centro-parietal ERPs across confidence levels. **Figure 7** exhibit ERPs and scalp topographies for confidence levels time-locked to the stimulus onset in low and high confidence in single-pulse trials.

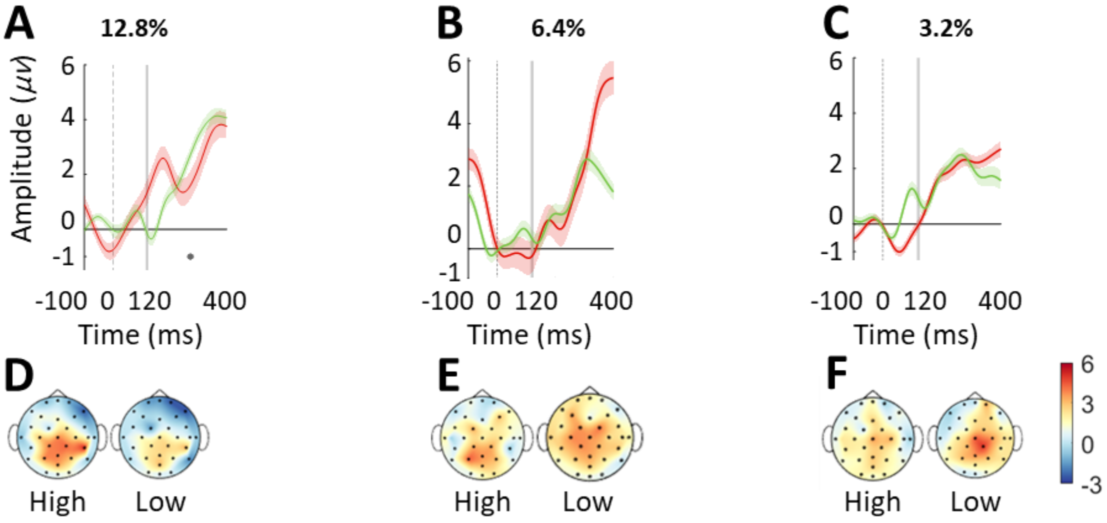


Figure 7. ERPs and scalp topographies in single-pulse trials. (A) (B) (C) ERPs of correct single-pulse trials shows an insignificant difference in weaker motion strength in high and low confidence level trials. **(D) (E) (F)** Scalp topographies in two levels of confidence (the mean amplitude in a time-window ranging from 200 ms to 500 ms after stimulus onset). The shading region around the mean indicates SEM. * indicate p<.05 from a *t*-test, of the difference between the two-time.

569

570 **Figure 8** exhibit ERPs and scalp topographies for confidence levels time-locked to the stimulus onset
571 in low and high confidence in double-pulse trials.

572

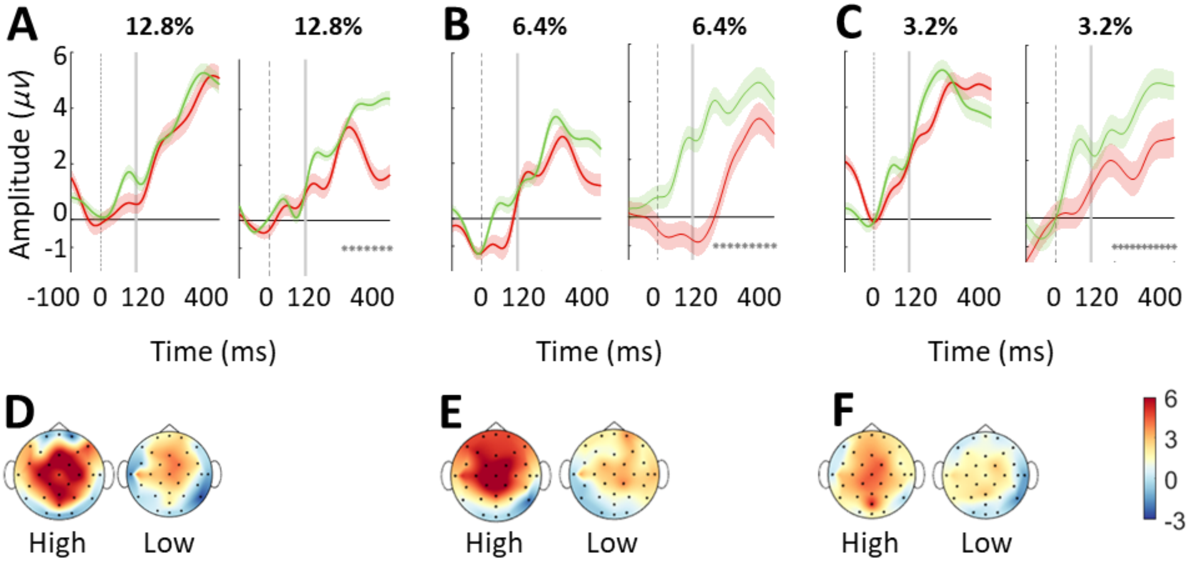


Figure 8. ERPs and scalp topographies in double-pulse trials. (A) (B) (C) ERPs in the two levels of confidence are distinct after the stimulus onset. **(D) (E) (F)** Scalp topographies in two levels of confidence (the mean amplitude in a time-window ranging from 200 ms to 500 ms after stimulus

onset of second pulse). The shading region around the mean indicates SEM. * indicate p<.05 from a *t*-test, of the difference between the two-time.

Interestingly the effect of different confidence profiles in centro-parietal was considerable in double-pulse trials (**Figure 8**) but not in single-pulse trials (**Figure 7**).

3.7 Pupil responses

We took the mean baseline-corrected pupil signal during 200 ms before feedback delivery as our measure of pupil response. In line with previous work (Urai et al., 2017) pupil responses reflect decision confidence in our double-pulse trials (**Figure 9A**; Eq.19, $\beta_1 = -.95$, $p < .001$, 95% CI = [-1.10, -.79] , $f^2 = .31$) while in single-pulse trials the confidence profile is not significant (**Figure 9B**; Eq.19, $\beta_1 = -.42$, $p = .21$, 95% CI = [-1.09, -.24], $f^2 = .22$).

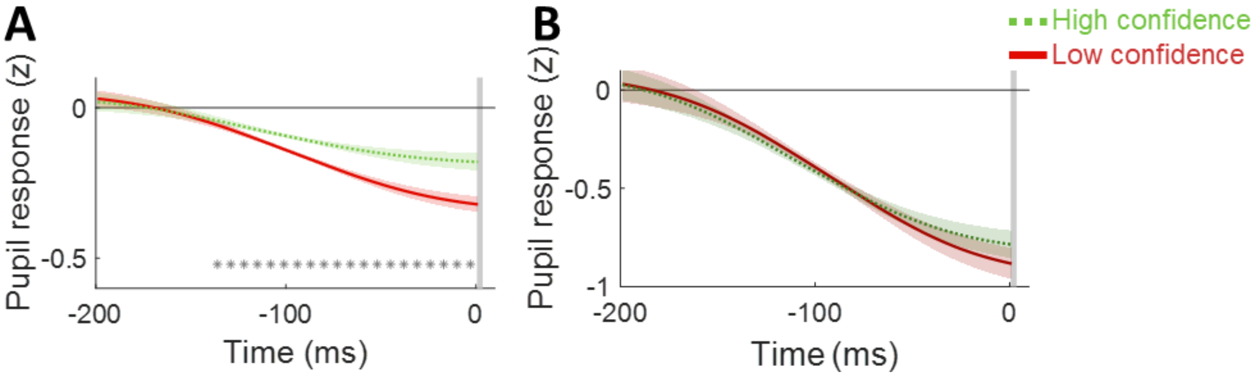


Figure 9. Standardized pupil response across time-window aligned to the feedback. (A) (B) Standardized pupil response, high confidence trials (green) vs low confidence trials (red) in double-pulse trials (A) and single-pulse trials (B). The shading region around the mean indicates SEM. * indicate p < .05 from a *t*-test.

