Boundary updating as a source of history effect on decision uncertainty

Heeseung Lee[1] & Sang-Hun Lee[1,*]

[1]Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Republic of Korea
[*]Lead contact:
visionsl@snu.ac.kr
1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea, 08826

SUMMARY

When sorting a sequence of stimuli into binary classes, current choices are often negatively correlated with recent stimulus history. This phenomenon—dubbed the repulsive bias—can be explained by boundary updating, a process of shifting the class boundary to previous stimuli. This explanation implies that recent stimulus history can also influence "decision uncertainty," the probability of making incorrect decisions, since it depends on the location of the boundary. However, there have been no previous efforts to elucidate the impact of previous stimulus history on decision uncertainty. Here, from the boundary-updating process that accounts for the repulsive bias, we derived a prediction that decision uncertainty increases as current choices become more congruent with previous stimuli. We confirmed this prediction in behavioral, physiological, and neural correlates of decision uncertainty. Our work demonstrates that boundary updating offers a principled account of how previous stimulus history concurrently relates to choice bias and decision uncertainty.

INTRODUCTION

We often characterize a stimulus for a particular dimension by assigning a binary class, such as 'small or large' for size, 'slow or fast' for speed, and 'light or heavy' for weight. This seemingly natural and basic linguistic act entails a cognitive computation called 'binary classification,' where we decide which side of the quantity distribution the stimulus falls on, with respect to the dimension of our interest[1-4]. Thus, successful binary classification requires not just accurately perceiving the stimulus quantity but also knowing the class boundary that divides the quantity distribution[1].

However, in everyday situations, the boundary is seldom presented explicitly at the moment of classification and, thus, needs to be learned and internally formed by observers themselves[5-12]. Importantly, it has been suggested that this internal boundary tends to continually shift towards recent stimuli[7-10,12] (Figure 1A). This "boundary shift towards recent stimuli" has accounted for a well-known history effect on choice called 'repulsive bias,' where the current choice is biased away from previous stimuli on a trial-to-trial basis[8-10,12,13] (Figure 1A,B). Recently, we formalized this boundary updating in the framework of Bayesian decision theory—called 'Bayesian Model of Boundary Updating (BMBU) — and demonstrated that BMBU readily accounts for the repulsive bias not just in behavioral choice but also in brain activity patterns[12].

Here, we note that the same boundary updating process that induces the repulsive bias in 'choice' should also confer another history effect on 'decision uncertainty.' Decision uncertainty is the probability that the current choice is incorrect, and this probability increases as the stimulus approaches nearer the boundary in binary classification tasks[11,14-19]. Suppose the boundary was initially neutral (Figure 1A, top), and a large stimulus was presented in the last trial (Figure 1A, middle). The boundary would then shift towards the large side according to the boundary updating hypothesis[7-10,12], bringing itself nearer to the stimuli associated with the 'large' choice but farther from those associated with the 'small' choice in the following trials (Figure 1A, bottom). Consequently, having previously viewed large stimuli, observers' 'large' choices are likely to be accompanied by high levels of uncertainty, whereas their 'small' choices by low levels of uncertainty. Generally put, the more congruent the current choice is with previous stimuli, the more uncertain it is (Figure 1C). We shall call this boundary-induced history effect on decision uncertainty *the pre-congruence effect*.

It is surprising that this seemingly straightforward, pre-congruence effect has not been recognized theoretically nor demonstrated empirically since decision uncertainty plays a fundamental role in diverse adaptive behaviors[20-22]. Decision uncertainty has been considered a crucial cognitive variable in adaptive human behavior, such as volatility monitoring[20], learning from errors[21], and executive control[22]. It should be noted that decision uncertainty differs from sensory uncertainty, which is influenced solely by the noise present in stimuli[23], whereas decision uncertainty can be modulated by the distance between the stimulus and the boundary even if sensory uncertainty is constant. Furthermore, decision uncertainty has wide-ranging positive associations with response time[11,16-18,24], pupillary responses[18,25-34], and neural activities in various brain regions[15,35-44].

To find empirical evidence for the pre-congruence effect, we probed the response time (RT), pupil size, and functional magnetic resonance imaging (fMRI) signals of human observers during binary

---

[1] The class boundary used in this paper is distinct from the 'decision boundary' used in signal detection theory or sequential analysis theory. We will explain this distinction in detail in the Discussion section. For the sake of brevity, the class boundary will also be referred to as the 'boundary.'

classification tasks with diverse experimental conditions. Initially, we confirmed that the current datasets display the history effect on choice (i.e., the repulsive bias) as predicted by BMBU. Then, by simulating BMBU, we quantitatively derived a set of implications of boundary updating on decision uncertainty, including the pre-congruence effect. Next, we confirmed these model implications by conducting multiple regression analyses on the RTs, the cortical activity in the dorsal anterior cingulate cortex (dACC), and the pupil-size measurements. The pre-congruence effect was significant in all datasets analyzed, and it remained robust despite changes in sensory modality, stimulus granularity, inter-trial interval, and trial-to-trial feedback. The magnitude of the pre-congruence effect was substantial, ranging from 10% to 45% of the effect of the current stimulus. Furthermore, the pre-congruence effect remained unaffected in the presence of previously known history effects on decision uncertainty. These results point to 'boundary updating towards recent stimuli' as a common source of two types of history effects: the repulsive bias, a well-known history effect on choice, and the pre-congruence effect, a previously unrecognized history effect on decision uncertainty.

RESULTS

History effects of previous stimuli on the current choice

Forty-one human individuals participated in our study and performed a binary classification task. They viewed a series of rings that appeared in a pseudo-randomized order[45] while classifying the size of each ring as 'small' or 'large' under moderate time pressure (Figure 2A). As previously reported[7-10,12], the participants showed the repulsive bias by making the 'small (large)' choice more often after seeing the large (small) stimulus than after seeing the small (large) stimulus in the previous trial (empty circles in Figure 2B). The repulsive bias was supported by the significant negative regression of the current choice C_t onto the previous stimulus S_{t-1} ($\beta_{S_{t-1}} = -0.57$; $P = 2.87 \times 10^{-14}$) in a multiple logistic regression model where C_t was simultaneously regressed onto the current stimulus S_t ($\beta_{S_t} = 1.30$; $P = 4.9 \times 10^{-22}$) and the previous choice C_{t-1} ($\beta_{C_{t-1}} = 0.12$; $P = 0.036$). The repulsive bias was not constrained to the immediately preceding stimulus but extended to the stimulus at the lag of three trials (Figure 2C).

Previously, our group demonstrated that the repulsive bias can be explained by the continual updating of the class boundary toward previous stimuli[12]. In doing so, we formalized such boundary updating with a Bayesian model, called the 'Bayesian model of boundary updating (BMBU),' and showed that BMBU quantitatively reproduces the history effects on choice with an exquisite level of finesse (, as indicated by the gray lines in Figure 2B and the bars in Figure 2C).

In the current study, we will demonstrate that boundary updating can also explain the history effects on decision uncertainty. To start, we will provide a brief overview of BMBU (see STAR methods) and describe how boundary updating relates to decision uncertainty.

Bayesian model of boundary updating

BUBU critically assumes that Bayesian observers have learned that their memory of stimuli becomes noisier as trials elapse[46,47]. Specifically, the learned generative model (Figure 3A) incorporates this memory decay by increasing the width of the distribution of noisy memory recalls $\sigma_{r_{t-i}}$ as a function of the number of elapsed trials i when the memory of the i-th previous stimulus r_{t-i} is recalled in the current trial t. The process of memory decay is approximated by a power function with a rate parameter κ ($\kappa > 0$):

$$\sigma_{r_{t-i}} = \sigma_m (1+\kappa)^i, \qquad \text{(Equation 1)}$$

where σ_m is the width of the distribution of sensory measurements of the current stimulus m_t.

With this generative model, the Bayesian observer can form two probabilistic beliefs, one regarding the current stimulus and the other regarding the current boundary, by inversely propagating the current sensory measurement m_t (Figure 3B) and the memory recalls of the previous stimuli r_{t-i} (Figure 3C), respectively, over the generative model. Note that the belief about the boundary can be formed in that way since the boundary for binary classification—a value that splits the stimulus distribution into even halves—is defined as the average size of the previous stimuli, which can be inferred from the noisy memory recalls of the observer. Here, the mean of this belief b_t is determined as the weighted average of recent stimulus memories r_{t-i} and a long-term stimulus memory μ_0:

$$b_t = \sum_{i=1}^{n} \beta_i r_{t-i} + \beta_0 \mu_0, \qquad \text{(Equation 2)}$$

where the weights are determined according to the generative model, as follows: $\beta_i \sim (1+\kappa)^{-2i}$; $\beta_0 = 1 - \sum_{i=1}^{n} \beta_i$. Consequently, past stimuli all attract b_t toward themselves, but with varying amounts: the attraction is maximal for the most recent one and decreases as trials elapse because $\beta_1 > \beta_2 > \cdots > \beta_n$.

Based on these beliefs about the current stimulus and boundary, the Bayesian observer computes the probability that the stimulus is larger than the boundary (the below-diagonal fraction of the bivariate distribution in Figure 3D), which is called decision variable. The decision variable v_t determines the current choice C_t: C_t is 'large' ($C_t = 1$) when v_t is greater than 0.5 and 'small' ($C_t = -1$) otherwise.

Importantly, the decision uncertainty u_t can also be computed by assessing the probability that the choice is incorrect from the same beliefs (the below-diagonal fraction of the bivariate distribution in Figure 3D), which is the reverse of confidence (f_t), the probability that the choice is correct ($u_t = 1 - f_t$)[48,49]. Thus, according to BMBU, the decision uncertainty is inversely proportionate to the distance between the stimulus and boundary estimates.

The attraction of the boundary to recent stimuli implies the repulsive bias for choice behavior and the pre-congruence effect for decision uncertainty

The way that the boundary is updated in BMBU implies that recent stimuli attract the boundary toward themselves on a trial-to-trial basis. To illustrate the trial-to-trial latent states of the inferred boundary, we fitted the parameters of BMBU to a single participant's choice data and simulated the boundary estimates using the best-fit parameters. A snapshot example of the resulting simulation is shown in Figure 3E, where the implication of BMBU can readily be appreciated by continuous and immediate shifts of the simulated boundary estimates toward the recent stimuli.

This boundary attraction toward recent stimuli, in turn, has two ramifications for decision behavior in the context of history effects, one regarding 'choice' and the other regarding 'decision uncertainty.' Its ramification for choice behavior is the repulsive bias: the boundary being attracted to previous stimuli causes the current choice to be biased toward the side opposite to the previous stimulus, as intuited early on (Figure 1A,B) and shown previously by our group[12].

The other—previously unidentified—ramification of the boundary attraction toward recent stimuli for decision uncertainty is the pre-congruence effect: the greater the congruency between the previous stimuli and the current choice ($S_{t-i} * C_t$) is, the greater the current level of decision uncertainty (u_t) is (Figure 3G), as we intuited early on (Figure 1A,C). Since the contribution of the previous stimulus to boundary updating diminishes as trials elapse due to memory decay (Figure 3A), BMBU predicts that the pre-congruence effect accordingly decreases as trials further lag (the purple bars in Figure 3H), similarly to the repulsive bias (the purple bars in Figure 2C).

It should be noted that besides the previously unidentified pre-congruence effect, BMBU readily captures the well-established link of decision uncertainty with the congruency between the current stimulus and the current choice ($S_t * C_t$). This phenomenon, where decision uncertainty decreases as the current stimulus becomes increasingly congruent with the current choice[14,18,48], will be referred to as the 'current-congruence effect' to distinguish it from the pre-congruence effect. BMBU predicts the current-congruence effect (as indicated by the progressive descent of the lines as a function of $S_t * C_t$ in Figure 3G and the yellow bar in Figure 3H) because the more congruent the current stimulus is with the current choice, the farther away it is from the current boundary, and thus the lower the decision uncertainty.

The above implication of BMBU regarding the history effects on decision uncertainty (as summarized in Figure 3G,H) was grounded on our model simulation of the trial-to-trial states of the latent variable u_t. And this simulation was carried out using the model best fit to the participants' choice data.

Thus, the validity of the simulated history effects critically relies on how accurately our model fitting procedure recovers the true states of u_t when the model complexity (e.g., number of parameters) and the experimental design (e.g., number of trials) are taken into account. To address this issue, we conducted a variable recovery test (see STAR methods for details) and confirmed that the recovered states of u_t well matched their true states ($R^2 = 0.91 \pm 0.0038$; mean $\pm$ 95% confidence interval).

As stressed earlier, BMBU's ability to exhibit the pre-congruence effect relies on the assumption that the memory recall of a stimulus becomes increasingly noisier as trials elapse. To examine the necessity of this assumption in accounting for the pre-congruence effect, we created three model variants in which the memory decay process is modified such that the boundary is inferred based on either only the immediately preceding stimulus ($\beta_i = 0$ for $i > 1$ in Equation 2), all the previous stimuli but with equal weights ($\kappa = 0$ in Equation 1), or the future stimuli instead of the previous stimulus (r_{t+i} is substituted for r_{t-i} in Equation 2). These model variants were all poorer than the original BMBU in predicting the trial-to-trial choice variability (Figure S1A). More importantly, they all failed to display neither the repulsive bias nor the pre-congruence effect on decision uncertainty (Figure S1B,C), which suggests that the assumption of memory decay is required for BMBU to account for the pre-congruence effect.

