



Confidence and central tendency in perceptual judgment

Yang Xiang¹ · Thomas Graeber² · Benjamin Enke³ · Samuel J. Gershman^{4,5}

Accepted: 11 March 2021
© The Psychonomic Society, Inc. 2021

Abstract

This paper theoretically and empirically investigates the role of noisy cognition in perceptual judgment, focusing on the *central tendency effect*: the well-known empirical regularity that perceptual judgments are biased towards the center of the stimulus distribution. Based on a formal Bayesian framework, we generate predictions about the relationships between subjective confidence, central tendency, and response variability. Specifically, our model clarifies that lower subjective confidence as a measure of posterior uncertainty about a judgment should predict (i) a lower sensitivity of magnitude estimates to objective stimuli; (ii) a higher sensitivity to the mean of the stimulus distribution; (iii) a stronger central tendency effect at higher stimulus magnitudes; and (iv) higher response variability. To test these predictions, we collect a large-scale experimental data set and additionally re-analyze perceptual judgment data from several previous experiments. Across data sets, subjective confidence is strongly predictive of the central tendency effect and response variability, both correlationally and when we exogenously manipulate the magnitude of sensory noise. Our results are consistent with (but not necessarily uniquely explained by) Bayesian models of confidence and the central tendency.

Keywords Visual perception · Bayesian modeling

Introduction

One of the most robust empirical regularities in studies of human perception is the *central tendency* (or regression) effect: across various perceptual domains, estimates of stimulus magnitude are consistently biased towards the center of the magnitude distribution (Hollingworth, 1910; Stevens & Greenbaum, 1966). One prominent explanation for the central tendency effect is that sensory signals are “regularized” to reduce the influence of noise. Intuitively, when the signal-to-noise ratio is very low, then the best guess is the center of the magnitude distribution. When

the signal-to-noise ratio is very high, then the best guess will ignore the magnitude distribution and only use the signal. At intermediate levels, the best guess will be somewhere in between the signal and the center of the magnitude distribution. These intuitions can be formalized in a Bayesian framework (Petzschner, Glasauer, & Stephan, 2015), where the prior expresses the effect of the magnitude distribution, and the likelihood expresses the effect of sensory noise. Bayes’ rule prescribes how these two sources of information should be optimally combined.

Central to the Bayesian framework is the role of noise in predicting the strength of the central tendency effect. However, typically we cannot directly measure the noise level. This empirical gap is frequently filled by *ad hoc* assumptions or by fitting free parameters to the data, rendering the Bayesian framework possibly unfalsifiable (Marcus & Davis, 2013; Jones & Love, 2011). The same issue vexes non-Bayesian models (e.g., Ratcliff & McKoon 2018, 2020) that implicitly or explicitly make assumptions about sensory or cognitive noise.

One way to get around this issue is to collect other measures that are hypothesized to capture the magnitude of noise. Even if we cannot measure the noise level directly, we can make predictions about the relationships between indirect measures and magnitude estimates as a test of

✉ Yang Xiang
yyx@mit.edu

- ¹ Sloan School of Management, MIT, Cambridge, MA, USA
- ² Harvard Business School, Boston, MA, USA
- ³ Department of Economics, Harvard University, Cambridge, MA, USA
- ⁴ Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA, USA
- ⁵ Center for Brains, Minds and Machines, MIT, Cambridge, MA, USA

model predictions. We pursue this strategy here, using subjective confidence and response variability as auxiliary measures to triangulate the effects of noise on perception. To make our predictions precise, we develop a simple Bayesian model of magnitude estimation and derive a number of generic predictions from it (i.e., predictions that don't depend strongly on the parameter values). However, our goal is not to advocate for the Bayesian model versus alternative models, but rather to formalize some hypothetical regularities which (if true) would need to be satisfied by any model of magnitude estimation.

Our model makes several key predictions, elaborated in the next section: (i) confidence should decrease with sensory noise; (ii) the sensitivity of magnitude estimates to actual magnitudes should increase with subjective confidence and decrease with sensory noise; (iii) the central tendency effect should decrease with confidence and increase with sensory noise; (iv) both confidence and sensitivity should decrease with stimulus magnitude, assuming sensory noise that grows with stimulus magnitude; and (v) response variability should increase with sensory noise and decrease with confidence when prior uncertainty is high relative to sensory noise.