4 Discussion

The current study was designed to clarify the confidence of decisions in more real-world contexts where the evidence arrives separately. Using an experimental design, we examined how human subjects combined the pieces of information to form their decision and confidence and how the two are related to each other. We performed two experiments with either single or double pulses of RDM stimuli. To this end, we investigated behavioral modeling, EEG responses and pupillometry. In summary, the results across experiments showed that participants used both pulses to decide about their confidence. Also, while their confidence was largely invariant to the gap interval, confidence scoring was not noticeably enhanced in double-pulse trials compared to single-pulse trials. Instead, participants reported their confidence with higher resolution and their metacognitive sensitivity improved. Furthermore, using RT, EEG and pupillometry analysis, we could considerably track the confidence profiles in double-pulse trials, unlike in single-pulse trials.

4.1 Behavioral and motion energy findings

Remarkably, unlike accuracy, confidence ratings in double-pulse trials have not increased significantly comparing to single-pulse trials. We hypothesize that participants mainly trust on the evidence of one of the pulses and ignore the other one. The trusted pulse can either be the first or second pulse; it also can simply be the stronger pulse. However, the effect of sequence and interaction of pulses on confidence was examined and no effect was observed. Moreover, motion stimulus fluctuations are known to influence the choice (Kiani et al., 2008; Resulaj et al., 2009) and confidence (Van Den Berg et al., 2016; Zylberberg et al., 2012), so they can inform us about the parts of the stimulus that bear more intensely on the choice and confidence (Kiani et al., 2013, 2008; Nienborg & Cumming, 2009). The motion energy analysis could not confirm the asymmetric influences of the two pulses for confidence. However, the motion energy analysis does provide independent confirmation of the unequal influences of the two pulses for choice (Kiani et al., 2013). Since, pervious research suggests that participants obtained more information from a second pulse (Kiani et al., 2013; Tohidi-Moghaddam et al., 2019), we hypothesized, in line with a large body of evidence (De Gardelle & Mamassian, 2015; Herce Castañón et al., 2019; Rahnev & Denison, 2018; Zylberberg et al., 2016, 2014), here observers do not make their decisions exactly in accordance with confidence rating.

Moreover, once comparing confidence in double-pulse trials grouped by accuracy, we show that the participants had lower confidence in double-pulse trials than single-pulse trials for incorrect choices but higher confidence for correct choices. In other words, compared with single-pulse trials, in double-pulse trials, participants adjusted their confidence by enhancing their confidence resolution or metacognitive sensitivity.

Typically, confidence facilitates evidence accumulation and drives a confirmation bias in perceptual decision-making (Rollwage et al., 2020). Likewise, we suggest that an extra brief and weak evidence can validate confidence and improve metacognitive sensitivity.

4.2 Computational modeling findings

To understand the nature of the differences in participants' metacognitive sensitivity in double-pulse vs single-pulse trials, we compared corresponding estimated metacognitive parameters. Likewise, we included the expected parameters that would be achieved in double-pulse trials under the assumption of perfect integration. Accordingly, we computed $Meta\text{-}d'/d'$ as a measure of 'metacognitive efficiency'. In the case of $Meta\text{-}d' = d'$, the observer is metacognitively 'ideal'. Indeed, all the information available for the decision would be used to report the confidence. Yet, in many cases, we might find that $Meta\text{-}d' < d'$, along with some degree of noise or suboptimality (Fleming & Lau, 2014; Maniscalco & Lau, 2012). Conversely, we may find that $Meta\text{-}d' > d'$, if subjects are able to draw on additional information such as hunches (Rausch & Zehetleitner, 2016), further processing of stimulus information (Charles, Van Opstal, Marti, & Dehaene, 2013) or extra prior knowledge on the task. In the model fitted to double-pulse trials, $Meta\text{-}d'/d'$ was around .8 and near to ideal for almost all participants. However, as in single-pulse trials, it varies considerably between participants, the value could not be adjusted in perfect integrator model similar to the behavioral model.

Previously, the better-than-expected performance in double-pulse trials was explained by underperformance in single-pulse trials (Kiani et al., 2013). Here, metacognitive sensitivity in double-pulse trials surpasses the value predicted by the perfect integrator model (**Figure 5C** and **Supplementary Figure 4**). This effect can be followed in all of our participant (except one of participants from EEG experiment) and can be explained by low confidence resolution in single-pulse trials.

Metacognitive noise is the noise that affects confidence estimates but not perceptual decisions (De Martino, Fleming, Garrett, & Dolan, 2013; Jang, Wallsten, & Huber, 2012; Maniscalco & Lau, 2016; Mueller & Weidemann, 2008; Rahnev, Nee, Riddle, Larson, & D'Esposito, 2016; Shekhar & Rahnev, 2018; Van den Berg, Yoo, & Ma, 2017). A recent work categorized sources of metacognitive inefficiency (Shekhar & Rahnev, 2020). Accordingly, metacognitive noise is a superordinate term for all noise sources that impact the confidence formation process (Shekhar & Rahnev, 2020, 2021) ranging from systematic to nonsystematic input and computation. Nevertheless, the exact source of metacognitive noise remains unclear (Shekhar & Rahnev, 2020). This noise can be tracked in our perfect integrator model, which was capable of accumulating decision evidence perfectly but could not predict confidence formation in our task. We suggest that the perfect integrator model was unable to adjust to confidence criteria when predicting confidence in double-pulse trials. However, an improved SDT capable of addressing metacognitive noise might be able to empower the employed perfect integrator model. Furthermore, SDT is not the only available model to implement a perfect integrator model. Previous studies suggested attractor models as a candidate model to implement the perfect integrator model (Kiani et al., 2013; Waskom & Kiani, 2018). Attractor models are a group of networks that formed a bridge between cognitive theory and biological data which exploits inhibition to achieve a competition among alternatives (Wang, 2002; Wong & Huk, 2008). Although these models can integrate momentary evidence to establish a decision, they have specific failure behaviors that would be apparent when the sources of evidence are separated by gaps in time (Kiani et al., 2013). Besides, when the stimulus is very short, mostly, none of the attractors could be reached and, the network would revert back to the resting state after the stimulus offset (Wang, 2002). Therefore, the choice would be assigned randomly. However, our experiments' data represent a noteworthy performance in single-pulse trials, which does not support this expectation. Consequently, to implement a perfect integrator model by implementing an attractor model, a mechanism for simulating a very short stimulus might be considered. Moreover, our behavioral assays highlighted different relationships between confidence and accuracy in the different conditions of the task. So, a dedicated neural module with a plausible circuit of confidence might be a better option to implement a perfect integrator model. Recently, multi-layer recurrent network models has been developed to account for decision confidence mechanisms (Atiya, Rañó, Prasad, & Wong-Lin, 2019; Paz, Insabato, Zylberberg, Deco, & Sigman, 2016). These models consist of multiple layers of neural integrators and in line with neural evidence of decision confidence (Kepecs, Uchida, Zariwala, & Mainen, 2008; Murphy, Robertson, Harty, & O'Connell, 2015), they are suggested to justify the observed behavior.