Pre-congruence effect in RT

To verify the pre-congruence effect predicted by BMBU, we first probed RT, a behavioral measure widely accepted as a correlate of decision uncertainty[16,18,24], by regressing it concurrently onto the pre-congruence ($S_{t-1} * C_t$), current-congruence ($S_t * C_t$), and choice-repetition ($C_{t-1} * C_t$) factors (Figure 4). Participants' RTs increased as the current choice became congruent with the previous stimulus, in support of the pre-congruence effect ($\beta_{S_{t-1}*C_t} = 0.11$, $P = 9.2 \times 10^{-9}$), and decreased as the current choice became congruent with the current stimulus, in support of the current-congruence effect ($\beta_{S_t*C_t} = -0.21$, $P = 2.7 \times 10^{-18}$) (Figure 4A,B). By contrast, the RTs were unaffected by whether the previous choice was repeated in the current trial ($\beta_{C_{t-1}*C_t} = -0.0046$, $P = 0.81$; Figure 4B). Notably, the size of the pre-congruence effect on the RTs was substantial and tantamount to 31% of the current-congruence effect (Figure 4B). The pre-congruence effect could be traced back to three trials (Figure 4C), consistent with the history effect on choice—the repulsive bias (Figure 2C).

Relationship between the pre-congruence effect and previously reported history effects involving decision confidence

Next, having confirmed the presence of the pre-congruence effect in RT, we considered the possibility that the pre-congruence effect is associated with, or can even be explained away by, previously reported history effects that involve decision uncertainty. Two such history effects are an autocorrelation in decision confidence between consecutive trials[50] and a tendency to repeat choices when decision confidence is high[17,18]. We indeed confirmed these two effects in our dataset: (i) the RT in the current trial (RT_t) increased as the RT in the previous trial (RT_{t-1}) increased (Figure 5A; $\beta_{RT_{t-1}} = 0.60, P = 3.0 \times 10^{-4}$); (ii) the probability of repeating choices ($C_{t-1} = C_t$) increased as the RT in the previous trial (RT_{t-1}) decreased and vice versa (Figure 5B; $t = 3.14, P = 0.0032$), which means that the repetitive and the alternative choice biases are facilitated after fast and slow RT in the preceding trial, respectively. If the facilitated choices involved faster RT_t, RT_t would be faster as choice is repetitive after fast RT_{t-1}, whereas RT_t would be faster as choice is alternative after slow RT_{t-1}. This modulatory effect of

the choice history on the association between RT_t and RT_{t-1} is observed by regressing RT_t concurrently onto RT_{t-1} and the interaction term $C_{t-1} * C_t \times RT_{t-1}$ (Figure 5C; $\beta_{C_{t-1}*C_t \times RT_{t-1}} = 0.043, P = 0.0046$; $\beta_{RT_{t-1}} = 0.061, P = 2.1 \times 10^{-4}$). Therefore, we examined whether the two history effects of RT identified in consistent with previous studies can explain away the pre-congruence effect.

The two effects require us to check whether the influence of the pre-congruence ($S_{t-1} * C_t$) on the current RT remains significant when the previous RT (RT_{t-1}) and the modulatory effects of the previous RT ($C_{t-1} * C_t \times RT_{t-1}$) are taken into account. To meet these two requirements, we conducted the previous regression analysis on RT_t while including three additional regressors: the previous RT (RT_{t-1}), its modulation of choice repetition ($C_{t-1} * C_t \times RT_{t-1}$), and its modulation of pre-congruence ($S_{t-1} * C_t \times RT_{t-1}$) (Figure 5D). We found that the current RT was significantly regressed onto the previous RT ($\beta_{RT_{t-1}} = 0.073,\ P = 4.9 \times 10^{-5}$) and its modulation of choice repetition ($\beta_{C_{t-1}*C_t \times RT_{t-1}} = 0.057,\ P = 2.3 \times 10^{-4}$), consistent with the two effects mentioned above, but not onto the previous RT's modulation of the pre-congruence ($\beta_{S_{t-1}*C_t \times RT_{t-1}} = 0.018,\ P = 0.13$). More importantly, as in the original regression, where those three additional regressors were not included, the influence of the pre-congruence on the current RT remained significant even after three trials elapsed (compare Figure 5C to Figure 4C). These results indicate that the two previously reported history effects involving decision uncertainty do not explain away but co-exist with the pre-congruence effect.

Robustness of the pre-congruence effect to sensory modality, stimulus granularity, inter-trial interval, and feedback presence

According to BMBU, the pre-congruence effect is to occur under various experimental conditions as long as the boundary for binary classification needs to be inferred from previously seen stimuli. To test this prediction, we analyzed three more datasets in our laboratory, which differ from the main dataset in sensory modality, stimulus granularity, inter-trial interval, or feedback presence (see Table 1 for specifications). Despite these differences, all these auxiliary datasets displayed the pre-congruence effect that is qualitatively similar to that in the main dataset for three important aspects (Figure 6). First, the current RT increased as the current choice became increasingly more congruent with the previous stimulus (Figure 6A1-3). Second, the pre-congruence effect was substantial in size, explaining 10% to 24% of the variance in RT that was explained by the current-congruence effect (Figure 6B1-3). Lastly, except for the dataset where trial-to-trial feedback was used, the pre-congruence effect remained significant even when trials elapsed by as far as five (Figure 6C1-3). These results suggest that the pre-congruence effect is a robust phenomenon, at least being generalizable to the experimental conditions used for our auxiliary datasets.

Pre-congruence effect in the fMRI activity of the dorsal anterior cingulate cortex

Previous studies have identified neural correlates of decision uncertainty in various non-human animal[14,49,51-53] and human[15,35-44] brain regions. Since we acquired fMRI measurements from eighteen of the participants while they performed the ring size classification task, we searched their brains for neural correlates of decision uncertainty and examined whether those neural correlates display the pre-congruence effects as BMBU implies.

We began by regressing the fMRI data solely onto the simulated trial-to-trial states of decision uncertainty produced by BMBU (u_t). As a result, we identified its neural correlates in the dorsal anterior cingulate cortex (dACC) and insula (Figure 7A; Table S1). In both regions, the fMRI responses showed

the highest correlation with u_t at 5.5 s after the stimulus onset (Figure S2, panels in the top and middle rows), which makes senses, given the hemodynamic delay in fMRI signal, because decision uncertainty is expected to be greatest at the moment of decision-making. Thus, as we did for the RT data, we regressed the trial-to-trial fMRI responses at that time point onto the three types of regressors (Figure 7B-D). The regression onto the current-congruence factor was significant in both regions ($\beta_{S_t*C_t} = -0.076,\ P = 0.0016$ for the dACC; $\beta_{S_t*C_t} = -0.079,\ P = 1.4 \times 10^{-5}$ for the insula). By contrast, the regression onto the pre-congruence factor was significant only in the dACC ($\beta_{S_{t-1}*C_t} = 0.053,\ P = 0.0037$ for the dACC; $\beta_{S_{t-1}*C_t} = 0.028,\ P = 0.083$ for the insula). The pre-congruence effect in the dACC was 46% of the size of the current-congruence effect (Figure 7C-1) and could be traced back to two trials (Figure 7D-1), similar to the RT findings. Furthermore, by expanding the regression analysis for the remaining within-trial time points, we confirmed that the pre-congruence effect is most pronounced at the moment of decision-making (Figure 7E-1).

Pre-congruence effect in pupillary responses

Previous studies have identified physiological correlates of decision uncertainty in pupillary responses[18,25-34]: the higher the decision uncertainty becomes, the larger the pupil grows in size. For the remaining twenty-three participants from whom fMRI data were not acquired, we tracked their pupillary responses during the task. Thus, we repeated the regression analysis on the pupil-size data to further verify the pre-congruence effect.

Consistent with the fMRI responses, the pupillary responses also showed the highest correlation with u_t at the moment of decision-making (Figure S2, panels in the bottom rows). Thus, as we did for the fMRI data, we regressed the trial-to-trial pupillary responses at that time point onto the three types of regressors (Figure 8A-C). The regressions onto the current-congruence and pre-congurence factors were both significant ($\beta_{S_t*C_t} = -0.076,\ P = 0.0016$; $\beta_{S_{t-1}*C_t} = 0.053,\ P = 0.0037$). The pre-congruence effect in pupil-size was 35% of the size of the current-congruence effect (Figure 8B) and could be traced back to one trial (Figure 8C). Similar to the fMRI data, we also confirmed that the pre-congruence effect in pupil-size is most pronounced at the moment of decision-making (Figure 8D).

DISCUSSION

Binary classification entails setting a boundary that divides quantities of interest into two even halves. Assuming that human brains are bounded by memory decay, BMBU prescribes a Bayesian solution of how the class boundary should be updated from their noisy recalls of recent stimuli. This solution implies that the class boundary is incessantly drawn towards recent stimuli over decision-making trials. As shown in our prior work[12], this boundary shift toward recent stimuli effectively explains the repulsive bias, a history effect of previous stimuli on current choices. Here, we deduced that the same boundary shift also leads to another, previously unrecognized, history effect: as the current choice becomes congruent with the previous stimulus, the decision uncertainty accompanying the current choice increases, which we refer to as the pre-congruence effect. We confirmed the pre-congruence effect in three well-known correlates of decision uncertainty: RT, dACC activity, and pupil size. Our analysis of the RT data indicates that the pre-congruence effect is not only robust to the changes in sensory modality, stimulus granularity, inter-trial interval, and trial-to-trial feedback but also cannot be attributed to the previously reported history effects associated with decision uncertainty. By concurrently verifying the implications of BMBU on the repulsive bias and pre-congruence effect, our work points to boundary updating as a common source of the history effects both on choice and decision uncertainty.

Novelty and uniqueness of the pre-congruence effect

Previous studies have shown that decision uncertainty, or its behavioral correlate such as RT, is influenced by various history factors. However, these factors differ from the pre-congruence factor, an extent to which the previous stimulus is congruent with the current choice ($S_{t-1} * C_t$). Below, we will identify these other history factors respectively and explain why they cannot explain away the pre-congruence effect.

Rahnev et al. (2015)[50] reported that subjective confidence reports are positively correlated across consecutive trials, dubbed 'confidence leak.' Here, the history factor is the subjective confidence report in the previous trial and thus includes neither the previous stimulus nor the current choice, unlike the pre-congruence factor. We showed that the current RT (RT_t) is regressed concurrently onto the previous RT (RT_{t-1}), a proxy of the previous confidence factor, and the pre-congruence factor (Figure 5A,D), which suggests the independence between the confidence leak and the pre-congruence effect.

Donner and his colleagues[17,18] reported that the tendency to repeat the previous choice in the current trial increases as the decision uncertainty accompanying the previous choice decreases. Here, decision uncertainty is not a target of the history effect but a history factor—a modulatory factor modifying the influence of the previous choice (C_{t-1}) on the current choice (C_t), to be exact. Nevertheless, the association between choice repetition ($C_{t-1}*C_t$) and decision uncertainty somehow might have contributed to the pre-congruence effect, which bears on the relationship between decision uncertainty and the pre-congruence ($S_{t-1}*C_t$), via the correlation between the previous stimulus (S_{t-1}) and choice (C_{t-1}). However, this possibility seems unlikely because the pre-congruence effect remained almost unaffected by the previous RT's modulatory effect (Figure 5D).

There are two other history effects on RT where the previous stimuli or their relationship with the current stimulus exert their influence on RT. First, patterns in preceding stimulus sequences, such as repetition and alternation, have been reported to affect the current RT[54]. The history factor is defined by the second-order relationship among past stimuli in this phenomenon, whereas the pre-congruence effect is defined by the respective relationships of past stimuli with the current choice. Second, RT is known to

decrease as the current stimulus becomes similar to the previous one, dubbed 'priming effect'[55,56]. The history factor is the congruence between the previous and current stimuli in the priming effect, whereas the history factor of pre-congruence effect is the congruence between the previous stimuli and the current choice. Even when considering the correlation between the current stimulus and choice, the bias in the priming effect is opposite to that in the pre-congruence effect: RT decreases in the priming effect but increases in the pre-congruence effect as a function of the congruence. Thus, the pre-congruence effect is distinct from these two history effects of stimuli on RT.

In a phenomenon called 'post-error slowing,' RT tends to increase following an error in the previous trial[57]. Here, the history factor is trial-to-trial feedback. However, we showed that the pre-congruence effect robustly occurred with and without trial-to-trial feedback (Figure 6).

Lastly, the pre-congruence effect should not be confused with a phenomenon where RT tends to be faster when the previous choice is repeated[58]. Here, the history factor is 'choice repetition'—i.e., the congruence between the previous and current choices ($C_{t-1}*C_t$). The regression of RT onto this choice repetition factor was also present in our datasets but much less substantial than the pre-congruence factor (Figures 4B,C and 6B,C). We suspect that these weak effects of choice repetition could have been caused by the counteract of the positive effect of motor repetition on RT[59] against the negative effect of choice repetition[58] because choices and motor responses were confounded in our experiment.

Grounded on the comparisons above, we conclude that the pre-congruence effect is a novel history effect that cannot be explained away by any previously reported history effects on decision uncertainty or RT.

Boundary updating, a principled and integrative account of the history effects on choice and decision uncertainty

Perceptual decision-making involves committing to a proposition about a particular world state. This commitment is expressed as a categorical choice and is accompanied by a continuous level of decision uncertainty. From a normative perspective, both 'choice' and 'decision uncertainty' are rooted in a common source, 'decision variable.' The choice is determined by which option is more supported by the decision variable, while the decision uncertainty is determined by how the chosen option is more supported by the decision variable than the unchosen option[23,48,49]. This close relationship between choice and decision uncertainty implies that the history effect on choice systematically relates to the history effect on decision uncertainty.