We test these model predictions in two ways. First, we implement a new large-scale magnitude estimation experiment. We elicit both magnitude estimates and subjective confidence. A key feature of our experiment is that we exogenously vary the objective stimulus, the mean of the stimulus distribution, and the magnitude of sensory noise. We collect these data because, as explained below, existing data sets on magnitude estimation and subjective confidence lack sufficient variation in some of the key elements of Bayesian models. Second, moving beyond our new experiment, we re-analyze data from several earlier studies that measured both continuous reports of stimulus magnitude and subjective confidence. Although these earlier studies are limited in several ways (which motivated our new experiment), the results from our re-analysis provide converging evidence for our model's main predictions.¹ Whether or not one accepts the Bayesian framework, these findings provide constraints on any model of confidence and central tendency in perceptual judgment.

Theoretical framework

To motivate our empirical predictions, we will first lay out a theoretical framework based on a simple Bayesian estimation problem, which mirrors the experimental tasks given to subjects.

¹The data and code are available at <https://github.com/yyxiang/confidence.central.tendency>.

Bayesian estimation

We model a task in which subjects are asked to estimate the magnitude of a stimulus x from a noise-corrupted signal $s = x + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is Gaussian-distributed sensory noise. For concreteness, consider the task of estimating the number of objects on a screen. In this case x is the true number of objects and s is the observer's sensory representation of the number.

If the subject has a Gaussian prior over the magnitude, $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$, then the posterior is also Gaussian:

$$P(x|s) = \mathcal{N}(x; \lambda s + (1 - \lambda)\mu_x, \sigma_x^2), \tag{1}$$

where the posterior variance is

$$\sigma_x^2 = (1 - \lambda)\sigma_x^2, \tag{2}$$

and the *sensitivity* is

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2}, \tag{3}$$

which takes values between 0 and 1. Intuitively, the magnitude estimate will be more sensitive (higher λ) to the signal when the sensory noise variance (σ_ϵ^2) is small relative to the prior uncertainty (σ_x^2).

We follow most of the literature in assuming that the prior approximates either the experienced magnitude distribution or the instructed distribution, depending on the paradigm. For analytical tractability, we assume a Gaussian prior even when the experimentally controlled stimulus distribution is non-Gaussian (e.g., uniform). In these cases, the mean is assumed to be equal to the center of the uniform distribution (e.g., Roach, McGraw, Whitaker, & Heron, 2017; Acerbi, Wolpert, & Vijayakumar, 2012). Our experimental predictions do not depend strongly on the choice of prior variance, provided that it is a fixed function of the magnitude distribution range.

We posit that the subjective estimate \hat{x} is chosen to minimize expected loss $\mathcal{L}(\hat{x}, x)$ conditional on the signal:

$$\hat{x} = \arg \min_{\hat{x}} \mathbb{E}[\mathcal{L}(\hat{x}, x)|s]. \tag{4}$$

The optimal subjective estimate under the quadratic loss $\mathcal{L}(\hat{x}, x) = (x - \hat{x})^2$ equals the posterior mean (see, e.g., Berger 1985):

$$\hat{x} = \lambda s + (1 - \lambda)\mu_x. \tag{5}$$

For the Gaussian estimation problem described here, this prediction holds true for several other loss functions. For example, the Bayes-optimal estimator for the absolute loss $\mathcal{L}(\hat{x}, x) = |x - \hat{x}|$ is the posterior median, and the Bayes-optimal estimator for the relaxed 0-1 loss ($\mathcal{L}_\kappa(\hat{x}, x) = 1$ if $|x - \hat{x}| \geq \kappa$, 0 otherwise) is the posterior mode in the

limit $\kappa \rightarrow 0$. Since the mean, median, and mode for a Gaussian distribution have the same value, our predictions are invariant across these choices of loss function.