Furthermore, perceptual decisions are often modeled using ideal observers (e.g., SDT). However, a source of suboptimal behavior in decision-making is 'lapse' (Gold & Ding, 2013; Pisupati, Chartarifsky-Lynn, Khanal, & Churchland, 2021). Lapses are an additional constant rate of errors independent of the evidence strength (Gold & Ding, 2013; Pisupati et al., 2021). Lapse rate has been shown to increase with higher perceptual uncertainty (Pisupati et al., 2021) and would be accounted by fitting extra parameter to psychometrics models. Accordingly, as the perfect integrator model was based on SDT, ignoring lapse in the single-pulse trials might lead to mis-estimation of decision parameters in double-pulse trials. Consequently, further models including the lapse parameters (Pisupati et al., 2021), may improve the perfect integrator model's predictivity.

4.3 Implicit confidence markers

Although research suggests faster decisions accompanied by higher confidence (Kiani et al., 2014; Vafaei Shooshtari et al., 2019; van den Berg et al., 2016; Zylberberg et al., 2016), our results do not show such an association in the presence of a brief piece of evidence. Moreover, our participants decide much faster in double-pulse trials comparing to single-pulse trials. We hypothesized that the decrease of response-time in double-pulse trials would be reflected with higher internal confidence. However, another hypothesis of this variation pointed to the extra time duration in double-pulse trials, which can be used to increase readiness to decide. We regress the delay-time before response cue onset and response-time in both single-pulse and double-pulse trials to explore the hypothesis. If the variation of response-time was primarily dependent on extra delay time, the delay time should have had a considerable effect on response-time, especially in our 120ms single-pulse trials when the stimulus duration was concise and the delay time varied. Nevertheless, the effects in both double-pulse and single-pulse trials are weak. Accordingly, the hypothesis that faster decision reflect higher confidence in double-pulse trials is supported. In addition, the confidence profile as a function of response-time was significant in double-pulse trials unlike in single-pulse trials.

Our findings furthermore suggest that reported confidence might not follow confidence marker in EEG response. We focused on the CPP —a neural correlate of perceptual processing believed to reflect evidence accumulation and correlated to confidence (Boldt et al., 2019; Herding et al., 2019; Rausch et al., 2020; Vafaei Shooshtari et al., 2019; Zizlsperger et al., 2014). However, our findings suggest that in the presence of a brief and weak stimulus, entirely unlike in double-pulse trials, CPP amplitudes show no significant variation in high and low level of confidence. As confidence in single and double-pulse trials did not vary significantly, we suggest that variation of CCP amplitude share more commonalities with implicit confidence measure rather than explicit confidence measures like ratings. Moreover, we propose that pupil response relation to confidence rating varies as the task condition changes; when participants access brief and weak stimuli, no association detected, unlike in the presence of a pair of separated stimuli. Our current observations are not easily reconciled with existing theoretical accounts of the impact of the confidence level on pupil response (Allen et al., 2016; Lempert et al., 2015; Urai et al., 2017).

To sum up, when participants access brief and mainly weak stimuli, the confidence ratings are not reliable and confidence profile could not be tracked from response-time, pupil and EEG response. In other words, implicit confidence markers, in some case, might be incapable of following the conscious confidence rating. This is in line with innovative findings abstracting implicit confidence measures from explicit confidence measures (Logan & Crump, 2010).

4.4 Limitations and future directions

To the best of our knowledge, how evidence accumulation processes improve the accuracy confidence association was not addressed using the combination of behavioral, neural, and pupillometry signatures before. Obviously, our results were grounded in assumptions of integration strategy in decision-making. However, this insight has recently been reconsidered (Carland, Marcos, Thura, & Cisek, 2016; Stine et al., 2020). Participants' decisions might be better explained by an urgency-gate model (Evans, Hawkins, Boehm, Wagenmakers, & Brown, 2017; Thura, Beauregard-Racine, Fradet, & Cisek, 2012) rather than an integration strategy such as perfect integrator. A participant’s strategy could be something between no integration and perfect integration or in a completely different space of models (Stine et al., 2020) and might be change depending on task paradigm or even subject’s internal state (Evans & Hawkins, 2019; Najafi & Churchland, 2018; Tsetsos, Gao, McClelland, & Usher, 2012). Consequently, further models to discuss the decision strategy in the presence of separated pulses could guide future works. In addition, future experiments could develop computational approaches and attempt to implement other scenarios in a discrete environment to study choice and confidence formation and examine the involved processes.

In addition, although the vast number of trials for each participant allowed us to do a robust subject-wise analysis and our EEG study replicated the same behavioral and modeling data, the small number of participants we used prevents us from making general claims. Future research might capitalize on our paradigm to provide a situation in which confidence remains persistent but metacognitive sensitivity improved. In this way, future research continues studying the neural basis of metacognitive ability and consciousness in addition to previous works (Feuerriegel, Blom, & Hogendoorn, 2021; Fleming & Dolan, 2012).

5 Conclusion

To sum, the present study sheds new light on confidence formation, especially in perceptual decision-making when a pair of visual cues separated by diverse temporal gaps. Our data suggest that accumulated evidence from both pulses shapes confidence but not in line with accuracy. Moreover, we showed that the classic perfect integrator model merely highlighted evidence accumulation which predict the choice and ignored the effect the metacognitive noise that affects confidence. Finally, integrating evidence from two separated pieces of information makes the confidence profiles in RT, EEG and pupil responses show up, unlike the situation in which participants have to decide based on a brief and weak pulse of information.

6 Conflict of Interest

The authors declare no conflict of interest.

7 Author Contributions

ZA: conceptualization, data acquisition, analysis, visualization, writing - original draft, writing - review and editing; SZ: conceptualization, supervision, writing - review and editing; AJ: supervision, writing - review and editing; RE: conceptualization, supervision, writing - review and editing.

8 Acknowledgments

This research was partially supported by Iran Cognitive Sciences & Technologies Council (code 8066). Data were recorded in the Cognitive Science Laboratory of Shahid Rajaei University. The authors would like to thank Samuel Klein for their constructive comments and proofreading of the manuscript. We thank all our participant and finally, special thanks to the open science movement and generous researchers who we had their helpful comments during the implementation and analysis.