However, previous studies have rarely inspected these two types of history effects together, instead mainly focusing on the history effect on choice[8,13,17,60-65]. Although a few studies have explored the relationship between the history effects on choice and RT in a descriptive manner[58,66], they have not provided any clear explanation for these relationships. Our work offers a principled account by pointing to "boundary updating towards recent stimuli" as a common source of the history effects on choice and decision uncertainty: previous stimuli concurrently induce the repulsive bias in choice and the pre-congruence effect on decision uncertainty by attracting the class boundary towards themselves (Figure 1).

It is worth noting the 'normative' nature of updating the class boundary towards recent stimuli, given memory decay in the brain. If the brain's memory recalls of past stimuli were perfect, then its knowledge about stimulus distribution would stabilize quickly after a reasonable number of trials, resulting in a boundary that is not easily influenced by recent stimuli. In such a situation, the pre-congruence effect is considered 'irrational' because any 'rational' decision-maker should only rely on current sensory input. However, due to memory decay, the brain has to rely on its memory recalls of more

recent, thus more reliable, stimuli to better infer the actual stimulus distribution, resulting in a boundary that is incessantly drawn to recent stimuli. In this regard, the pre-congruence effect can be considered a by-product of the memory-bounded brain's 'normative' solution to infer the actual stimulus distribution from noisy memory recalls of recent stimuli. Looking at the pre-congruence effect from this perspective offers a fresh understanding of how history effects on decision uncertainty can be reinterpreted as rational decision-making. For instance, the presence of history effects on confidence reports has been considered an indication of decision-makers' metacognition inefficiency[67]. However, if we consider the normative nature of boundary updating, the absence, not the presence, of the history effect on confidence reports may indicate their metacognitive inefficiency.

Lastly, it is important to note that the pre-congruent effect is not only statistically significant but also substantial in size. In fact, it ranged from 10% to 46% of the current-congruency effect in our datasets. This implies that if boundary updating is not incorporated into the estimation of decision uncertainty, the trial-to-trial estimates of decision uncertainty substantially deviate from their actual states. Consistent with this implication, cortical activity in any single voxel did not significantly correlate with decision uncertainty when a model variant lacking the boundary updating process was used to define decision uncertainty (Figure S3). Additionally, incorporating boundary updating into the definition of decision uncertainty made our model estimates of decision uncertainty sensitive enough to identify the dACC, but not the insula, as a genuine neural correlate of decision uncertainty based on the presence of the pre-congruence effect (Figure 7). Given that previous efforts to identify the neural correlate of task difficulty have found it difficult to distinguish between these two areas[11,16,68,69], considering the pre-congruence effect is expected to facilitate the search for the 'genuine' neural substrate of decision uncertainty.

Limitations of the Study

It is important not to confuse the 'class boundary' with the 'decision boundary' in signal detection theory[70,71]. The class boundary represents the center of a stimulus distribution, while the decision boundary is the optimal boundary value that maximizes the objective function of any given decision-making task. Therefore, unlike the class boundary, the decision boundary can be adjusted not only by previous stimuli[72] but also by the value and base rate of choices[73,74]. This means that decision uncertainty can be influenced by non-stimulus history factors as well. In this regard, BMBU should be considered an account of the history effects that arise from previously experienced stimuli. More work needs to be done to comprehensively understand the history effects that come from diverse sources.

In addition, the formalism of BMBU has a restriction of only being suitable for binary classification on quantities that vary along a single dimension. Therefore, it would be worth incorporating the decision boundary into BMBU because the formalism of signal detection theory is proficient in dealing with quantities defined in more than one dimension[70].

Our results can be said to address the history effect on decision confidence because the statistical definition of decision confidence (f_t) is the probability that the current choice is correct[23,48,49] which is the reciprocal relationship to decision uncertainty ($u_t = 1 - f_t$). However, it is hard to argue that our work addressed the history effect on the decision confidence that is subjectively defined, which is often measured by the ratings provided by decision-makers[48]. This is because the metacognitive ability may interfere with the performance of the confidence rating in this self-report paradigm[67]. Further research is needed to investigate whether the pre-congruence effect is present in self-rated decision confidence.

ACKNOWLEDGEMENTS

This research was supported by the Research Grant from Seoul National University (339-20220013), the Brain Research Program (No. NRF-2021R1F1A1052020, NRF-2017M3C7A1047860), and the Basic Research Laboratory Program (NRF-2018R1A4A1025891) and the Basic Science Research Program (RS-2023-00276729) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT.

AUTHOR CONTRIBUTIONS

Conceptualization, H.L. and S.-H.L.; Methodology, H.L. and S.-H.L.; Investigation, H.L. and S.-H.L.; Writing – Original Draft, H.L.; Writing – Review & Editing, H.L. and S.-H.L.; Funding Acquisition, H.L. and S.-H.L.; Resources, S.-H.L.; Supervision, S.-H.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the initial drafting process, H.L. utilized ChatGTP to verify the naturalness of his phrases and sentences. Afterward, the corresponding author, S.-H.L., thoroughly revised the initial draft and made the necessary edits to the content. S.-H.L. takes full responsibility for the publication's content.

MAIN FIGURES

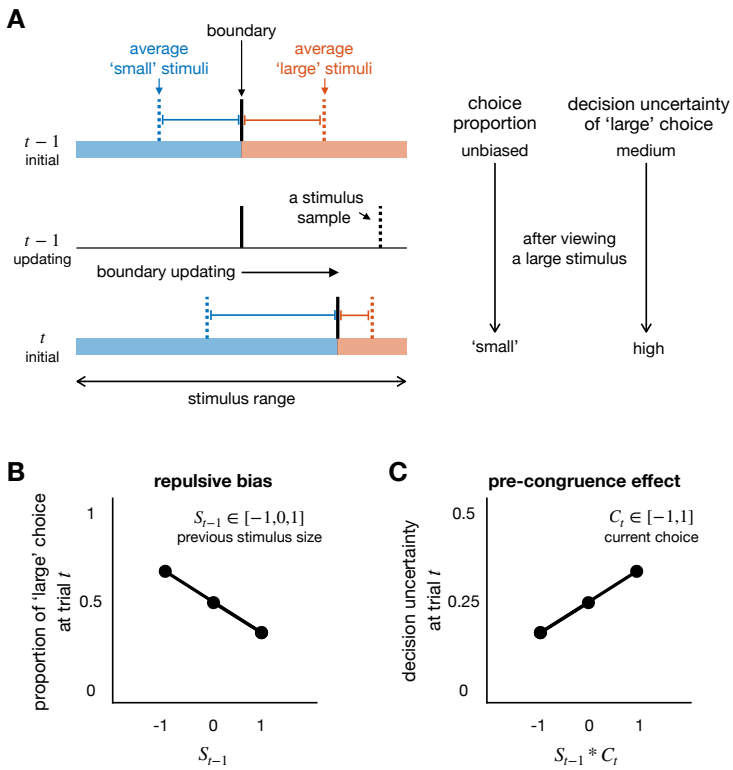

Figure 1. Boundary updating, as a common source of bias in choice and decision uncertainty during binary classification

(A) Schematic illustration of how boundary updating leads to the repulsive bias in choice and the pre-congruence effect on decision uncertainty. When the class boundary (solid vertical black bars) is unbiased, the stimulus sizes that lead to 'large' choices and those to 'small' choices are equal in proportion (as indicated by a large blue rectangle and a small red rectangle) and in distance (as indicated by a long blue horizontal line and a short red horizontal line) to the boundary (top row, left). Thus, both choice and decision uncertainty are unbiased (top row, right). According to the hypothesis of boundary updating, the boundary shifts toward the large side after encountering a large stimulus (middle row, left). As a result, the stimulus sizes that lead to 'large' choices and those to 'small' choices are no longer equal both in proportion and in distance to the boundary (bottom low, left). The imbalance in proportion leads to more 'small' choices in the following trial (the repulsive bias; bottom row, right), while the imbalance in distance to the boundary leads to an increase in decision uncertainty for the 'large' choice in the following trial (the pre-congruence effect; bottom row, right). Horizontal rectangles demarcate the stimulus ranges that are smaller (blue) or larger (red) than the boundary. The dotted blue and red vertical bars demarcate the average stimulus sizes smaller and larger than the boundary, respectively.

(B) Repulsive bias in choice proportion. The repulsive bias can be signified by the negative correlation between the choice in the current trial and the stimulus in the previous trial. S_{t-1} indicates the stimulus size at trial $t-1$ (the immediately preceding trial), where small, medium, and large sizes are denoted by -1, 0, and 1, respectively.

(C) Pre-congruence effect on decision uncertainty. The pre-congruence effect can be signified by the positive correlation between the decision uncertainty at trial t (the current trial) and the congruence of

the choices between the previous and current trials. C_t indicates the choice at trial t, where 'small' and 'large' choices are denoted by -1 and 1, respectively.

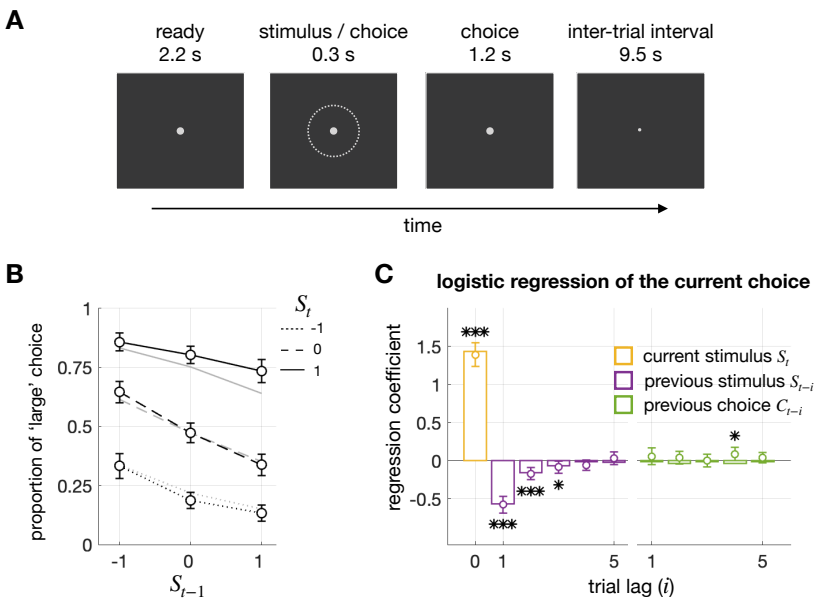

Figure 2. Task structure and repulsive bias

(A) Task structure. On each trial, the size of a ring was classified as 'small' or 'large' within 1.5 sec after the stimulus onset.

(B) Repulsive bias in psychometric curves. The proportions of 'large' choices conditioned on current stimuli (S_t) are plotted against previous stimuli (S_{t-1}), in black lines for human observers and in gray lines for the Bayesian model for boundary updating (BMBU). The negative slopes of the curves signify the repulsive bias.

(C) Repulsive bias in logistic regressions. The coefficients in the multiple logistic regression of current choices (C_t) onto current and previous stimuli (S_{t-i}) and previous choices (C_{t-i}) are plotted in circles and bars for humans and BMBU, respectively. The significant negative coefficients of previous stimuli, not previous choices, signify the repulsive bias. Here and thereafter, error bars represent the 95% confidence interval of the means, and asterisks indicate *P* values of two-sided Student's *t*-test: $*< 0.05$, $**< 0.01$, $***< 0.001$.

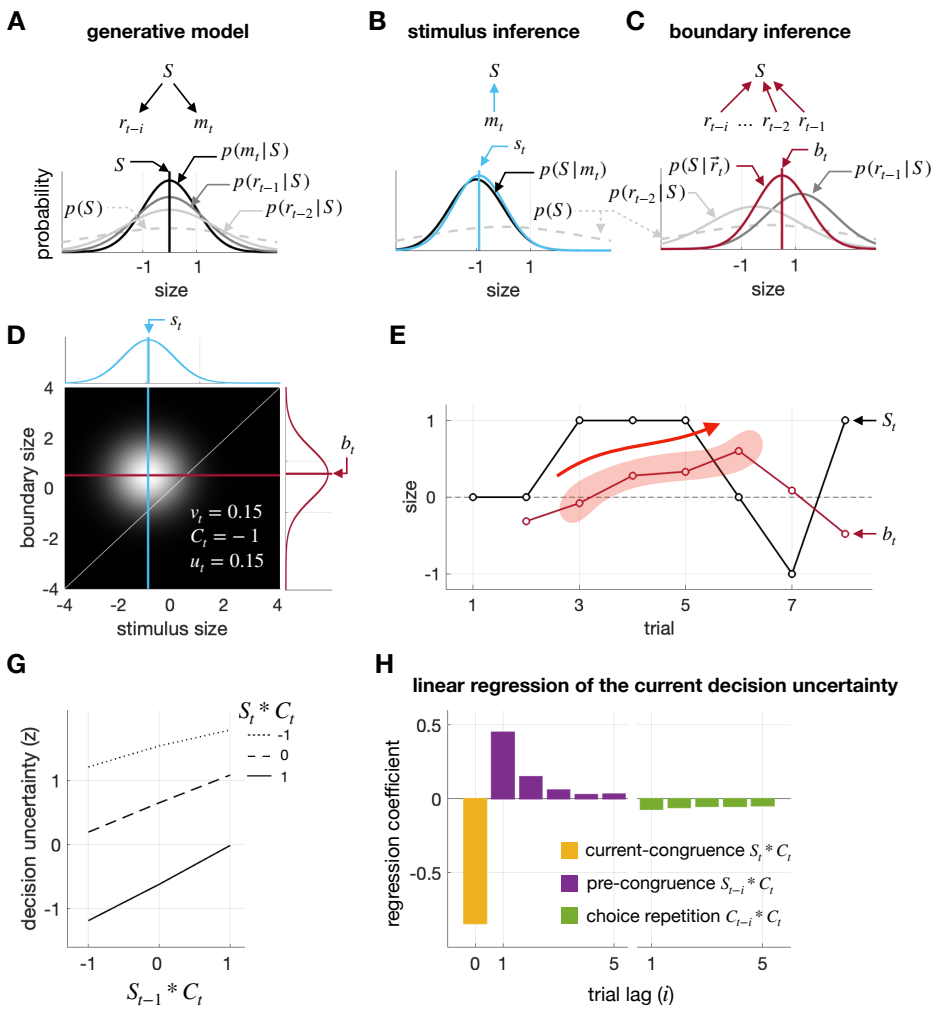

Figure 3. Bayesian model of boundary updating and its simulation of pre-congruence effect

(A) Generative model. BMBU posits that the sensory measurement of the current stimulus (m_t) and the memory evidence of the previous stimulus in the i-th preceding trial (r_{t-i}) are both normally distributed, as denoted by the solid curves $p(m_t|S_t)$ and $p(r_{t-i}|S_{t-i})$, respectively. For illustrative purposes, the stimulus sizes are set at a single value ($S_t = S_{t-i} = S$). To account for the decay in memory precision, BMBU posits that the distribution of memory measurements $p(r_{t-i}|S_{t-i})$ becomes broad as trials elapse. Additionally, BMBU posits that observers have prior knowledge of the distribution of physical stimulus size, as denoted by the dashed curve $p(S)$.