Subjective confidence

Magnitude estimation tasks require subjects to report a single point estimate of the magnitude, but Bayesian models hypothesize that subjects are representing an entire distribution over magnitudes. Subjective confidence judgments can potentially provide a probe of this distributional representation. According to the Bayesian confidence hypothesis (Meyniel, Sigman, & Mainen, 2015; Pouget, Drugowitsch, & Kepecs, 2016; Sanders, Hangya, & Kepecs, 2016; Aitchison, Bang, Bahrami, & Latham, 2015; Fleming & Daw, 2017; Rahnev, Koizumi, McCurdy, D’Esposito, & Lau, 2015), subjective confidence corresponds to the posterior probability that an action is optimal (in this context, the probability that the subjective estimate equals the objective stimulus magnitude). Note that for continuous magnitudes, the probability that the point estimate \hat{x} equals the objective magnitude x is 0. However, we can evaluate the posterior probability that the true magnitude falls within an infinitesimally small region around the posterior mean estimate. In the limit, this probability becomes the density of the normally distributed estimate, evaluated at its mean: $\frac{1}{\sigma_{\hat{x}}\sqrt{2/\pi}} \equiv c$. Thus, we see that Bayesian confidence (c) for the Gaussian estimation problem is inversely proportional to the posterior standard deviation. This result motivates the use of the standard deviation as a measure of “cognitive uncertainty” (Enke & Graeber, 2020).² Most importantly, note that under a constant prior variance σ_x^2 , Bayesian confidence c decreases in the variance of sensory noise, σ_ϵ^2 . Because our experimental design controls the prior variance as argued above, we will in the following interpret Bayesian confidence as an approximate measure of sensory noise.

Predictions

Using this framework, we can now lay out a set of theoretical predictions, which we will empirically test in subsequent sections.

Prediction 1 *Sensitivity (λ) monotonically increases with confidence (c).*

²Note that the measure of Bayesian confidence described here does not directly map onto our experimental instructions, which simply asked subjects to report their subjective confidence using a Likert scale. However, this formulation has the advantage of being directly related to earlier formulations for discrete decisions.

This prediction follows from the expression relating sensitivity to confidence:

$$\lambda = 1 - \frac{2}{\pi c^2 \sigma_x^2}. \tag{6}$$

Although the model quantity λ is not directly observable, we can approximate it empirically as described in the experimental analysis below. The next prediction concerns the relationship between confidence and the central tendency effect, a straightforward corollary of Prediction 1 because $\frac{\partial \hat{x}}{\partial \mu_x} = 1 - \lambda$. As discussed above, we follow previous work in assuming that the prior is determined by the experienced or instructed magnitude distribution, and hence μ_x is the mean of the magnitude distribution.

Prediction 2 *The central tendency effect ($\frac{\partial \hat{x}}{\partial \mu_x}$) monotonically decreases with confidence.*

Here we have defined the central tendency effect formally as the degree to which the perceptual estimate changes with the average stimulus magnitude.

We now state the causal analogs of these predictions based on exogenous changes in sensory noise.

Prediction 3 *An exogenous increase in sensory noise (σ_ϵ^2), which reduces subjective confidence, decreases sensitivity.*

Prediction 4 *An exogenous increase in sensory noise increases the central tendency effect.*

Next, we theoretically explore how the strength of the central tendency effect depends on the stimulus magnitude. A well-known phenomenon is that response variability increases with stimulus magnitude according to Weber’s law. A common interpretation is that the signal-to-noise ratio decreases with stimulus magnitude, due either to a non-linear transformation of magnitude (e.g., Petchschner & Glasauer, 2011; Fechner 1860; Nieder & Miller 2003; Stevens 1961; Roach et al. 2017) or to magnitude-dependent scaling of sensory noise (e.g., Treisman 1964; Gibbon 1977; Gibbon & Church 1981). For concreteness, we examine the implications of the latter assumption:

$$\frac{\partial \sigma_\epsilon^2}{\partial x} > 0 \Rightarrow \frac{\partial c}{\partial x} < 0 \quad \text{and} \quad \frac{\partial \lambda}{\partial x} < 0 \tag{7}$$

Prediction 5 *As the stimulus magnitude increases, confidence and sensitivity decrease. The latter effect implies a stronger central tendency effect for larger magnitudes.*

Finally, we turn to response variability. Theoretically, variability is affected both by the effect of sensory noise on estimates (which increases variability) and the counteracting central tendency effect (which decreases

variability; see Enke & Graeber 2020). The expression for response variability is given by:

$$\text{Var}(\hat{x}|x) = \lambda^2 \sigma_\epsilon^2 = \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2} \right)^2 \sigma_\epsilon^2. \quad (8)$$

The relationship between response variability and other quantities depends on the degree of prior uncertainty relative to sensory noise.

Prediction 6 *When sensory noise is small relative to prior uncertainty ($\sigma_\epsilon^2 < \sigma_x^2$), response variability increases in sensory noise variance and decreases in confidence. When sensory noise is large relative to prior uncertainty ($\sigma_\epsilon^2 > \sigma_x^2$), response variability decreases in sensory noise variance and increases in confidence.*

Intuitively, when sensory noise is zero, the response is exactly equal to the stimulus and there is no residual variability. In the limit of large sensory noise, the response equals the prior mean and again there is no residual variability. Thus there is response variability only for intermediate values of sensory noise.