9 Ethics

The ethics committee of the Iran University of Medical Sciences (protocol #IR.IUMS.REC1399648) approved the experimental protocol, and subjects gave written informed consent.

10 Supplementary Material

The Supplementary Material for this article can be found online at:

11 Data Availability Statement

The datasets generated for this study are available on request to the corresponding author.

12 References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Josa A*, *2*(2), 284–299.

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *Elife*, *5*, e18103.

Atiya, N. A. A., Rañó, I., Prasad, G., & Wong-Lin, K. F. (2019). A neural circuit model of decision uncertainty and change-of-mind. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10316-8

Atiya, N. A. A., Zgonnikov, A., O'Hora, D., Schoemann, M., Scherbaum, S., & Wong-Lin, K. (2020). Changes-of-mind in the absence of new post-decision evidence. *PLoS Computational Biology*, *16*(2), e1007149.

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1), 1–11.

Baranski, J. V, Petrusic, W. M., Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., … Dolan, R. J. (2017). The relationship between perceptual decision variables and confidence in the human brain. *Journal of Neuroscience*, *32*(18), 412–428.

Boldt, A., Schiffer, A.-M., Waszak, F., & Yeung, N. (2019). Confidence predictions affect performance confidence and neural preparation in perceptual decision making. *Scientific Reports*, *9*(1), 1–17.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Carland, M. A., Marcos, E., Thura, D., & Cisek, P. (2016). Evidence against perfect integration of sensory information during perceptual decision making. *Physiology*, *115*(1), 1–13.

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage*, *73*, 80–94.

Cohen, J. (1970). Approximate power and sample size determination for common one-sample and two-sample hypothesis tests. *Educational and Psychological Measurement*, *30*(4), 811–831.

Colizoli, O., De Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports*, *8*(1), 1–13.

De Gardelle, V., & Mamassian, P. (2015). Weighting mean and variability during confidence judgments. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0120870

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Evans, N. J., & Hawkins, G. E. (2019). When humans behave like monkeys: Feedback delays and extensive practice increase the efficiency of speeded decisions. *Cognition*, *184*, 11–18.

Evans, N. J., Hawkins, G. E., Boehm, U., Wagenmakers, E.-J., & Brown, S. D. (2017). The computations that support simple decision-making: A comparison between the diffusion and urgency-gating models. *Scientific Reports*, *7*(1), 1–13.

Feuerriegel, D., Blom, T., & Hogendoorn, H. (2021). Predictive activation of sensory representations as a source of evidence in perceptual decision-making. *Cortex*, *136*, 140–146.

Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1), nix007.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338–1349.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443.

Fleming, S. M., Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, 1.

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *Neuroimage*, *106*, 134–143.

Gold, J. I., & Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. *Progress in Neurobiology*, *103*, 98–114.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J.-D. (2014). The relationship between perceptual decision variables and confidence in the human brain. *Cerebral Cortex*, *26*(1), 118–130.

Herce Castañón, S., Moran, R., Ding, J., Egner, T., Bang, D., & Summerfield, C. (2019). Human noise blindness drives suboptimal cognitive inference. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-09330-7

Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., & Blankenburg, F. (2019). Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. *NeuroImage*, *201*, 116011.

Hoeks, B., & Ellenbroek, B. A. (1993). A neural basis for a quantitative pupillary model. *Journal of*

Psychophysiology, *7*, 315.

Jang, Y., Wallsten, T. S., & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *Psychological Review*, *119*(1), 186.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*, *33*(50), 19434–19441.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1594), 1322–1337.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227.

Kiani, R., Churchland, A. K., & Shadlen, M. N. (2013). Integration of direction cues is invariant to the temporal gap between them. *Journal of Neuroscience*, *33*(42), 16483–16489.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, *28*(12), 3017–3029.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764. https://doi.org/10.1126/science.1169405

Kira, S., Yang, T., & Shadlen, M. N. (2015). A neural implementation of Wald's sequential probability ratio test. *Neuron*, *85*(4), 861–873.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27.

Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating pupil dilation and metacognitive confidence during auditory decision-making. *PLoS One*, *10*(5), e0126588.

Logan, G. D., & Crump, M. J. C. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science*, *330*(6004), 683–686.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430.

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.

Maniscalco, B., & Lau, H. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neuroscience of Consciousness*, *2016*(1).

Mathôt, S. (2013). A simple way to reconstruct pupil size during eye blinks. *Retrieved From*, *10*, m9.

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, *88*(1), 78–92.

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240.

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, *15*(3), 465–494.

Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications*, *7*, 13526.

Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *Elife*, *4*, e11946.

Najafi, F., & Churchland, A. K. (2018). Perceptual Decision-Making: A Field in the Midst of a Transformation. *Neuron*, *100*(2), 453–462.

Nienborg, H., & Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, *459*(7243), 89–92.

O'connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729.

Paz, L., Insabato, A., Zylberberg, A., Deco, G., & Sigman, M. (2016). Confidence through consensus: a neural mechanism for uncertainty monitoring. *Scientific Reports*, *6*, 21830.

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., & Churchland, A. K. (2021). Lapses in perceptual decisions reflect exploration. *Elife*, *10*, e55490.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*.

Rahnev, D., Nee, D. E., Riddle, J., Larson, A. S., & D'Esposito, M. (2016). Causal evidence for frontal cortex organization for perceptual decision making. *Proceedings of the National Academy of Sciences*, *113*(21), 6059–6064.

Rausch, M., & Zehetleitner, M. (2016). Visibility is not equivalent to confidence in a low contrast orientation discrimination task. *Frontiers in Psychology*, *7*, 591.

Rausch, M., Zehetleitner, M., Steinhauser, M., & Maier, M. E. (2020). Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring. *NeuroImage*, *218*, 116963.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*(21), 9475–9489.

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications*, *11*(1), 1–11.

Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*(3), 791–806.

Shekhar, M., & Rahnev, D. (2018). Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. *Journal of Neuroscience*, *38*(22), 5078–5087.

Shekhar, M., & Rahnev, D. (2020). Sources of Metacognitive Inefficiency. *Trends in Cognitive Sciences*.

Shekhar, M., & Rahnev, D. (2021). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review*, *128*(1), 45.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*(1), 1–13.

Stine, G. M., Zylberberg, A., Ditterich, J., & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making. *Elife*, *9*, e55365.

Tagliabue, C. F., Veniero, D., Benwell, C. S. Y., Cecere, R., Savazzi, S., & Thut, G. (2019). The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. *Scientific Reports*, *9*(1), 1–12.

Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology*, *108*(11), 2912–2930.

tickle, hannah, Tsetsos, K., Speekenbrink, M., & Summerfield, C. (2020). Optional Stopping in a Heteroscedastic World. *PsyArXiv*. https://doi.org/10.31234/OSF.IO/T7DN2

Tohidi-Moghaddam, M., Zabbah, S., Olianezhad, F., & Ebrahimpour, R. (2019). Sequence-dependent sensitivity explains the accuracy of decisions when cues are separated with a gap. *Attention, Perception, and Psychophysics*, *81*(8), 2745–2754. https://doi.org/10.3758/s13414-019-01810-8

Tsetsos, K., Gao, J., McClelland, J. L., & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, *6*, 79.