(B) Inference of stimulus size. BMBU infers the current stimulus size by constructing the posterior distributions of stimulus size based on the current sensory measurement ($p(S|m_t)$). The estimate of the stimulus s_t is the mean of the posterior distribution, as indicated by the blue vertical line.

(C) Inference of boundary location. BMBU infers the boundary by constructing the posterior distribution of stimulus size based on the memory measurements of the previous stimuli ($p(S|r_{t-1}, r_{t-2}, \dots) = p(S|\vec{r}_t)$). The estimate of boundary b_t is the mean of the posterior distribution, as indicated by the red vertical line. For illustrative purposes, only the memory measurements of the stimuli within trial lags of two (r_{t-1} and r_{t-2}) are considered. Note that the boundary estimate (b_t) is more biased toward the

memory measurement of the more recent stimulus (r_{t-1}) than the farther one (r_{t-2}), which explains why the attraction of the boundary toward previous stimuli diminishes as trials elapse.
(D) Binary classification with decision uncertainty. The posterior distributions of stimulus size and class boundary are projected onto a bivariate space, where the image intensity corresponds to the probability density of the product of the two posterior distributions ($p(S|m_t) \times p(S|\vec{r}_t)$). BMBU classifies the current stimulus as 'large' if the probability that the inferred stimulus is larger than the boundary (the decision variable, v_t) is greater than 0.5, and as 'small' otherwise. Decision uncertainty (u_t) is quantified as the probability that the current choice (C_t) is incorrect and increases as the stimulus gets nearer the boundary. In this particular example, v_t is 0.15, C_t is 'small' (-1), and u_t is 0.15.
(E) An example trajectory of the boundary in a representative subject 04. The black and red lines track the stimulus size and the boundary inferred by BMBU, respectively, over eight trials. The shaded red area highlights a gradual shift of the boundary towards the large side due to a streak of large stimuli.
(G) Model simulation of pre-congruence and current-congruence effects in psychometric curves. The simulated decision uncertainty increases with the congruence between the current choice and the previous stimuli ($S_{t-1} * C_t$) but decreases with the congruence between the current choice and the current stimulus ($S_t * C_t$), the former and latter signifying the pre-congruence and current-congruence effects on decision uncertainty, respectively.
(H) Model simulation of pre-congruence and current-congruence effects in linear regressions. The vertical bars represent the coefficients of the multiple linear regression of the simulated decision uncertainty onto the congruences of current choices with the current stimulus ($S_t * C_t$), the previous stimuli ($S_{t-i} * C_t$), and the previous choices ($C_{t-i} * C_t$). The pronounced positive coefficients of the congruence between the current choice and recent stimuli (purple bars) signify the pre-congruence effect.
See also Figure S1.

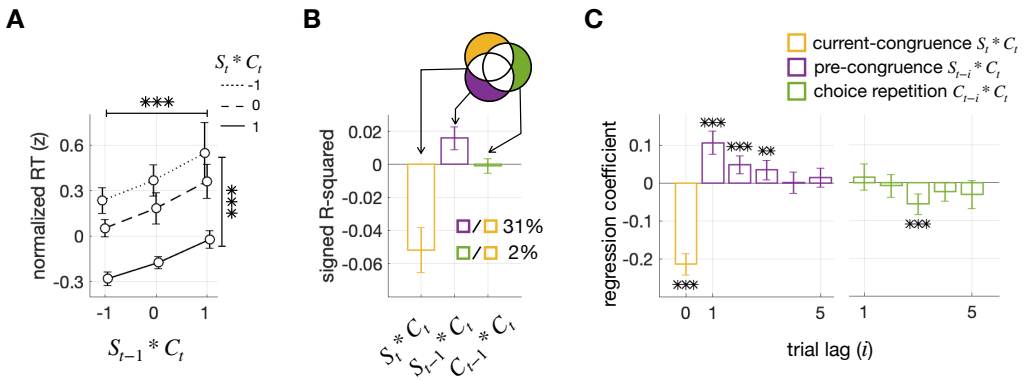

Figure 4. Pre-congruence and current-congruence effects on RT

(A) Changes in RT as a function of the pre-congruence and current-congruence factors. Normalized measures of RT are plotted against the levels of the congruence between the current choice and the immediately preceding stimulus ($S_{t-1} * C_t$), separately for each level of the congruence between the current choice and the current stimulus ($S_t * C_t$). The asterisks on top and side indicate the statistical significance ($P < 0.001$) of the pre-congruence ($S_{t-1} * C_t$) and current-congruence ($S_t * C_t$) effects, respectively. Their significance was assessed by the multiple regressions of RT onto the two congruence factors ($S_{t-1} * C_t$ and $S_t * C_t$) and the choice-repetition factor ($C_{t-1} * C_t$).

(B) Explained variance analysis. Bars plot the RT variance uniquely explained by the current-congruence factor ($S_t * C_t$; yellow), the pre-congruence factor ($S_{t-1} * C_t$; purple), and the choice-repetition factor ($C_{t-1} * C_t$; green), respectively, as illustrated by the Venn diagram. The percentage scores quantify the ratio of the variances explained by the pre-congruence and choice-repetition factors, respectively, to the variance explained by the current-congruence factor.

(C) Multiple regression analysis. Bars plot the regression coefficients for the current-congruence factor (($S_t * C_t$; yellow), the pre-congruence factor with trial lag i ($S_{t-i} * C_t$; purple), and the choice-repetition factor with trial lag i ($C_{t-i} * C_t$; green). The asterisks indicate the significance of regression coefficients: $*, P < 0.05$; $**, P < 0.01$; $***, P < 0.001$.

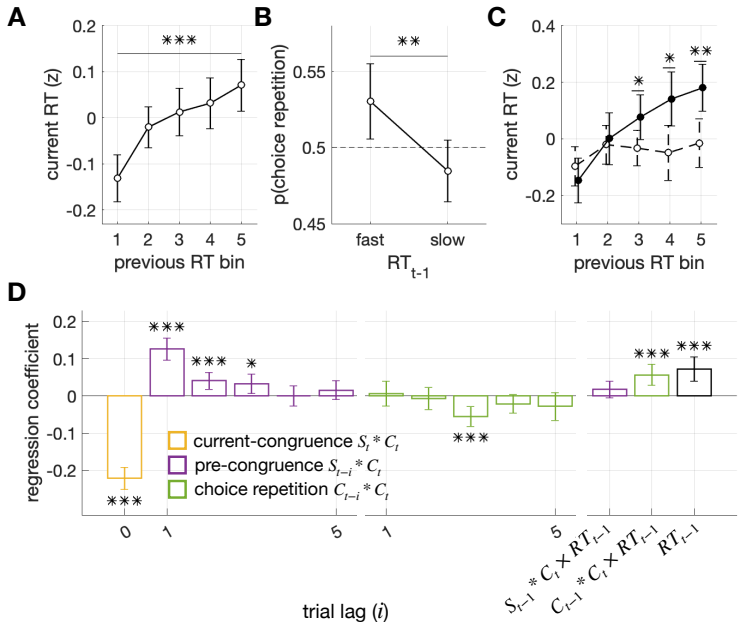

Figure 5. Co-existence of pre-congruence and current-congruence effects with other confidence-related history effects

(A) Positive correlation in RT between consecutive trials. The RT in the current trial was plotted against the binned RTs in the previous trial. The asterisks indicate the significance ($P < 0.001$) of the linear regression of the current RT (RT_t) onto the previous RT (RT_{t-1}).

(B) Effect of previous RT on choice repetition. The probability of repeating the previous choice in the current trial is plotted against the split-half levels of the previous RT. The asterisks indicate the significant difference in choice repetition between these two levels ($P < 0.01$, two-sided pair-wise t-test).

(C) Modulatory effects of choice repetition on the relationship between previous and current RTs. The RT in the current trial was plotted against the binned RTs in the previous trial, separately when consecutive choices were repeated ($C_{t-1} * C_t = 1$; solid circles with solid line) and when alternated ($C_{t-1} * C_t = -1$; empty circles with dashed line). The asterisks indicate the significant differences in current RT between the choice repetition and alternation conditions at given bins of previous RT (two-sided pair-wise t-tests; $*, P < 0.05$; $**, P < 0.01$).

(D) Multiple regression analysis with additional regressors. The multiple regression analysis depicted in Figure 4C was repeated but with three more regressors: the RT in the previous trial (RT_{t-1}) and the two interaction terms, one reflecting the modulation of the previous RT by the pre-congruence factor ($(S_{t-i} * C_t) \times RT_{t-1}$) and the other reflecting the modulation of previous RT by the choice-repetition factor ($(C_{t-i} * C_t) \times RT_{t-1}$). The asterisks indicate the significance of regression coefficients: $*, P < 0.05$; $**, P < 0.01$; $***, P < 0.001$.

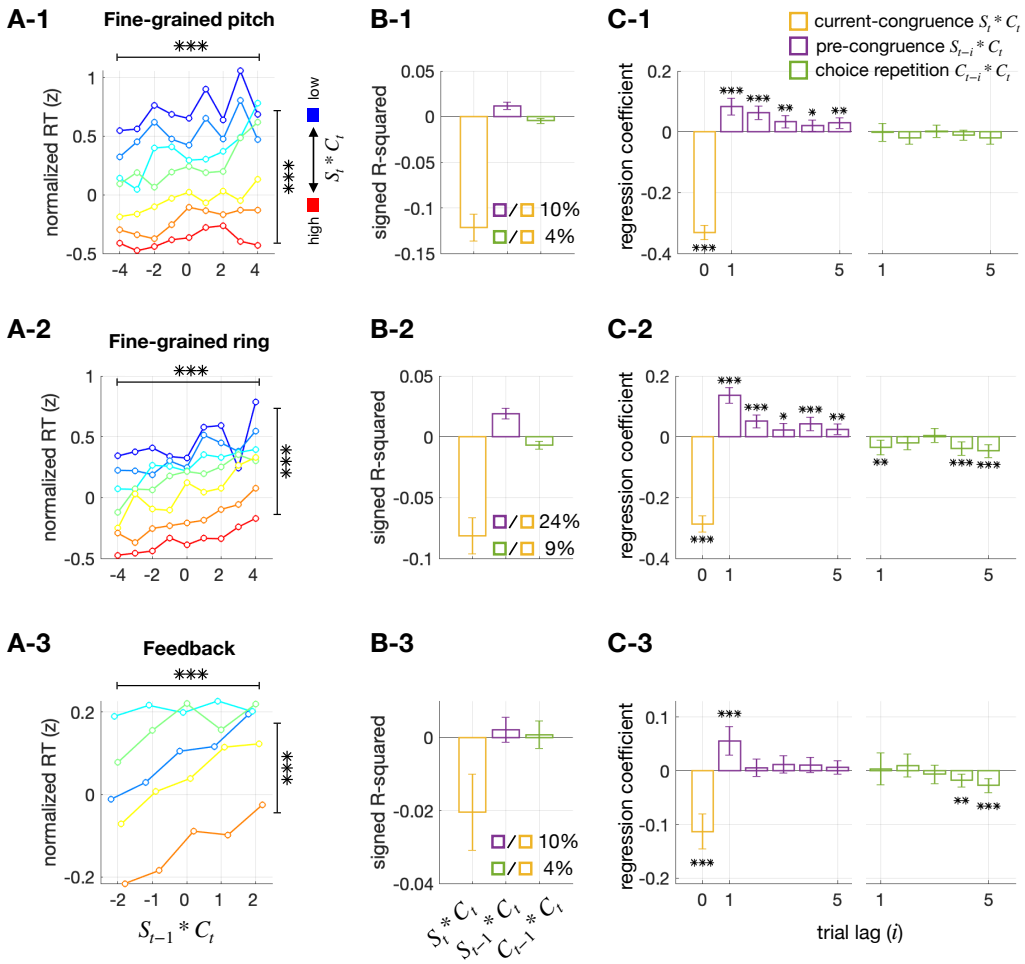

Figure 6. Pre-congruence and current-congruence effects on RT in three other datasets acquired with different task protocols

The analyses shown in Figure 4 were repeated on three additional datasets. While these datasets used the same binary classification task as the main dataset, they varied in several task protocols, including stimulus feature, stimulus granularity, inter-trial interval, and presence of trial-to-trial feedback. The name and details of each dataset are specified in Table 1. The summarized results of the datasets named 'fine-grained pitch,' 'fine-grained ring,' and 'feedback' in Table 1 are presented in the top (1), middle (2), and bottom (3) rows of panels. For datasets 'fine-grained pitch' and 'fine-grained ring,' where continuous stimuli were used, the stimuli were discretized into nine bins for illustrative purposes. In doing so, the data points at the two lowest bins defined for the current-congruence factor ($S_t * C_t$) are not shown because they were the trials where incorrect choices were made despite strong stimulus evidence and thus very rare.