Experiment

Although a considerable number of studies have implemented magnitude estimation tasks while also measuring subjective confidence, none of these studies explicitly vary all of the variables that are of interest in light of our theoretical predictions above. Specifically, we require a study setup that features (i) variation in stimuli; (ii) a meaningful degree of exogenous variation in the mean of the stimulus distribution; (iii) exogenous variation in sensory noise; and (iv) a large sample size to allow for sufficiently powered statistical analyses of the relationship between confidence and the precise mechanics of central tendency. Earlier studies typically had sample sizes of fewer than 50 participants (see Table 2).

Materials and methods

Participants

We recruited 300 participants from Amazon Mechanical Turk (MTurk). All participants gave informed consent prior to testing. To ensure that participants fully understood the experiment, they completed a comprehension check immediately after the instructions. Participants who failed the comprehension check were asked to leave the study and were compensated (\$0.5) for their time. Participants who passed the comprehension check proceeded with the experiment and were paid (\$4.5) for their participation. Participants also had the opportunity to earn a performance-

based bonus payment (\$1) if their estimate on a randomly chosen trial was within 2 of the stimulus magnitude. 79 out of 300 participants received the bonus payment. The hourly rate was between \$9/hour and \$11/hour depending on whether they received a bonus. The experiment was approved by the Harvard Institutional Review Board.

Stimuli

The stimuli were arrays of black dots on a white background. The number of dots ranged between 15 and 65. Note that we did not control for density, so we cannot make a strong claim about numerosity *per se* using these stimuli.

Each participant completed six blocks with a total of 240 arrays. For each participant, the average stimulus magnitude within each block was randomly drawn from a discrete uniform distribution on the integers from 30 to 50 (sampled without replacement). For each block, the within-block distribution was a uniform distribution centered at the average stimulus magnitude, ranging between ± 15 (sampled with replacement). This procedure ensured that our trials feature substantial variation both in actual stimuli within-block and in average stimuli across blocks. We conceptualize the latter as the perceptual prior and use the variation to manipulate the central tendency. To make the prior salient to subjects, they were informed about the average stimulus value within a block at the beginning of each block, before making any estimates.

The stimulus duration on each trial was either 100 ms or 2000 ms, intermixed randomly within each block; each of the two duration conditions appeared on half of the trials in each block. This feature of our experiment allows us to leverage exogenous variation in sensory noise and confidence to test Prediction 3. This is motivated by the common assumption that sensory evidence is accumulated across time, such that the signal-to-noise ratio is higher for longer durations (Inglis & Gilmore, 2013; Cheyette & Piantadosi, 2019; 2020).

Procedure

As illustrated in Fig. 1, each trial began with the presentation of a fixation cross at the center of the screen for 500 ms. Then a random dot array was presented for either 100 ms or 2000 ms. Next, participants had 10 s to type in their numerosity estimate using the number pad on the keyboard. Participants were then prompted to provide a confidence rating by clicking on a discrete slider from 0-10.

Data analysis

We excluded trials on which participants did not respond. We analyzed the data using linear mixed-effects models.

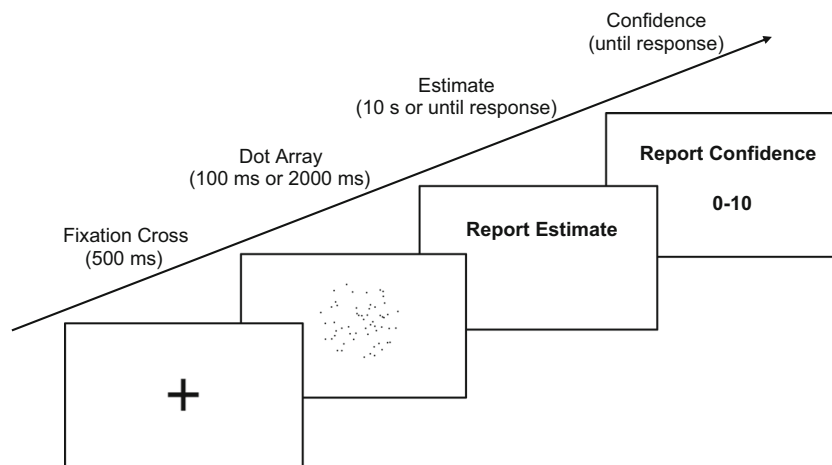


Fig. 1 Illustration of the experiment. Each trial begins with a fixation cross, followed by a dot array. Participants then report their numerosity estimate and confidence