Twomey, D. M., Kelly, S. P., & O'Connell, R. G. (2016). Abstract and effector-selective decision signals exhibit qualitatively distinct dynamics before delayed perceptual reports. *Journal of Neuroscience*, *36*(28), 7346–7352.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643.

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 1–11.

Vafaei Shooshtari, S., Esmaily Sadrabadi, J., Azizi, Z., & Ebrahimpour, R. (2019). Confidence Representation of Perceptual Decision by EEG and Eye Data in a Random Dot Motion Task. *Neuroscience*, *406*. https://doi.org/10.1016/j.neuroscience.2019.03.031

Van Den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *ELife*, *5*(FEBRUARY2016), 1–21. https://doi.org/10.7554/eLife.12192

Van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological Review*, *124*(2), 197.

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. *Current Biology*, *26*(23), 3157–3168. https://doi.org/10.1016/j.cub.2016.10.021

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*(5), 955–968.

Waskom, M. L., & Kiani, R. (2018). Decision Making through Integration of Sensory Evidence at Prolonged Timescales. *Current Biology*, *28*(23), 3850–3856.e9. https://doi.org/10.1016/j.cub.2018.10.021

Wong, K.-F., & Huk, A. C. (2008). Temporal dynamics underlying perceptual decision making: Insights from the interplay between an attractor model and parietal neurophysiology. *Frontiers in Neuroscience*, *2*, 28.

Zizlsperger, L., Sauvigny, T., Händel, B., & Haarmeier, T. (2014). Cortical representations of confidence in a visual perceptual decision. *Nature Communications*, *5*, 3940.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*, 79.

Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *ELife*, *5*(OCTOBER2016), e17688. https://doi.org/10.7554/eLife.17688

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition*, *27*(1), 246–253. https://doi.org/10.1016/j.concog.2014.05.012

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual reasoning underlies the learning of priors in decision making. Neuron*, 99(5), 1083–1097.*

Supplementary Material

13 Supplementary Appendix 1: Figures and Tables

13.1 Supplementary Tables

Supplementary Table 1. Subtraction of confidence in double-pulse from single-pulse trials was significantly affected by choice accuracy.

Participant	$\boldsymbol{\beta_1}$
P_1	0.17 (p < 0.01) CI = [.15, .19]
P_2	0.21 ± 0.01 (p < 0.01) CI = [.19, .23]
P_3	0.10 ± 0.01 (p < 0.01) CI = [.08, .12]
P_4	0.12 ± 0.01 (p < 0.01) CI = [.10, .14]

Each row shows the coefficients of Eq.10 of manuscript, their related p values and a 95% confidence interval.

Supplementary Table 2. Performance was largely unaffected by interpulse interval for double-pulse trials with equal pulse strength and with unequal pulse strength.

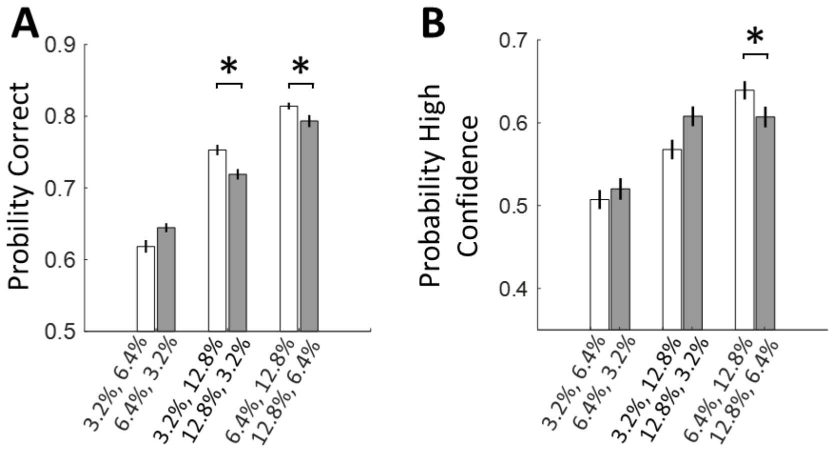
	Equal strength		Unequal strength		
Participant	β_3	β_4	β_3	β_4	β_5
P_1	-0.45 ± 1.07 (p = 0.67)	-0.01 ± 0.01 (p = 0.91)	0.39 ± 0.84 (p = 0.63)	-0.10 ± 0.09 (p = 0.29)	-0.01 ± 0.01 (p = 0.77)
	CI = [-2.55, 1.65]	95% CI = [-.03, .01]	CI = [-1.26, 2.03]	CI = [-.28, .08]	CI = [-.03, .01]
P_2	-2.58 ± 1.50 (p = 0.09)	-0.01 ± 0.08 (p = 0.93)	0.78 ± 1.23 (p = 0.52)	0.01 ± 0.15 (p = 0.95)	0.02 ± 0.07 (p = 0.77)
	CI = [-5.52, .36]	CI = [-.17, .15]	CI = [-1.63, 3.19]	CI = [-.28, .30]	CI = [-.11, .17]
P_3	0.48 ± 1.08 (p = 0.66)	0.07 ± 0.11 (p = 0.52)	2.18 ± 0.95 (p = 0.03)	- 0.19 ± 0.10 (p = 0.06)	0.14 ± 0.10 (p = 0.17)
	CI = [-1.64, 2.60]	CI = [-.15, .28]	CI = [.32, 4.04]	CI = [-.39, .01]	CI = [-.06, .34]
P_4	0.74 ± 0.97 (p = 0.44)	-0.12 ± 0.08 (p = 0.15)	-0.57 ± 0.75 (p = 0.45)	0.02 ± 0.08 (p = 0.81)	-0.07 ± 0.07 (p = 0.33)
	CI = [-1.16, 2.64]	CI = [-.28, .04]	CI = [-2.04, .90]	CI = [-.14, .18]	CI = [-.21, .07]

Each row shows the coefficients of Eq.4 and 5, their related p values and a 95% confidence interval of β_i.