(A) Changes in RT as a function of the pre-congruence and current-congruence factors. The format is identical to that of Figure 4A.

(B) Explained variance analysis. The format is identical to that of Figure 4B.

(C) Multiple regression analysis. The format is identical to that of Figure 4C.

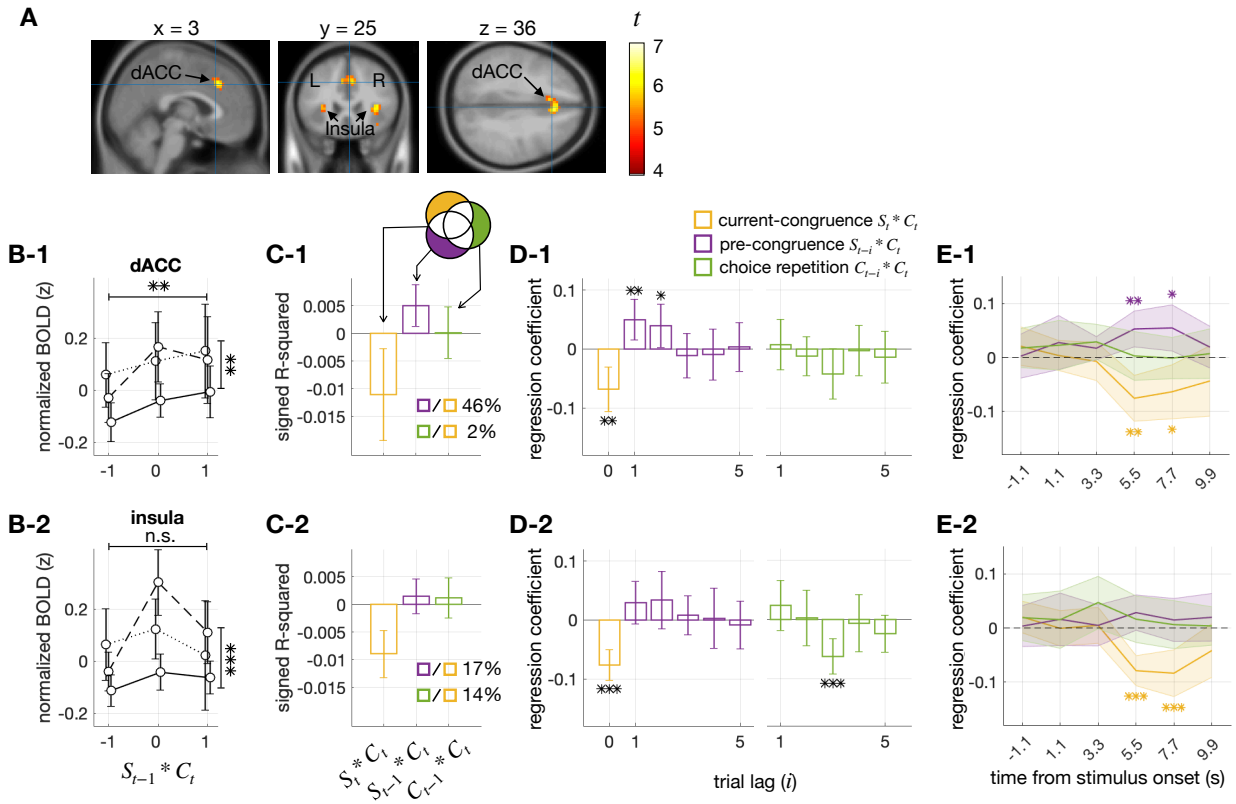

Figure 7. Pre-congruence and current-congruence effects on cortical activity

(A) Definition of the cortical regions of interest (ROIs). The ROIs were defined by carrying out a two-sided Student's t-test with a false discovery rate of 0.05 and identifying voxel clusters with a minimum number of 15 contiguous voxels whose BOLD responses significantly predicted the model estimates of decision uncertainty. The *t*-statistics of these clusters are overlaid on the sagittal (left), coronal (middle), and axial (right) slices of the anatomical image of a template brain. The panels below (B-E) present the results for the ROIs in the dACC and insula in the top (1) and bottom (2) rows, respectively.

(B) Changes in BOLD responses as a function of the pre-congruence and current-congruence factors. The format is identical to that of Figure 4A. Note that both congruence effects were significant in the dACC (1), whereas only the current-congruence effect was significant in the insula (2).

(C) Explained variance analysis. The format is identical to that of Figure 4B. Note that the ratio of the variances explained by the pre-congruence to the variance explained by the current-congruence factor was more substantial in the dACC (1) than in the insula (2).

(D) Multiple regression analysis. The format is identical to that of Figure 4C. Note that the regression coefficients of the pre-congruence factor were significant in the dACC (1) but not in the insula.

(E) Time courses of the current-congruence, pre-congruence, and choice-repetition effects on cortical activity. The regression coefficients of the current-congruence (yellow; $S_t * C_t$), pre-congruence (purple; $S_{t-1} * C_t$), and choice-repetition (green; $C_{t-1} * C_t$) factors are plotted over the time frames relative to the stimulus onset. Note that the emergence of the significant pre-congruence effect is aligned with the time frames associated with decision-making only in the dACC not in the insula.

See also Figure S2,3 and Table S1.

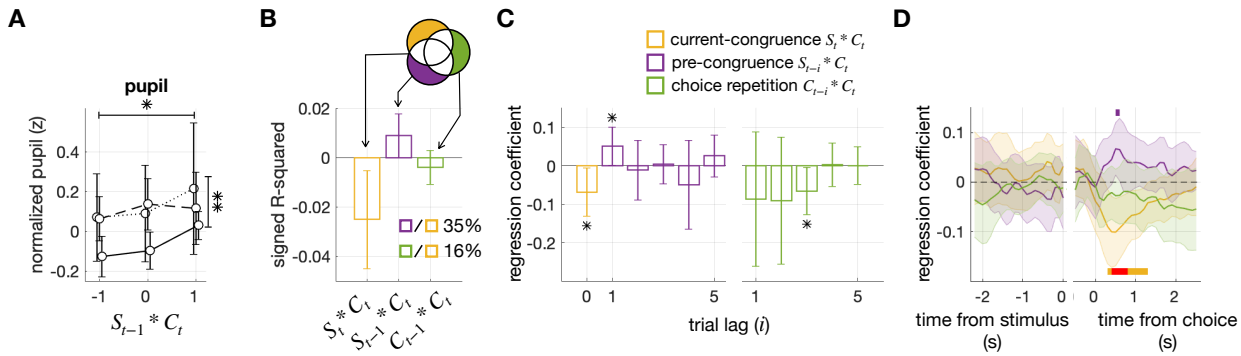

Figure 8. Pre-congruence and current-congruence effects on pupil size

(A) Changes in pupil size as a function of the pre-congruence and current-congruence factors. The format is identical to that of Figure 4A.

(B) Explained variance analysis. The format is identical to that of Figure 4B.

(C) Multiple regression analysis. The format is identical to that of Figure 4C.

(D) Time courses of the current-congruence, pre-congruence, and choice-repetition effects on pupil size. The regression coefficients of the current-congruence (yellow; $S_t * C_t$), pre-congruence (purple; $S_{t-1} * C_t$), and choice-repetition (green; $C_{t-1} * C_t$) factors are plotted over the time frames relative to the stimulus onset. The colored horizontal bars demarcate the time points where the regression coefficients of the pre-congruence and current congruence factors were significant (two-sided Student's *t*-test; purple, $P < 0.05$ for the pre-congruence factor; yellow, $P < 0.05$ and red, $P < 0.01$ for the current-congruence factor).

See also Figure S2.

MAIN TABLES AND LEGENDS

dataset name (number of subjects)		stimulus	stimulus number	trial-to-trial feedback	inter-trial interval (s)
main dataset	fMRI (18)	ring	3	x	13.2
	eye tracking (23)	ring	3	x	13.2
fine-grained pitch (58)		pitch	continuous	x	2
fine-grained ring (58)		ring	continuous	x	2
feedback (30)		ring	5	o	2.5

Table 1. Task protocols for the five datasets used in the current study

STAR METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and analyzed data	This paper	https://github.com/Heeseung-Lee/iScience-2023.git
Software and algorithms		
MATLAB	MathWorks	https://www.mathworks.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact Sang-Hun Lee (visionsl@snu.ac.kr).

Materials availability

Besides data and MATLAB codes, this study did not generate any new reagents or materials.

Data and code availability

- All data have been deposited at GitHub and are publicly available. The DOI is listed in the key resources table.
- All original code has been deposited at GitHub and is publicly available. The DOI is listed in the key resource table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We utilized three datasets used for previous papers of our group[12,75-77] and two new datasets using fine-grained size and pitch stimuli. The fMRI dataset[12] consisted of 18 participants (9 females) aged 20-30 years. The eye tracking dataset[75,76] comprised 23 participants (11 females) aged 18-36 years. The trial-to-trial feedback dataset[77] included 30 participants (13 females) aged 18-30 years. The two new datasets using fine-grained stimuli shared the same subjects which included 58 (38 females) aged 18-34 years. The experimental procedures were approved by the Research Ethics Committee of Seoul National University, and all participants provided informed consent and were unaware of the study's objectives.

METHOD DETAILS

Task procedure

The main data set used the following task procedure (Figure 2A): Participants were instructed to fixate at the center of the screen and classify a brief (0.3s) ring-shaped stimulus as ‘small’ or ‘large’ within 1.5s after stimulus onset by pressing the left or right key, respectively. The timing and identity of each key press were recorded. Trials were separated by 13.2s, and participants were given feedback on their performance after each run of 26 trials.

Before the main runs, participants completed 54 practice trials followed by 180 threshold-calibration trials. During the threshold-calibration trials, which were separated by 2.7s, participants received trial-to-trial feedback based on the boundary with a radius of 2.84°. A Weibull function was fit to the psychometric curves obtained from the threshold-calibration trials using a maximum-likelihood procedure. The size threshold $\Delta°$ (i.e., the size difference between medium ring and large or small ring) associated with a 70.7% correct proportion was estimated from the fitted Weibull function. The mean and standard deviation of the estimated size threshold were 0.023 and 0.0078, respectively. One of three

estimated ring sizes ($2.84 - \Delta°, 2.84°, 2.84 + \Delta°$) was shown on each trial. Participants had extensive training on the task before participating in the main experiments.

Two of the three supplementary datasets were collected from 58 participants, who performed pitch and ring-size classification tasks on separate sessions, respectively, with 315 trials using fine-grained stimuli randomly sampled from normal distributions with an inter-trial interval of 2s and without trial-to-trial feedback for each task. The other dataset was collected from 30 participants, who performed the same classification task over 5 daily sessions with 1,700 trials using ring stimuli of 5 discrete sizes ($3.84°, 3.92°, 4.00°, 4.08°, 4.16°$), with trial-to-trial feedback and an inter-trial interval of 2.5s.

A Bayesian model for boundary updating (BMBU)

The generative model

The generative model is the observers' causal account for noisy sensory measurements. Specifically, the true size of the ring, S, is the cause of a noisy measurement, m_t, on the current trial. This measurement becomes noisier as more trials pass, resulting in a noisier measurement of the value of S on trial $t - i$, which is denoted as r_{t-i} (Figure 3A). The generative model can be defined using three probabilistic terms: the prior probability of S, $p(S)$, the likelihood of S given m_t, $p(m_t|S)$, and the likelihood of S given r_{t-i}, $p(r_{t-i}|S)$. All three terms are modeled as normal distribution functions, which are characterized by mean and standard deviation parameters, μ and σ. Specifically, the prior is characterized by μ_0 and σ_0, the likelihood for m_t is characterized by μ_{m_t} and σ_{m_t}, and the likelihood for r_{t-i} is characterized by $\mu_{r_{t-i}}$ and $\sigma_{r_{t-i}}$. The mean parameters of the likelihood functions are identical to the observed sensory measurements, m_t and r_{t-i}. Therefore, the four parameters that need to be learned are μ_0, σ_0, σ_{m_t}, and $\sigma_{r_{t-i}}$. The parameter σ_{m_t} is assumed to be constant across different values of m_t and across trials and is therefore reduced to a constant value σ_m. The parameter $\sigma_{r_{t-i}}$ is assumed to increase with the number of elapsed trials, and it is modeled as a parametric function of κ, where $\kappa > 0$: $\sigma_{r_{t-i}} = \sigma_m(1+\kappa)^i$. As a result, the generative model is fully specified by the four parameters, which are represented as $\Theta = [\mu_0, \sigma_0, \sigma_m, \kappa]$.