Model 1 regresses subjective magnitude estimates on the true stimulus magnitude value (Stimulus) and the average stimulus magnitude in a block (AveStim), with random effects for the intercept, Stimulus, and AveStim grouped by participants. In the absence of a central tendency effect, the regression coefficient of the true stimulus should be one, while the coefficient of average stimulus should be zero. The central tendency effect is indicated by a stimulus coefficient of less than one and an average stimulus coefficient of greater than zero.

To explore the role of sensory noise and confidence for the central tendency effect, we ran Model 2. Here, we added as additional regressors subjective confidence, the interaction between confidence and the stimulus, and the interaction between confidence and average stimulus, and added random effects for subjective confidence, the interaction between confidence and the stimulus, and the interaction between confidence and average stimulus grouped by participants. As derived above, our hypothesis is that higher confidence (as a consequence of lower sensory noise) is associated with a higher responsiveness to the stimulus and a lower responsiveness to the average stimulus, meaning that the first interaction effect should be positive and the second one negative.

In addition, Model 2 also accounts for our exogenous variation in stimulus duration, which we conceptualize as exogenous variation in sensory noise that translates into confidence. We include as regressors a binary Condition indicator (0 for the 100 ms condition, 1 for the 2000 ms condition), the interaction between the stimulus and Condition, and the interaction between average stimulus magnitude and Condition, and random effects for Condition, the interaction between the stimulus and

Condition, and the interaction between average stimulus magnitude and Condition grouped by participants.

To see how the proportional regression to the prior changes in stimulus magnitude, we specified Model 3. Here, we regress an empirical estimate of λ on stimulus magnitude with random effects for the intercept and stimulus magnitude grouped by participants, and hypothesize a negative coefficient.³

Turning to the analysis of response variability, in Model 4, we regressed Variability (response standard deviation⁴ across multiple repetitions of the same magnitude) on Confidence (averaged across stimulus repetitions), with random effects for the intercept and Confidence grouped by participants. In Model 5, we regressed Variability on Condition, with random effects for the intercept and Condition grouped by participants. For better interpretability, we standardized the Confidence coefficients.

³Note that since we assume that the noise-corrupted signal to be unbiased, $\mathbb{E}[s|x] = x$, the average estimate of a given objective stimulus magnitude x converges to $\mathbb{E}[\hat{x}|x] = \lambda x + (1-\lambda)\mu_x$ (following Eq. 5). Under the simplifying assumption of $s = x$ for a given signal realization (which holds in expectation), we can empirically approximate an analogue of λ for each individual estimate \hat{x} :

$$\hat{\lambda} := \frac{\hat{x} - \mu_x}{x - \mu_x} \tag{9}$$

In regression analyses, we excluded $\hat{\lambda}$ values of $+/-$ infinity and those where $\hat{\lambda}$ was undefined due to a denominator of zero. We excluded 3% of the sample this way.

⁴We used standard deviation instead of variance because we found it to be slightly better behaved. The predictions are qualitatively the same for both standard deviation and variance.

Results

Preliminaries

Our study rests on two prerequisites: (i) The existence of a central tendency effect; and (ii) variation in subjective confidence as a function of stimulus duration.

The results of Model 1 (column 1 of Table 1) confirm the existence of a central tendency effect in our data: the coefficient of the true stimulus is substantially smaller than one [$F(1, 71405) = 1372.2, p < 0.0001$], and the coefficient of the average stimulus is considerably larger than zero [$F(1, 71405) = 71.56, p < 0.0001$]. To visualize the central tendency effect, we plotted subjective estimates as a function of stimulus magnitude. As shown in Fig. 2a, the slopes of the estimation functions are considerably smaller than 1, and the level of bias in the estimation functions increases in the mean of the stimulus distribution.

Our regression analysis revealed that confidence was higher for longer durations [$t(71405) = 16.20, p < 0.0001$]⁵, consistent with the hypothesis that lower sensory noise registers as higher confidence (Fig. 2b).