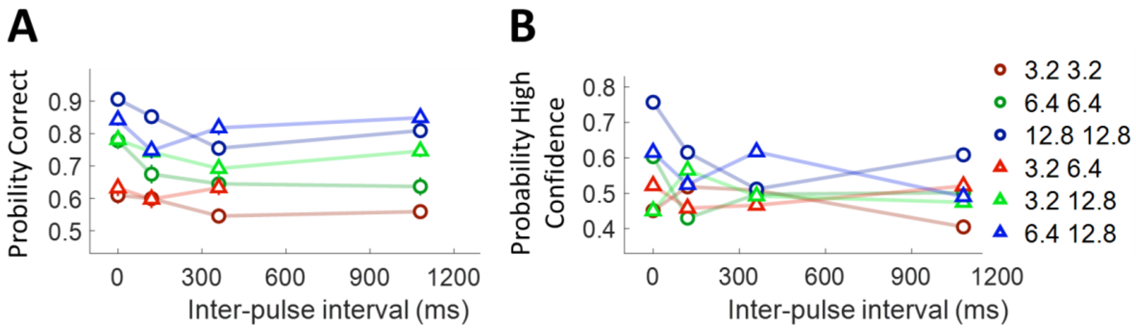
Supplementary Table 3. Pairwise comparisons across models (1: single-pulse trials, 2: double-pulse trials, 3: perfect integrator) for SDT parameters.

	d'			$Meta\text{–}d'$			$Meta\text{–}d'/d'$		
	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
tstat	3.79	4.05	1.25	1.98	0.21	2.48	1.74	0.12	3.58
df	22	22	22	22	22	22	22	22	22
pValue	0.001	0.52×10^{-5}	0.22	0.05	0.83	0.02	0.09	0.90	0.002
95% CI	[.13, .44]	[.20, 0.62]	[-.08, 0.32]	[-.02, 1.12]	[-.66, 0.82]	[.10, 1.15]	[-.17, 1.98]	[-1.09, 1.23]	[.35, 1.32]
Cohen's *d*	1.55	1.65	- 0.51	0.81	0.09	1.01	0.71	0.05	1.46

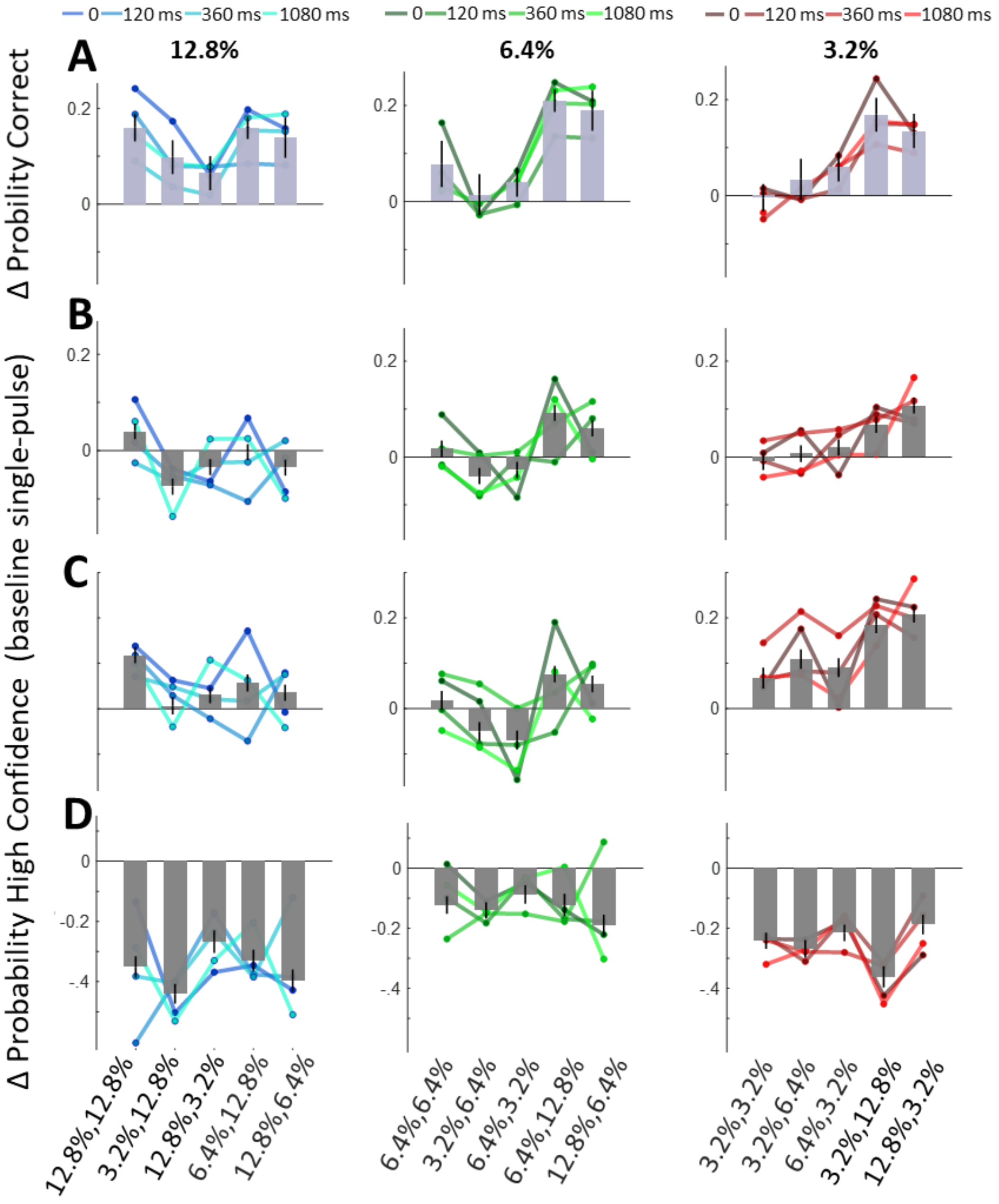
13.2 Supplementary Figures



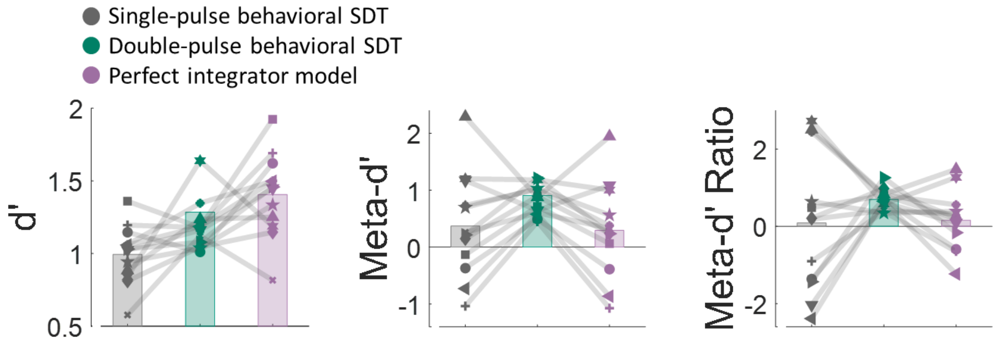
Supplementary Figure 10. Choice confidence was not depended on the sequence of motion pulses. (A) The weak–strong pulse sequence contributed higher accuracy than the strong–weak sequence. **(B)** The weak–strong pulse sequence did not contribute higher confidence comparing to the strong–weak sequence. In all panels, data are represented as group mean ± SEM. (*p<0.05)



Supplementary Figure 11. Interplay between confidence/accuracy and interpulse interval in double-pulse trials. (A) Choice accuracy for double-pulse trials grouping in all possible interval conditions. **(B)** Confidence of double-pulse trials was calculated by pooling data across all time intervals. In (A) and (B) each data point addresses pooled data from indicated sequence pulse and its reverse order (e.g., 12.8– 3.2% and 3.2 –12.8%).