Stimulus inference (s)

To estimate the value of S on a current trial based on the sensory measurement m_t, a Bayesian approach was employed, where the mean of the posterior distribution of S given m_t, denoted as $p(S|m_t)$, was considered as the Bayesian estimate s_t. This posterior distribution is a conjugate normal distribution of the prior and the likelihood of S given the sensory evidence m_t. The Bayesian estimate s_t and its standard deviation σ_{s_t} were calculated as follows:

$$s_t = \frac{\sigma_0^2 m_t + \sigma_m^2 \mu_0}{\sigma_0^2 + \sigma_m^2}; \quad \sigma_{s_t} = \frac{\sigma_0 \sigma_m}{\sqrt{\sigma_0^2 + \sigma_m^2}}.$$

Class boundary inference (b)

The Bayesian observer estimates the value of the class boundary on a current trial, b_t, by taking the mean of the posterior function based on a given set of retrieved sensory measurements, $\vec{r}_t = [r_{t-1}, r_{t-2}, \dots r_{t-n}]$: $(S|\vec{r}_t)$, The maximum number of measurements that can be retrieved, n, is set to 7 because it is much longer than the effective trial lags of the previous stimulus effect (Figure 2). The posterior function, $p(S|\vec{r}_t)$, is a conjugate normal distribution of the prior and likelihoods of S given the evidence $\vec{r}_t$:

$$\begin{aligned} p(S|\vec{r}_t) &\propto p(\vec{r}_t|S)p(S) \\ &= p(r_{t-1}|S)p(r_{t-2}|S) \dots p(r_{t-7}|S)p(S) \end{aligned}$$

, whose mean and standard deviation were calculated based on the knowledge of how the retrieved stimulus becomes noisier as trials elapse:

$$b_t = \beta_0\mu_0 + \sum_{i=1}^{7}\beta_i r_{t-i};\ \sigma_{b_t} = \sqrt{\beta_0^2\sigma_0^2 + \sum_{i=1}^{7}\beta_i^2\sigma_{r_{t-i}}^2}$$

, where $\beta_0 = \frac{\sigma_0^{-2}}{\sigma_0^{-2}+\sum_{i=1}^{7}\sigma_{r_{t-i}}^{-2}}$ and $\beta_i = \frac{\sigma_{r_{t-i}}^{-2}}{\sigma_0^{-2}+\sum_{i=1}^{7}\sigma_{r_{t-i}}^{-2}}$.

Deduction of choice (c) and decision uncertainty (u)

For each trial, if s_t exceeds b_t, the simulated choice c_t is recorded as 'large' (or 1); otherwise, it is recorded as 'small' (or -1). In addition, we defined the decision uncertainty, u_t, which represents the likelihood that the current decision will be incorrect. Specifically, u_t is calculated as follows:

$$u_t = \Phi\left[\frac{-|s_t - b_t|}{\sqrt{\sigma_{s_t}^2 + \sigma_{b_t}^2}}\right],$$

where Φ stands for the standard normal cumulative function.

A constant-boundary model

The constant-boundary model comprises two parameters, the bias of the class boundary, μ_0, and the measurement noise, σ_m. In this model, we assumed that stimulus estimates, s_t, were sampled from a normal distribution, $\mathcal{N}(S_t, \sigma_m)$, where S_t represents the physical size of the stimulus. The uncertainty in each stimulus sample was assumed to be equal to σ_m, that is, $\sigma_{s_t} = \sigma_m$. Furthermore, we assumed that the boundary estimate, b_t, was a constant value, μ_0, with $\sigma_{b_t} = 0$.

BOLD signal

MRI data were collected using a 3 Tesla Siemens Tim Trio scanner equipped with a 12-channel Head Matrix coil at the Seoul National University Brain Imaging Center. Stimuli were generated using MATLAB (MathWorks) in conjunction with MGL (http://justingardner.net/mgl) on a Macintosh computer. Observers looked through an angled mirror attached to the head coil to view stimuli displayed via an LCD projector (Canon XEED SX60) onto a back-projection screen at the end of the magnet bore at a viewing distance of 87cm, yielding a field of view of 22×17°.

For each observer, we acquired three types of MRI images, as follows: (1) 3D, T1-weighted, whole-brain images (MPRAGE; resolution, 1×1×1mm; field of view (FOV), 256mm; repetition time (TR), 1.9s; time for inversion, 700ms; time to echo (TE), 2.36; and flip angle (FA), 9°), (2) 2D, T1-weighted, in-plane images (MPRAGE; resolution, 1.08×1.08×3.3mm; TR, 1.5s; T1, 700ms; TE, 2.79ms; and FA, 9°), and (3) 2D, T2*-weighted, functional images (gradient EPI; TR, 2.2s; TE, 40ms; FA, 73º; FOV, 208mm; image matrix, 90×90; slice thickness, 3mm with 10% space gap; slice, 32 oblique transfers slices; bandwidth, 790Hz/ps; and effective voxel size, 3.25×3.25×3.3mm).

The images of individual observers were normalized to the MNI template using the following steps: motion correction, coregistration to whole-brain anatomical images via the in-plane images[78], spike elimination, slice timing correction, normalization using the SPM DARTEL Toolbox[79] to 3×3×3mm voxel size, and smoothing with 8×8×8mm full-width half-maximum Gaussian kernel. All the procedures were implemented with SPM8 and SPM12 (http://www.fil.ion.ucl.ac.uk.spm)[80,81], except for spike elimination for which we used the AFNI toolbox[82]. The first six frames of each functional scan (the first trial of each run) were discarded to allow hemodynamic responses to reach a steady state. Then, the normalized BOLD time series at each voxel, each run, and each subject were preprocessed using linear detrending and high-pass filtering (132s cut-off frequency with a Butterworth filter), conversion into percent-change signals, correction for non-neural nuisance signals by regressing out the mean BOLD

activity of the cerebrospinal fluid (CSF), and standardization by subtracting the mean activity and by dividing the standard deviation.

To define the anatomical mask of CSF, probability tissue maps for individual participants were generated from T1-weighted images, normalized to the MNI space, and smoothed as were done for the functional images by using SPM12, and then averaged across participants. Finally, the location of CSF was defined as respective groups of voxels in which the probability was more than 0.5.

Unfortunately, in a few of the sessions, functional images did not cover the entire brain. We investigated a brain region only when it does not loss any session. As a result, some voxels in the temporal pole, ventral orbitofrontal, and posterior cerebellum were excluded from data analysis.

Pupillometry

Stimuli were presented in a dimly lit room on a gamma-linearized 22-inch CRT monitor (Totoku CV921X CRT monitor) operating at vertical refresh rate of 180Hz and a spatial resolution of 800×600 pixels. Stimuli were generated using MATLAB (MathWorks) in conjunction with MGL (http://justingardner.net/mgl) on a Macintosh computer. Observers viewed the monitor at a distance of 90cm while their binocular eye positions were sampled at 500Hz by an infrared eye tracker (EyeLink 1000 Desktop Mount, SR Research; instrument noise, 0.01° RMS). The LED illuminator and camera were positioned side by side, at a distance of 65cm from the observer, and angled toward the observer's face to ensure that infrared light illuminated both eyes and was being reflected from both eyes and imaged on the camera sensor.

When measuring gaze position using video-based methods, eye blinks can interfere with the accuracy of the data, as pupil and gaze information are not available during these periods. We identified eye blinks based on three criteria: (i) missing pupil data for either eye, (ii) pupil-size measurements with unrealistically large fluctuations (>50 units per sample), or (iii) substantial deviation (>20°) of gaze position from the screen center. Data collected immediately before and after an eye blink (±200 ms) were likely contaminated and therefore we linearly interpolated the pupil-sizes and the three gaze positions (x, y, and d $(=\sqrt{x^2+y^2})$) before and after the blink events.

To minimize confounding effects from gaze position on pupil-size measurements, we processed the blink-free samples of pupil-size and three gaze positions time series by linear detrending, band-pass filtering (0.01Hz to 4Hz cut-off frequency with a Butterworth filter), and resampling to 10Hz. The resampled time series of pupil-size was then orthogonalized from the three resampled gaze positions up to their fourth powers (total 12 regressors) to extract any confounding in the pupil-size originating from the gaze position[75]. The orthogonalized pupil-size time series was then standardized by subtracting the mean and dividing the standard deviation of the time series. The pupil-sizes of the eyes were averaged, and trials were epoched and baseline corrected by subtracting the baseline pupil-size averaged across 0ms to 500ms from the stimulus onset. We chose this baseline time window because we assumed that the decision-related pupil signal would be initiated between -200ms to 300ms from the stimulus onset and because the latency of the empirical responses was around 200ms[83]. Finally, the baseline corrected pupil-sizes were aligned by the stimulus onset or the choice following previous studies[18,84].

QUANTIFICATION AND STATISTICAL ANALYSIS

Model fitting

To estimate the parameters of the generative model for each human participant, we used maximum likelihood estimation. Specifically, we sought to find the values of $\Theta = [\mu_0, \sigma_0, \sigma_m, \kappa]$ that maximized the sum of log-likelihoods across T individual choices made by the human observer, $\vec{C}_T = [C_1, C_2, \dots, C_T]$:

$$\widehat{\Theta} = \arg\max_{\Theta} \sum_{t=1}^{T} \log p(C_t|\Theta).$$

For each participant, we estimated the parameters of the generative model in the following steps. Initially, we used a MATLAB function, fminseachbnd, to find local minima of parameters with the iterative evaluation number set to 50. To acquire a set of candidate parameter estimates, we repeated this step by choosing 1,000 different initial parameter sets randomly sampled within uniform prior bounds. In the second step, we selected the top 20 candidate sets based on goodness-of-fit (sum of log-likelihoods) and searched for the minima using each of those 20 sets as initial parameters. We increased the iterative evaluation number to 100,000 and set tolerances of function and parameters to 10^{-7} for reliable estimation. Using the parameters fitted via the second step, we repeated the second step once more. Then, we selected the parameter set that showed the largest sum of likelihoods as the final parameter estimates. We excluded the first trial of each run and trials with reaction times less than 0.3 seconds from further analyses. This exclusion is necessary because the first trial of each run does not have its previous trial, which is necessary for investigating the repulsive influence. Additionally, a response made during the stimulus's initial presentation (0~0.3s) may be too hasty to reflect a normal cognitive decision-making process.

Definition of congruence

The term 'congruence' refers to the product of the variable of interest and the current choice, where choices were represented as 1 for 'large' and -1 for 'small' and stimuli were represented as 1, 0, and -1 for large, medium, and small sizes, respectively. For instance, if the current choice was 'large (1)' and the previous stimulus was a 'small' size (-1), then the congruence would be -1 (1 * -1). In the supplementary data where there were five stimuli, the stimuli were encoded as [-2, -1, 0, 1, 2].

Estimation of decision uncertainty

By fitting the model parameters ($\widehat{\Theta} = [\hat{\mu}_0, \hat{\sigma}_0, \hat{\sigma}_m, \hat{\kappa}]$) separately for each human participant, we were able to create a Bayesian observer tailored to each individual. By conducting the experiment again with these Bayesian observers using the same stimulus sequences presented to their human counterparts, we obtained a sufficient number (10^6 repetitions) of simulated choices, c_t, and decision uncertainty values, u_t. These values were determined based on the corresponding number of stimulus estimates, s_t, and boundary estimates, b_t, for each Bayesian observer. Finally, we took the averages across those 10^6 simulations as the final outcomes. When estimating u_t for the observed choice C_t, we only included simulation outcomes where the simulated choice c_t matched the observed choice C_t.

Variable recovery test

To ascertain the validity of our procedure of estimating the trial-to-trial states of decision uncertainty u_t, we ran simulations to check how accurately the true states of u_t are recovered by the procedure. This variable recovery test is crucial because the accuracy of the model estimates of u_t is a prerequisite for correctly identifying the brain regions where BOLD responses correlate with decision uncertainty and the time points when pupil size measurements correlate with decision uncertainty. The variable recovery test was carried out with the following aim and procedure.

The variable recovery test aimed to evaluate the correspondence between the *true* and *recovered* states of u_t on a trial-to-trial basis for each of the individual observers, who are likely to vary in parameters. To achieve this aim, it is important to generate a comprehensive set of *true* trial-to-trial states of u_t using a realistic set of model parameters that reflect the range of parameters of actual observers. To do so, we first created 256 different sets of parameter values by taking the possible combinations of the four different values of each of the four model parameters ($\mu_0, \sigma_0, \sigma_m, \kappa$), where the four different values corresponded to the 20, 40, 60, and 80 percentiles of the parameter values fitted to the choices of the observers (N=41) who participated in either the fMRI (N=18) or the eye-tracking (N=23) experiments. Second, we acquired the synthetic choices and the *true* states of decision uncertainty (u_t) by plugging one of the 256 parameter sets into BMBU and simulating it on the actual stimulus sequence used for one of the 41 observers. Third, we fitted the 4 parameters of BMBU to those synthetic choices for each observer.

Fourth, we simulated the *recovered* states of u_t using those fitted model parameters on the identical stimulus sequence used for generating the *true* states of u_t. Fifth, we calculated the R-squared between the *true* and *recovered* u_ts to assess how reliably our model fitting procedure can recover the *true* states of the model variables. Sixth, we repeated the second to fifth steps for all the remained 255 parameter sets and used the R-squared averaged across all the 256 parameter sets as the performance measure of the recovery test for the one observer of interest. Finally, we repeated the second to sixth procedures for all the remaining 40 observers and used the average R-squared and its confidence interval across all the 41 observers as the performance measure of the recovery test.