Confidence and the central tendency effect

We begin our main analysis by testing Predictions 1 and 2. The results of Model 2 (column 2 of Table 1) show that the central tendency effect is strongly moderated by subjective confidence. The positive interaction effect of confidence and stimulus implies that, for every standard deviation of confidence, the responsiveness of subjective estimates to the stimulus value increases by 5 percentage points [$t(71399) = 6.71, p < 0.0001$]. This result confirms Prediction 1. Similarly, the negative interaction effect of confidence and average stimulus shows that confident subjects place substantially lower weight on the mean of the stimulus distribution [$t(71399) = -3.18, p < 0.01$], supporting Prediction 2 that subjective estimates are pulled towards the prior mean to a greater extent when confidence is low. These two interaction effects correspond to the central predictions of the model about the role of sensory noise for the central tendency effect. Figure 2c visualizes the results by revealing stronger central tendency effects when confidence is low, i.e., the implied response curve is substantially flatter.

Sensory noise and the central tendency effect

To confirm our correlational findings, we leverage the exogenous variation in confidence that is induced by stimulus

Table 1 Regression coefficients and standard errors for stimulus estimates in the new data set (Models 1 and 2)

	Model 1	Model 2
(Intercept)	7.435**** (0.464)	6.666**** (0.560)
Stimulus	0.637**** (0.010)	0.537**** (0.011)
AveStim	0.110**** (0.013)	0.210**** (0.017)
Confidence		-1.483**** (0.370)
Stimulus:Confidence		0.049**** (0.007)
AveStim:Confidence		-0.036** (0.011)
Condition		2.481**** (0.496)
Stimulus:Condition		0.178**** (0.009)
AveStim:Condition		-0.176**** (0.014)
Observations	71408	71408

Condition refers to the 2000 ms stimulus duration condition

The Confidence coefficients are standardized

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

duration. As shown in column 2 of Table 1, longer stimulus durations increased the effect of objective stimulus magnitude [$t(71399) = 20.91, p < 0.0001$] and decreased the effect of average stimulus magnitude [$t(71399) = -12.24, p < 0.0001$], consistent with a weaker central tendency effect when sensory noise is lower. In combination with the strong effect of duration on confidence, these results suggest that sensory noise causally influences both confidence and central tendency effects, confirming Prediction 3 and Prediction 4. Figure 2d visualizes these patterns by showing that the central tendency effect is considerably more pronounced under shorter stimulus duration, i.e., the implied response curve is flatter.

Sensory noise and the magnitude-dependent central tendency effect

We find strong support for Prediction 5. First recall from Fig. 2b that confidence decreases in the stimulus magnitude, indicating larger sensory noise. Our model therefore predicts proportionally larger compression (lower λ) at higher magnitudes. The results of Model 3 show that the empirical analog of λ indeed decreases in stimulus magnitude [$t(69087) = -17.17, p < 0.0001$], meaning that participants display proportionally stronger central tendency effects at higher stimulus magnitudes. We can also intuitively gauge this effect in Fig. 2a, where estimates are further away from the 45-degree line at higher magnitudes.

Sensory noise and response variability

Figures 2e and f visualize the results on the relationship between sensory noise and response variability. Note that

⁵This result came from an additional regression, where we regressed Confidence on Stimulus and Condition, with random effects for the intercept, Stimulus, and Condition grouped by participants.