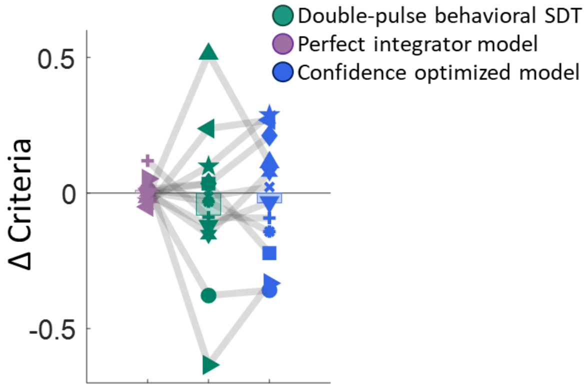


Supplementary Figure 12. Variation of accuracy or confidence in double-pulse trials baselined by corresponding coherence (3.2%, 6.4% and 12.8% for each column). (A) Considering all the trials, accuracy improved in combination with almost all pulses comparing to the baseline. **(B)** Considering all the trials, confidence improved in combination with stronger pulses while the confidence in sequence with a weaker pulse either decreased or remained constant. **(C)** In correct-choice trials, the increasing effect of stronger pulses is more significant and the confidence even slightly improved in combination with weaker pulses comparing to corresponding baseline. **(D)** Interestingly, in incorrect trials, the confidence decreased in every condition. The colored line representing matching data for each of four possible gaps. In bar graph, the data are represented as group mean ± SEM.



Supplementary Figure 4. Comparison of models and human behavior. Stimulus sensitivity (d'), metacognitive sensitivity ($Meta\text{-}d'$) and, metacognitive efficiency ($Meta\text{-}d'/d'$) estimated for single-pulse trials, double-pulse trials and the perfect integrator models.

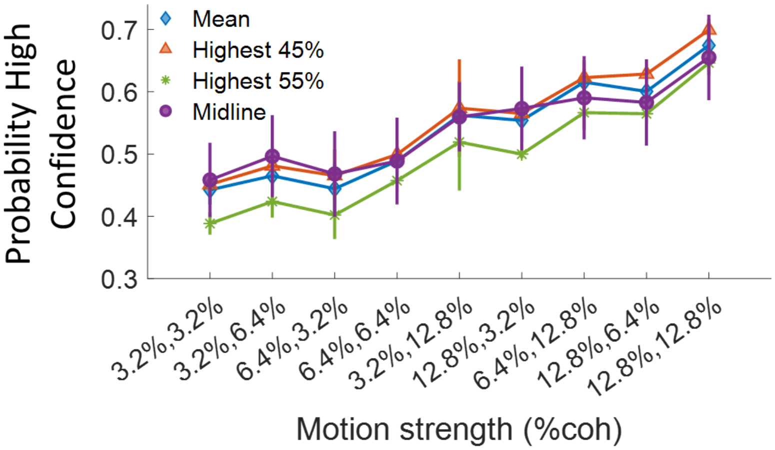
A univariate ANOVA showed that d' between models fit to double/single-pulse trials and the perfect integrator model significantly differed (F(2,33) = 9.99; p = 0.41 × 10^{-4}). Also, a univariate ANOVA showed that $Meta\text{-}d'$ between models fit to double/single-pulse trials and the perfect integrator model partialy differed (F(2,33) = 1.04; p = 0.09). We also computed metacognitive efficiency ($Meta\text{-}d'/d'$), A univariate ANOVA revealed a significant difference on all three models (F(2,33) = 2.50; p = 0.10).), We also applied the t-test as a post hoc procedure to compare all pairs of d', $Meta\text{-}d'$, $Meta\text{-}d'/d'$ from three models (Supplementary **Supplementary Table 3.** Pairwise comparisons across models (1: single-pulse trials, 2: double-pulse trials, 3: perfect integrator) for SDT parameters.Table 3).



Supplementary Figure 5. Variation of confidence criteria comparing to single-pulse trials in perfect integrator vs double-pulse trials and optimized model.

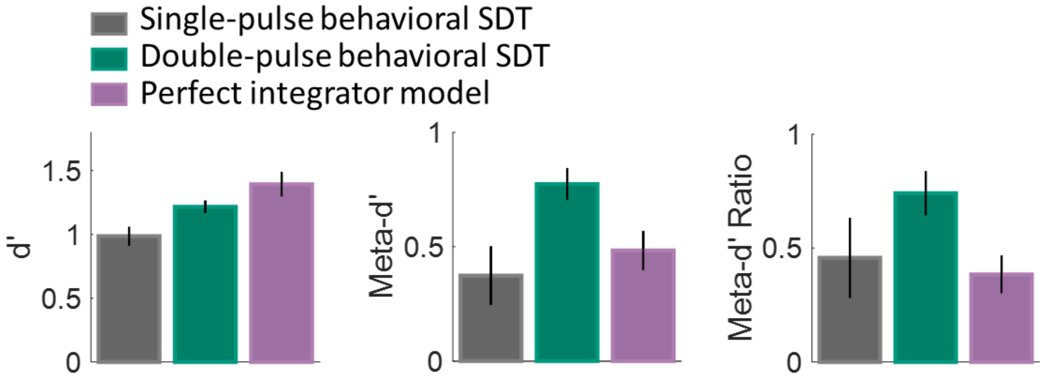
1000

1001



Supplementary Figure 6. A univariant Anova showed that confidence categorized by four different approaches in double-pulse trials not significantly differed (F(3,140) = 9.99; *p* = 0.34). Three paired-samples *t*-tests between our confidence categorization with other methods showed no difference (all *ps* > 0.36).

1002



Supplementary Figure 7. Comparison of models and human behavior considering the same numbers of trials. (a) Stimulus sensitivity (d'), metacognitive sensitivity ($Meta\text{-}d'$), metacognitive efficiency ($Meta\text{-}d'/d'$) estimated for single-pulse trials, double-pulse trials and the perfect integrator models.

We compared d', $Meta\text{-}d'$ and, $Meta\text{-}d'/d'$ of fitted models to single/double-pulse trials and simulated data by perfect integrator model, following up with three Dunn pair tests. A Kruskal-Wallis test showed that d' between models fit to double/single-pulse trials and the perfect integrator model not significantly differed ($H(3) = 3.23$; $p = 0.20$). We also applied the Dunn test as a post hoc procedure to compare all pairs of d' from three models. No d' in models significantly differed from others (all *ps* > 0.21).

Also, a Kruskal-Wallis test showed that $Meta\text{-}d'$ between models fit to double/single-pulse trials and the perfect integrator model significantly differed ($H(3) = 6.96$; $p = 0.03$). Post-hoc Dunn were used to compare all pairs of $Meta\text{-}d'$ from three models. The difference of $Meta\text{-}d'$ of single-pulse trials and double-pulse was significant ($p = 0.03$, CI = [- 12.58, -0.41]). However, the difference of $Meta\text{-}d'$ was insignificant for single-pulse trials and perfect integrator model ($p = 0.87$, CI = [-7.83, 4.33]) and for double-pulse trials and perfect integrator model ($p = 0.17$, CI = [4.75, 10.83]).