Definition of the signed R-squared

The signed R-squared is a measure that takes into account the direction of the relationship between a variable of interest and other variables in a multiple regression model. It is calculated by multiplying the sign of the regression coefficient with the uniquely explained variance of the variable. The uniquely explained variance is obtained through variance partitioning analysis[85,86], which involves comparing the R-squared values of the full regression model that includes all variables and the reduced regression model that excludes the variable of interest. For instance, to calculate the signed R-squared of variable x, we first calculate the R-squared of the full model (R^2_{xyz}) that includes variables x, y, and z. Next, we compute the R-squared of the reduced model (R^2_{yz}) that includes variables y and z only. The uniquely explained variance of x ($R^2_{x,yz}$) is obtained by subtracting R^2_{yz} from R^2_{xyz}. Finally, the signed R-squared of x is defined as $sign(\beta_x)R^2_{x,yz}$, where $sign(\beta_x)$ is the sign of the regression coefficient β_x of the multiple regression model with variables x, y, and z as regressors.

Searching for the brain signals correlated with u_t

To explore the brain regions that showed a correlation between BOLD signals and u_t, we conducted a regression analysis for each participant and voxel. The response variable for the regression model was the preprocessed BOLD signals concatenated across runs. We used three explanatory vectors in the regression model. The first vector was created by convolving the canonical hemodynamic function with u_t for each trial. The second vector was created by convolving the canonical hemodynamic function with a constant value of 1, and both vectors were standardized. The third vector contained only a single value of 1. The first regression coefficient indicated how well u_t predicted the BOLD signal for each voxel, and by calculating this coefficient for every voxel, we generated a map of regression coefficients for each participant and the entire brain.

To determine whether the average coefficient of u_t across participants significant, we performed a two-sided Student *t*-test on each voxel. We then corrected the *P* values of the entire brain for false discovery rate (FDR) [87] (Figure S3). The total number of voxels in the entire brain was 90481, and we defined the brain areas significantly correlated with u_t as the voxel clusters covering a region larger than 15 contiguous voxels and having FDR-corrected *P* values less than 0.05. For the ROI analysis, we averaged the preprocessed BOLD signals across individual voxels within an ROI.

REFERENCES

1. Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. Linguistics and philosophy *30*, 1-45.
2. Rips, L.J., and Turnbull, W. (1980). How big is big? Relative and absolute properties in memory. Cognition *8*, 145-174.
3. Solt, S. (2015). Vagueness and imprecision: Empirical foundations. Annu. Rev. Linguist. *1*, 107-127.
4. Tribushinina, E. (2011). Once again on norms and comparison classes. Linguistics *49*, 525-553.
5. Lages, M., and Treisman, M. (2010). A criterion setting theory of discrimination learning that accounts for anisotropies and context effects. Seeing and perceiving *23*, 401-434.
6. Morgan, M., Watamaniuk, S., and McKee, S. (2000). The use of an implicit standard for measuring discrimination thresholds. Vision research *40*, 2341-2349.
7. Raviv, O., Lieder, I., Loewenstein, Y., and Ahissar, M. (2014). Contradictory behavioral biases result from the influence of past stimuli on perception. PLOS Comput Biol *10*, e1003948.
8. Hachen, I., Reinartz, S., Brasselet, R., Stroligo, A., and Diamond, M. (2021). Dynamics of history-dependent perceptual judgment. Nature communications *12*, 1-15.
9. Lages, M., and Treisman, M. (1998). Spatial frequency discrimination: visual long-term memory or criterion setting? Vision research *38*, 557-572.
10. Treisman, M., and Williams, T.C. (1984). A theory of criterion setting with an application to sequential dependencies. Psychological Review *91*, 68.
11. White, C.N., Mumford, J.A., and Poldrack, R.A. (2012). Perceptual criteria in the human brain. Journal of Neuroscience *32*, 16716-16724.
12. Lee, H., Lee, H.-J., Choe, K.W., and Lee, S.-H. (2023). Neural Evidence for Boundary Updating as the Source of the Repulsive Bias in Classification. The Journal of Neuroscience *43*, 4664-4683. 10.1523/jneurosci.0166-23.2023.
13. Bosch, E., Fritsche, M., Ehinger, B.V., and de Lange, F.P. (2020). Opposite effects of choice history and evidence history resolve a paradox of sequential choice bias. Journal of vision *20*, 9-9.
14. Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature *455*, 227. 10.1038/nature07200 https://www.nature.com/articles/nature07200#supplementary-information.
15. Hebart, M.N., Schriever, Y., Donner, T.H., and Haynes, J.-D. (2014). The relationship between perceptual decision variables and confidence in the human brain. Cerebral Cortex *26*, 118-130.
16. Grinband, J., Hirsch, J., and Ferrera, V.P. (2006). A neural representation of categorization uncertainty in the human brain. Neuron *49*, 757-763.
17. Braun, A., Urai, A.E., and Donner, T.H. (2018). Adaptive history biases result from confidence-weighted accumulation of past choices. Journal of Neuroscience *38*, 2418-2429.
18. Urai, A.E., Braun, A., and Donner, T.H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. Nature communications *8*, 14637.
19. Hebart, M.N., Donner, T.H., and Haynes, J.-D. (2012). Human visual and parietal cortex encode visual choices independent of motor plans. Neuroimage *63*, 1393-1403.
20. Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. Nature neuroscience *10*, 1214.
21. Drugowitsch, J., Mendonça, A.G., Mainen, Z.F., and Pouget, A. (2019). Learning optimal decisions with confidence. Proceedings of the National Academy of Sciences *116*, 24872-24880.
22. Shenhav, A., Cohen, J.D., and Botvinick, M.M. (2016). Dorsal anterior cingulate cortex and the value of control. Nature neuroscience *19*, 1286-1291.
23. Pouget, A., Drugowitsch, J., and Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. Nature neuroscience *19*, 366-374.

24. Fietz, J., Pöhlchen, D., Binder, F.P., Group, B.W., Czisch, M., Sämann, P.G., and Spoormaker, V.I. (2022). Pupillometry tracks cognitive load and salience network activity in a working memory functional magnetic resonance imaging task. Human Brain Mapping *43*, 665-680.
25. de Gee, J.W., Tsetsos, K., Schwabe, L., Urai, A.E., McCormick, D., McGinley, M.J., and Donner, T.H. (2020). Pupil-linked phasic arousal predicts a reduction of choice bias across species and decision domains. Elife *9*, e54014.
26. Pajkossy, P., Gesztesi, G., and Racsmány, M. (2023). How uncertain are you? Disentangling expected and unexpected uncertainty in pupil-linked brain arousal during reversal learning. Cognitive, Affective, & Behavioral Neuroscience, 1-22.
27. Burlingham, C.S., Mirbagheri, S., and Heeger, D.J. (2022). A unified model of the task-evoked pupil response. Science Advances *8*, eabi9979.
28. Wang, H., Ortega, H.K., Atilgan, H., Murphy, C.E., and Kwan, A.C. (2022). Pupil correlates of decision variables in mice playing a competitive mixed-strategy game. Eneuro *9*.
29. Graves, J.E., Egré, P., Pressnitzer, D., and de Gardelle, V. (2021). An implicit representation of stimulus ambiguity in pupil size. Proceedings of the National Academy of Sciences *118*, e2107997118.
30. de Gee, J.W., Correa, C.M., Weaver, M., Donner, T.H., and van Gaal, S. (2021). Pupil dilation and the slow wave ERP reflect surprise about choice outcome resulting from intrinsic variability in decision confidence. Cerebral Cortex *31*, 3565-3578.
31. Ganea, D.A., Bexter, A., Günther, M., Gardères, P.-M., Kampa, B.M., and Haiss, F. (2020). Pupillary dilations of mice performing a vibrotactile discrimination task reflect task engagement and response confidence. Frontiers in Behavioral Neuroscience *14*, 159.
32. Filipowicz, A.L., Glaze, C.M., Kable, J.W., and Gold, J.I. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. Elife *9*, e57872.
33. van Kempen, J., Loughnane, G.M., Newman, D.P., Kelly, S.P., Thiele, A., O'Connell, R.G., and Bellgrove, M.A. (2019). Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. Elife *8*, e42541.
34. Colizoli, O., de Gee, J.W., Urai, A.E., and Donner, T.H. (2018). Task-evoked pupil responses reflect internal belief states. Scientific reports *8*, 13702.
35. Fleck, M.S., Daselaar, S.M., Dobbins, I.G., and Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. Cerebral Cortex *16*, 1623-1630.
36. Fleming, S.M., Huijgen, J., and Dolan, R.J. (2012). Prefrontal contributions to metacognition in perceptual decision making. Journal of Neuroscience *32*, 6117-6125.
37. Boldt, A., and Gilbert, S.J. (2022). Partially overlapping neural correlates of metacognitive monitoring and metacognitive control. Journal of Neuroscience *42*, 3622-3635.
38. Bang, D., and Fleming, S.M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. Proceedings of the National Academy of Sciences *115*, 6082-6087.
39. Sherman, M.T., Seth, A.K., and Kanai, R. (2016). Predictions shape confidence in right inferior frontal gyrus. Journal of Neuroscience *36*, 10323-10336.
40. Braunlich, K., and Seger, C.A. (2016). Categorical evidence, confidence, and urgency during probabilistic categorization. Neuroimage *125*, 941-952.
41. Morales, J., Lau, H., and Fleming, S.M. (2018). Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. Journal of Neuroscience *38*, 3534-3546.
42. Heereman, J., Walter, H., and Heekeren, H.R. (2015). A task-independent neural representation of subjective certainty in visual perception. Frontiers in human neuroscience *9*, 551.
43. Hilgenstock, R., Weiss, T., and Witte, O.W. (2014). You'd better think twice: post-decision perceptual confidence. NeuroImage *99*, 323-331.
44. Gherman, S., and Philiastides, M.G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. elife *7*, e38293.

45. Buracas, G.T., and Boynton, G.M. (2002). Efficient design of event-related fMRI experiments using M-sequences. Neuroimage *16*, 801-813.
46. Zokaei, N., Burnett Heyes, S., Gorgoraptis, N., Budhdeo, S., and Husain, M. (2015). Working memory recall precision is a more sensitive index than span. Journal of Neuropsychology *9*, 319-329.
47. Gorgoraptis, N., Catalao, R.F., Bays, P.M., and Husain, M. (2011). Dynamic updating of working memory resources for visual objects. Journal of Neuroscience *31*, 8502-8511.
48. Sanders, J.I., Hangya, B., and Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. Neuron *90*, 499-506.
49. Lak, A., Costa, G.M., Romberg, E., Koulakov, A.A., Mainen, Z.F., and Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron *84*, 190-201.
50. Rahnev, D., Koizumi, A., McCurdy, L.Y., D'Esposito, M., and Lau, H. (2015). Confidence leak in perceptual decision making. Psychological science *26*, 1664-1680.
51. Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., and Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. Nature neuroscience *16*, 749-755.
52. Kiani, R., and Shadlen, M.N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. science *324*, 759-764.
53. Teichert, T., Yu, D., and Ferrera, V.P. (2014). Performance monitoring in monkey frontal eye field. Journal of Neuroscience *34*, 1657-1671.
54. Cho, R.Y., Nystrom, L.E., Brown, E.T., Jones, A.D., Braver, T.S., Holmes, P.J., and Cohen, J.D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. Cognitive, Affective, & Behavioral Neuroscience *2*, 283-299.
55. Kristjánsson, Á., and Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. Attention, Perception, & Psychophysics *72*, 5-18.
56. Galluzzi, F., Benedetto, A., Cicchini, G.M., and Burr, D.C. (2022). Visual priming and serial dependence are mediated by separate mechanisms. Journal of Vision *22*, 1-1.
57. Jentzsch, I., and Dudschig, C. (2009). Short article: Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. Quarterly Journal of Experimental Psychology *62*, 209-218.
58. Urai, A.E., De Gee, J.W., Tsetsos, K., and Donner, T.H. (2019). Choice history biases subsequent evidence accumulation. Elife *8*, e46331.
59. Zhang, H., and Alais, D. (2020). Individual difference in serial dependence results from opposite influences of perceptual choices and motor responses. Journal of Vision *20*, 2-2.
60. Busse, L., Ayaz, A., Dhruv, N.T., Katzner, S., Saleem, A.B., Schölvinck, M.L., Zaharia, A.D., and Carandini, M. (2011). The detection of visual contrast in the behaving mouse. Journal of Neuroscience *31*, 11351-11361.
61. Odoemene, O., Pisupati, S., Nguyen, H., and Churchland, A.K. (2018). Visual evidence accumulation guides decision-making in unrestrained mice. Journal of Neuroscience *38*, 10143-10155.
62. Akaishi, R., Umeda, K., Nagase, A., and Sakai, K. (2014). Autonomous mechanism of internal choice estimate underlies decision inertia. Neuron *81*, 195-206.
63. Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., Urai, A.E., Donner, T.H., Carandini, M., Tonegawa, S., and Uchida, N. (2020). Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. ELife *9*, e49834.
64. Cicchini, G.M., Mikellidou, K., and Burr, D. (2017). Serial dependencies act directly on perception. Journal of vision *17*, 6-6.
65. Fründ, I., Wichmann, F.A., and Macke, J.H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. Journal of vision *14*, 9-9.