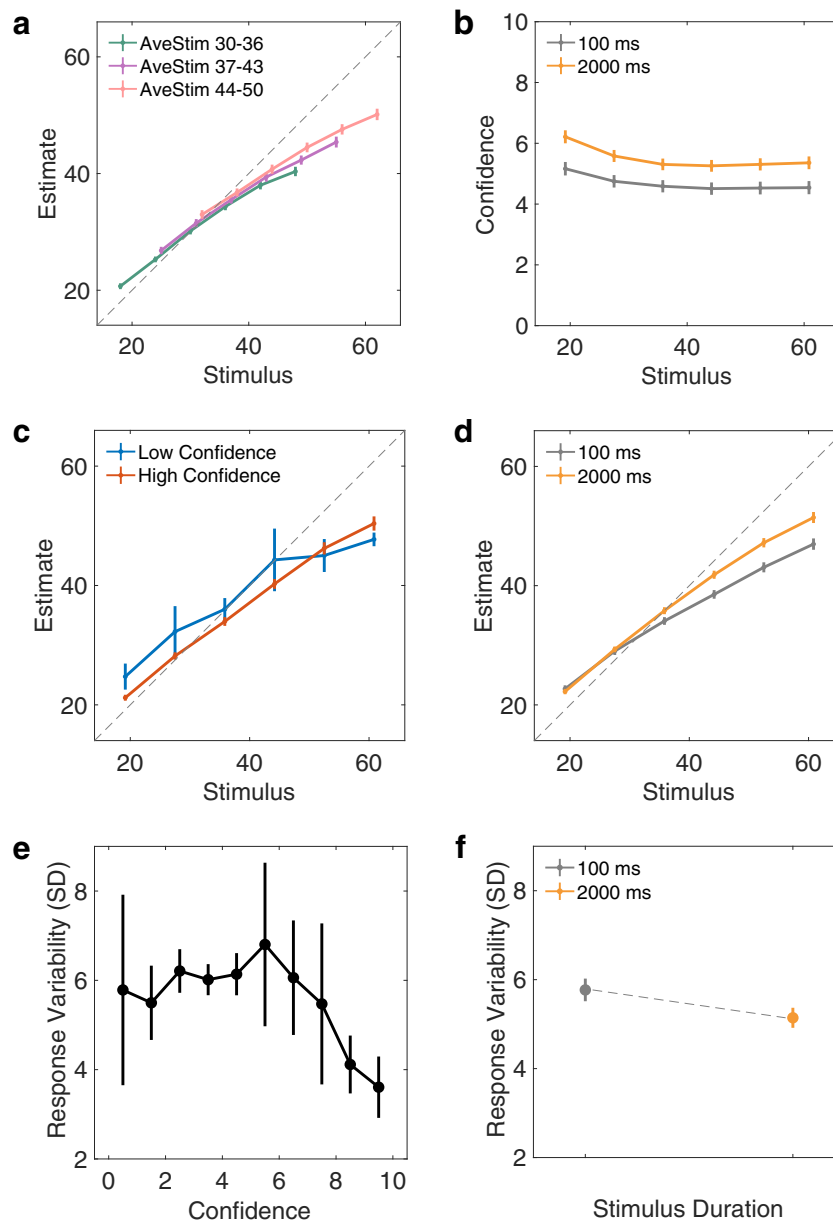


Fig. 2 **a** Subjective estimates as a function of objective stimulus magnitude, shown separately for three within-block average stimulus ranges. **b** Confidence as a function of objective stimulus magnitude, shown separately for different stimulus duration conditions. **c** Subjective estimates as a function of objective stimulus magnitude, shown separately for low confidence (confidence levels 0-3) and high confidence (confidence levels 7-10). **d** Subjective estimates as a function of objective stimulus magnitude, shown separately for different stimulus durations. **e** Response standard deviation as a function of confidence. **f** Response standard deviation under short and long stimulus duration. Error bars indicate 95% confidence intervals

Prediction 6 implies a hump-shaped relationship between response variability and confidence: response variability should increase at low confidence levels (sensory noise variance high relative to prior uncertainty) but decrease at high confidence levels (sensory noise variance relatively low). Figure 2e strongly supports this distinctive prediction of the Bayesian model. A plausible assumption is that prior uncertainty usually exceeds sensory noise variance.

Correspondingly, we find that response variability decreases across the upper half of the confidence range when sensory noise variance is low relative to prior uncertainty (Fig. 2e), and results from our regression Model 4 show a negative effect of confidence [$t(11979) = -6.93, p < 0.0001$]. Turning to the causal manipulation, we see from Model 5 results that variability is higher under shorter stimulus duration [$t(18181) = -5.36, p < 0.0001$; also see

Table 2 Descriptive information about the included studies

Data set	Reference	Stimuli	# of Participants
AB17	Akdoğan and Balcı (2017)	Duration of Time	76
DB18	Duyan and Balcı (2018)	Number of Beeps	44
DB19	Duyan and Balcı (2019)	Number of Dots	18
DB20	Duyan and Balcı (2020)	Length of Lines	41
RZ14	Rausch and Zehetleitner (2014)	Direction of Coherent Motion	20
SP17	Samaha and Postle (2017)	Orientation of Grating	40

Experiment 4 in data set AB17 was excluded because each participant was presented with a single stimulus magnitude. Experiment 3 in data set SP17 was excluded because it did not elicit continuous report of stimulus magnitude

Fig. 2f]. Thus the data clearly support the model predictions about the relationship between variability, sensory noise and subjective confidence.