We also computed metacognitive efficiency ($Meta\text{-}d'/d'$), A Kruskal-Wallis test revealed a significant difference on all three models ($H(3) = 7.42$, $p = 0.02$), metacognitive efficiency in double-pulse and perfect integrator differed significantly ($p = 0.04$, CI = [0.16 12.33]) while in double-pulse and single-pulse, it partially differed ($p = 0.07$, CI = [-11.83, 0.33]). The difference of single-pulse and perfect integrator was not significant ($p = 0.99$, CI = [-5.58 6.58]).

14 Supplementary Appendix 2: Signal detection theory models

In the binary decision, the observer must discriminate between stimuli labeled S_2 or labeled S_1. Each stimulus presentation generates a value on an internal decision axis (Figure 1b), corresponding to the evidence in favor of S_1 or S_2. Evidence generated by each stimulus class is normally distributed across the decision axis, and the distance between these distributions in standard deviation units (d') measures how well the observer can discriminate S_1 from S_2. The observer sets a decision criterion cr, such that all signals exceeding cr are labeled S_2 and all those failing to exceed cr are labeled S_1. The observer also sets criteria $cr_{2,``S1"}$ and $cr_{2,``S2"}$ to determine confidence ratings around the decision criterion cr. These two thresholds must be well-ordered so that $\mathrm{cr}_{2,``S1"} < cr < \mathrm{cr}_{2,``S2"}$ (Figure 1b). When a S_2 response is made, a confident S_2 response requires the evidence also to have surpassed the $cr_{2,``S2"}$ threshold [1]. Consider only trials where the observer responds S_2, which means the decision axis exceeding cr. Then the S_2 distribution corresponds to the distribution of evidence for correct responses (i.e., S_2 stimuli classified as S_2), and the S_1 distribution corresponds to the distribution of evidence for incorrect responses (i.e., S_1 stimuli classified as S_2).

14.1 Confidence *Hit Rate* and *False Alarm Rate*

Sweeping the $\mathrm{cr}_{2,``S2"}$ criterion across the decision axis generates different values for confidence false alarm rate ($Prob(conf = "h" \mid stim \neq resp$) and confidence hit rate ($Prob(conf = "h" \mid stim = resp$). A summary of the observer's confidence performance is provided by hit rate (Hit Rate2) and false alarm rate (False Alarm Rate2)[1]:

$$\begin{aligned} \text{Hit Rate2} &= \text{Prob}(\text{conf} = "h" \mid \text{stim} = \text{resp}) = \frac{\text{n(high conf correct)}}{\text{n(correct)}}, \\ \text{False Alarm Rate2} &= \text{Prob}(\text{conf} = "h" \mid \text{stim} \neq \text{resp}) = \frac{\text{n(high conf incorrect)}}{\text{n(incorrect)}}, \end{aligned} \tag{1}$$

where $n(cond)$ denotes a count of the total number of trials satisfying the condition $cond$.

14.2 Decision *Hit Rate* and *False Alarm Rate*

In the SDT model, the decision hit rate (Hit Rate1) and the decision false alarm rate (False Alarm Rate1) are also calculated as follows:

$$\begin{aligned} \text{Hit Rate1} &= \frac{\text{n(resp=}Si\text{, stim=}Si\text{)}}{\text{n(stim = }Si\text{)}}, \ i = 1,2 \\ \text{False Alarm Rate1} &= \frac{\text{n(resp=}Si\text{, stim=}Sj\text{)}}{\text{n(stim = }Sj\text{)}}, \ i = 1, j = 2 \text{ or } i = 2, j = 1 \end{aligned} \tag{2}$$

where i and j represent the stimulus classification. After calculating the Hit Rate1 and False Alarm Rate1 of each participant, d' and cr are calculated as follows for each participant:

$$d' = \phi^{-1}(\text{Hit Rate1}, 0, 1) - \phi^{-1}(\text{False Alarm Rate1}, 0, 1)$$

$$cr = -0.5 * [\phi^{-1}(\text{Hit Rate1}, 0, 1) + \phi^{-1}(\text{False Alarm Rate1}, 0, 1)] \tag{3}$$

here, ϕ^{-1} is the inverse of a function that represents a normal cumulative distribution and is calculated as follows:

$$\phi(s, \mu, \sigma) = \int_{-\infty}^{0} N(v, \mu, \sigma) dv, \tag{4}$$

where $N(v, \mu, \sigma)$ is a Normal distribution with mean (μ) and standard deviation (σ). After the above calculations, to simplify the next calculations, we may consider the value of cr as zero point and move the distribution diagrams related to each option on the axis of the evidence.

By setting d', cr and two criteria $cr_{2,\text{“}S1\text{”}}$ and $cr_{2,\text{“}S2\text{”}}$ (Figure 1B), the probabilities of each confidence rating conditional on a given stimulus and response (Hit Rate2 and False Alarm Rate2) can be calculated theoretically according to the following equations:

$$\text{Prob}(\text{conf} = \text{”}h\text{”} | \text{stim} = S1, \text{resp} = \text{“}S1\text{”}) = \text{HitRate2}_{\text{”S1”}} = \frac{\phi\left(\text{cr}_{2,\text{“}S1\text{”}}, -\frac{\text{d}'}{2}\right)}{\phi\left(\text{cr}, -\frac{\text{d}'}{2}\right)}$$

$$\text{Prob}(\text{conf} = \text{”}h\text{”} | \text{stim} = S2, \text{resp} = \text{“}S1\text{”}) = \text{FalseAlarmRate2}_{\text{”S1”}} = \frac{\phi\left(\text{cr}_{2,\text{“}S1\text{”}}, \frac{\text{d}'}{2}\right)}{\phi\left(\text{cr}, \frac{\text{d}'}{2}\right)} \tag{5}$$

$$\text{Prob}(\text{conf} = \text{”}h\text{”} | \text{stim} = S1, \text{resp} = \text{“}S2\text{”}) = \text{HitRate2}_{\text{”S2”}} = \frac{1 - \phi\left(\text{cr}_{2,\text{“}S2\text{”}}, \frac{\text{d}'}{2}\right)}{1 - \phi\left(\text{cr}, \frac{\text{d}'}{2}\right)}$$

$$\mathrm{Prob}(\mathrm{conf} = ”h”|\mathrm{stim} = S2, \mathrm{resp} = “S2”) \; = \mathrm{FalseAlarmRate2}_{”\mathrm{S2}”} = \frac{1 - \phi\left(\mathrm{cr}_{2,“S2”}, -\frac{\mathrm{d}'}{2}\right)}{1 - \phi\left(\mathrm{cr}, -\frac{\mathrm{d}'}{2}\right)}$$

In the SDT model, there are different methods for adjusting the model with the data obtained from the experiments. In the method we used, d' and cr were calculated from the participants' performance (Eq. 3). Then, using maximum likelihood estimation (MLE) and Eq. 1 and 5 and by altering the value of the confidence criteria while holding d' and cr constant, a set of (Hit Rate2, False Alarm Rate2) pairs ranging between (0, 0) and (1, 1) were generated. Moreover, $Meta\text{–}\,d'$ was found by fitting the decision SDT model to response-specific confidence.

15 Supplementary references

[1] B. Maniscalco and H. Lau, “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*, Springer, 2014, pp. 25–66.