66. Cicchini, G.M., Mikellidou, K., and Burr, D.C. (2018). The functional role of serial dependence. Proceedings of the Royal Society B *285*, 20181722.
67. Shekhar, M., and Rahnev, D. (2021). Sources of metacognitive inefficiency. Trends in Cognitive Sciences *25*, 12-23.
68. Lamichhane, B., Adhikari, B.M., and Dhamala, M. (2016). The activity in the anterior insulae is modulated by perceptual decision-making difficulty. Neuroscience *327*, 79-94.
69. Shenhav, A., Straccia, M.A., Cohen, J.D., and Botvinick, M.M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. Nature neuroscience *17*, 1249.
70. Green, D.M., and Swets, J.A. (1966). Signal detection theory and psychophysics (John Wiley).
71. Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annual review of neuroscience *30*.
72. Norton, E.H., Fleming, S.M., Daw, N.D., and Landy, M.S. (2017). Suboptimal criterion learning in static and dynamic environments. PLoS computational biology *13*, e1005304.
73. Maddox, W.T. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. Journal of the experimental analysis of behavior *78*, 567-595.
74. Luo, T.Z., and Maunsell, J.H. (2018). Attentional changes in either criterion or sensitivity are associated with robust modulations in lateral prefrontal cortex. Neuron *97*, 1382-1393. e1387.
75. Choe, K.W., Blake, R., and Lee, S.-H. (2016). Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. Vision research *118*, 48-59.
76. Choe, K.W., Blake, R., and Lee, S.-H. (2014). Dissociation between neural signatures of stimulus and choice in population activity of human V1 during perceptual decision-making. Journal of Neuroscience *34*, 2725-2743.
77. Lee, H.-J., Lee, H., Lim, C.Y., Rhim, I., and Lee, S.-H. (2023). What humans learn from corrective feedback for perceptual decision-making. bioRxiv, 2023.2001. 2011.523567.
78. Nestares, O., and Heeger, D.J. (2000). Robust multiresolution alignment of MRI brain volumes. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine *43*, 705-715.
79. Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. Neuroimage *38*, 95-113.
80. Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., and Turner, R. (1996). Movement-related effects in fMRI time-series. Magnetic resonance in medicine *35*, 346-355.
81. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825-841.
82. Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical research *29*, 162-173.
83. Korn, C.W., and Bach, D.R. (2016). A solid frame for the window on cognition: Modeling event-related pupil responses. Journal of vision *16*, 28-28.
84. de Gee, J.W., Knapen, T., and Donner, T.H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. Proceedings of the National Academy of Sciences *111*, E618-E625.
85. Borcard, D., Legendre, P., and Drapeau, P. (1992). Partialling out the spatial component of ecological variation. Ecology *73*, 1045-1055.
86. Perquin, M.N., Heed, T., and Kayser, C. (2022). Variance (un) explained: Experimental conditions and temporal dependencies explain similarly small proportions of reaction time variability in perceptual and cognitive tasks. bioRxiv, 2022.2012. 2022.521656.
87. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological), 289-300.

SUPPLEMENTAL INFORMATION

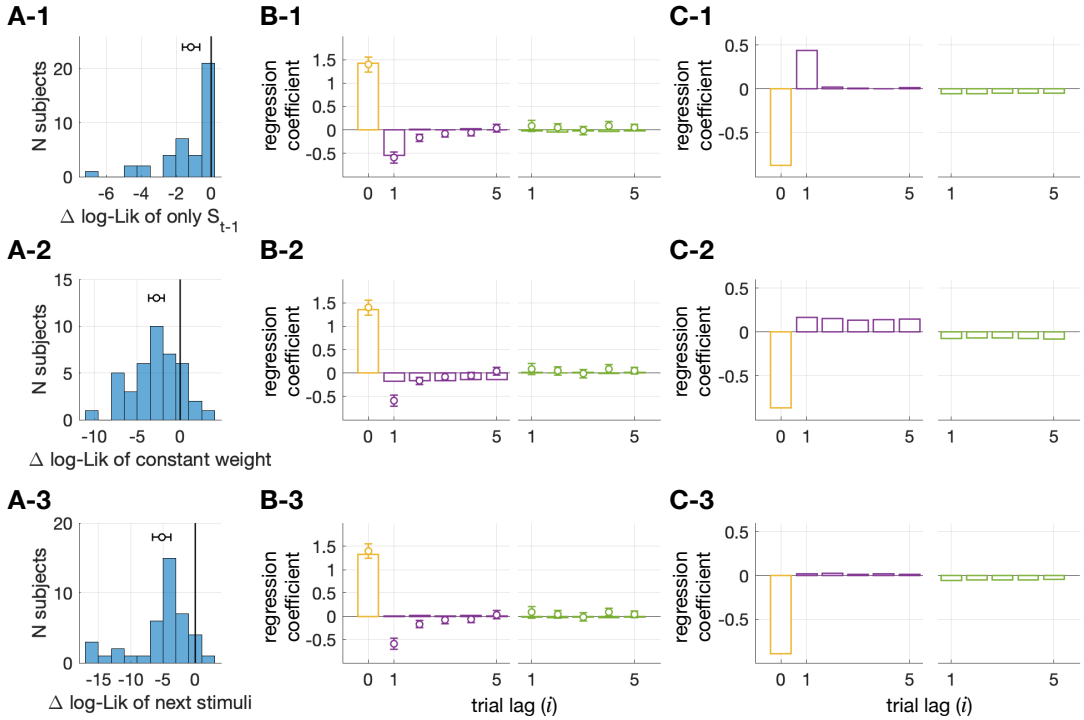

Figure S1. Necessity of the 'memory-decay' assumption in explaining the pre-congruence effect, related to Figure 3

To evaluate the necessity of the assumption that the memory recall of a stimulus becomes increasingly noisier as trials elapse (Equations 1 and 2) in accounting for the pre-congruence effect, we created three variants of the original model (BMBU) by modifying the process of boundary inference, as follows. In the first model variant called 'limited memory access,' the memory recall was assumed to be accessible only for the immediately preceding stimulus ($\beta_i = 0$ for $i > 1$ in Equation 2). In the second model variant called 'no memory decay,' the noisiness of memory recall was assumed not to increase after one trial lag ($\sigma_{r_{t-i}} = \sigma_m(1+\kappa)$ in Equation 1). In the third model variant called 'prospective memory recall,' the boundary inference was based on the future stimuli instead of the past stimuli (r_{t-i} is replaced with r_{t+i} in Equation 2). These three model variants were compared with BMBU for goodness-of-fit evaluation, the results of which are summarized in A. The multiple regression analyses depicted in Figure 2C and 3H were repeated using the three model variants, the results of which are summarized in B and C, respectively. The summarized results of the 'limited memory access,' 'no memory decay,' and 'prospective memory recall' model variants are presented in the top (1), middle (2), and bottom (3) rows of panels.

(A) Goodness-of-fit analysis. The histograms show the frequency (number of observers) distributions of the differences in log-likelihood between BMBU and the model variants. For each observer, the predictive power (AIC) of BMBU was subtracted from those of the model variant. The circles and horizontal error bars represent the means and their 95% confidence intervals of the log-likelihood differences. The log-likelihoods of the model variants were all significantly smaller than that of BMBU ('limited memory access,' $t = -4.7, (P = 2.7 \times 10^{-5})$; no memory decay,' $t = -6.2, (P = 2.2 \times 10^{-7})$; 'prospective memory recall,' $t = -7.3, (P = 7.9 \times 10^{-9})$).

(B) Model simulation of repulsive bias demonstrated in logistic regressions. The coefficients in the multiple logistic regression of the current choice (C_t) onto the current and previous stimuli (S_{t-i}) and the previous choice (C_{t-i}) are plotted in circles and bars for humans and the models, respectively.
(C) Model simulation of pre-congruence and current-congruence effects in linear regressions. The vertical bars represent the coefficients of the multiple linear regression of the simulated decision uncertainty onto the congruences of current choices with the current stimulus ($S_t * C_t$), the previous stimuli ($S_{t-i} * C_t$), and the previous choices ($C_{t-i} * C_t$).

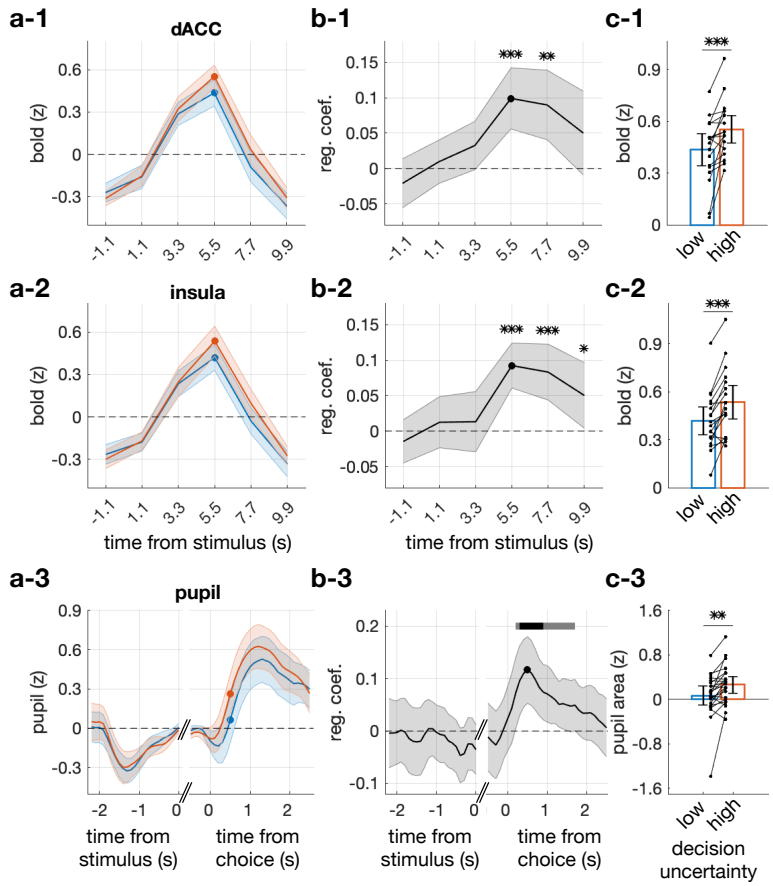

Figure S2. Time courses of the neural and pupillary correlates of the model estimates of decision uncertainty (u_t), related to Figure 7,8

To identify the time points where the neural and pupillary correlates of the model estimates of decision uncertainty (u_t) are maximally pronounced, we carried out the following analyses on the time series of the BOLD responses in the dACC and insula and of the pupil size measurements.

(A) Time courses of BOLD (1,2) and pupil size (3). To visualize the effect of u_t on the BOLD responses and pupil size measurements, we split the trials into the low and high halves in terms of u_t. The colors of the symbols, lines, and shades correspond to the two levels of u_t (blue and red for the low and high levels, respectively).

(B) Time courses of regression coefficients. We linearly regressed the BOLD responses and pupil size measurements onto u_t in each time point and identified the time points where the significance of those regressions reached their maximum values (the black dots). The coefficients are plotted as a function of time relative to the stimulus onset (1-2) or the time of making choices (3). The *P* values of two-sided Student's *t*-test of the coefficients are indicated by the horizontal bar (gray: $*< 0.05$; black: $**< 0.01$).

(C) Differences in BOLD (1,2) and pupil size (3) between the two levels of u_t. In the identified time points in B, we compared the signals between trials of low and high halves of u_t. The color scheme was identical to that used in A. Each dot pair with a line represents a single observer.

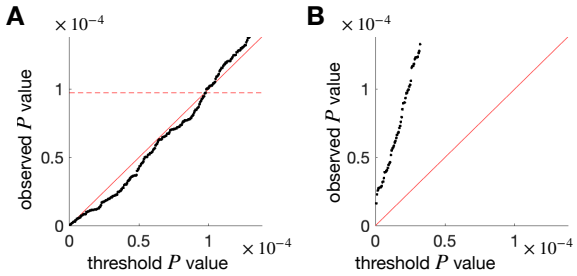

Figure S3. Importance of incorporating the 'boundary updating' process into the definition of decision uncertainty in search of its neural correlate, related to Figure 7

To demonstrate how critical the boundary updating process (as depicted in Figure 3A-E) is for the accurate definition of decision uncertainty, we created a constant-boundary model that lacks this process (see STAR Methods for details) and assessed its effectiveness in discovering BOLD correlates of decision uncertainty. Specifically, we regressed the BOLD responses of all brain voxels onto the model estimates of decision uncertainty as defined by BMBU or the constant-boundary model. We then plotted the lowest *P* values of each voxel against the threshold *P* values to control for the false discovery rate (FDR). By doing so, we could determine the significant voxels after controlling for FDR by defining the critical *P* value, which is the maximum *P* value lower than the FDR threshold (indicated by the horizontal dashed line). When BMBU's estimates of decision uncertainty were used, we found that up to 177 voxels were significant (A), whereas none were significant when the constant-boundary model's estimates of decision uncertainty were used (B).

cortical area	continuous voxel (N)	peak voxel		
		MNI coordinate	t-stat	P-value
dorsal anterior cingulate cortex	64	$[6, 27, 36]$	6.29	8.2×10^{-6}
left insula	27	$[-30,\ 21, 9]$	6.50	5.7×10^{-6}
right insula	66	$[36, 18, 6]$	7.60	7.5×10^{-7}

Table S1. Brain regions where BOLD responses significantly correlate with the model estimates of decision uncertainty (u_t), related to Figure 7

To define these regions, we used two criteria: (1) the regression coefficient between the BOLD responses and the model estimates of decision uncertainty was significant (two-sided Student's t-test $P < 0.05$, after controlling for the false discovery rate), and (2) there were at least 15 contiguous voxels that met the first criterion.