Discussion

Taken together, the results establish support for all six predictions spelled out above. In particular, (i) lower sensory noise (high confidence) is associated with a *higher* sensitivity of estimates to the true stimulus; (ii) lower sensory noise (higher confidence) is associated with a *lower* sensitivity of estimates to the average stimulus; (iii) these results hold both in correlational analyses and when we exogenously manipulate confidence; (iv) central tendency increases with the stimulus magnitude; and (v) response variability decreases with confidence and increases under shorter stimulus duration.

Re-analysis of earlier studies

While our experiment has the advantage of being specifically tailored to investigate the predictions associated with the core elements of a Bayesian model, we sought to test the validity of our hypothesis more generally. Owing to the recent publication of the Confidence Database (Rahnev et al., 2020), we were able to additionally address the relationship between confidence and the central tendency effect by re-analyzing data from several earlier studies.

Materials and methods

Data sets

Out of a total of 145 studies contained in the Confidence Database (Rahnev et al., 2020), we identified six studies that elicited *continuous* reports of both stimulus magnitude and subjective confidence (Table 2). Two of these studies use circular stimuli (motion direction and grating orientation),

and hence are not relevant for analyses of magnitude-dependent noise. Nonetheless, these studies are still informative about the interplay between objective magnitude and confidence in determining subjective judgments.

Data analysis

To measure the central tendency effect, we specified the average stimulus magnitude as the cumulative mean of all across-block stimulus magnitudes that preceded the current trial. We excluded each participant's first trial in regressions involving average stimulus, because first trials do not have preceding trials. In our analysis, we excluded estimates that were clearly out of range and trials on which participants did not comply with the experiment rules.⁶

To assess the relationship between confidence and central tendency, we constructed linear mixed-effects models akin to the ones used to analyze our own experiment, except that here we do not have treatment conditions that exogenously vary sensory noise and confidence through stimulus duration.

Results and discussion

Subjective estimates correlate more strongly with objective stimulus magnitude under high confidence

Consistent with the model and our own experimental results, we document significant positive interactions between stimulus and confidence in all data sets (Table 3): AB17 [$t(15254) = 4.66, p < 0.0001$], DB18 [$t(14432) = 3.78, p < 0.001$], DB19 [$t(8563) = 3.58, p < 0.001$], DB20 [$t(9651) = 6.16, p < 0.0001$], RZ14 [$t(8974) = 22.84, p < 0.0001$], and SP17 [$t(14059) = 11.57, p < 0.0001$]. In other words, subjective estimates correlate

⁶These are trials with estimates 999 or NaN in DB18, trials with estimates 9999 or 999 in DB19, and trials on which participants did not respond.

Table 3 Regression coefficients and standard errors for the re-analysis of stimulus estimates (Model 2)

Dataset	AB17	DB18	DB19	DB20	RZ14	SP17
(Intercept)	0.512**** (0.116)	0.363 (0.929)	6.367**** (1.253)	93.025**** (14.287)	72.216**** (16.608)	123.56**** (8.076)
Stimulus	0.849**** (0.017)	0.900**** (0.042)	0.631**** (0.040)	0.671**** (0.021)	0.352**** (0.031)	0.537**** (0.019)
AveStim	0.027 (0.036)	0.153 (0.078)	-0.103 (0.082)	-0.023 (0.035)	0.244** (0.083)	0.072 (0.094)
Confidence	-0.030 (0.037)	-0.357 (0.243)	-3.527** (1.237)	-2.558 (8.303)	-49.055**** (14.176)	-26.668**** (7.255)
Stimulus:Confidence	0.035**** (0.007)	0.058**** (0.015)	0.076**** (0.021)	0.056**** (0.009)	0.244**** (0.011)	0.190**** (0.016)
AveStim:Confidence	-0.054**** (0.014)	-0.027 (0.022)	0.086 (0.080)	-0.035 (0.022)	0.036 (0.077)	0.108 (0.085)
Observations	15260	14438	8569	9657	8980	14065

The Confidence coefficients are standardized

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

more strongly with the objective stimulus magnitude when confidence is high, as predicted by Bayesian models of perception.

In contrast to the consistent interaction between objective stimulus magnitude and confidence across studies, the interaction between average stimulus magnitude and confidence is not consistently observed across studies (Table 3). We find a significant effect in only one of the studies [AB17, $t(15254) = -3.88$, $p < 0.001$], which has the predicted sign. We conjecture that this is due to insufficient variation in the average stimulus magnitude in prior experiments. In our new study reported above, we remedied this limitation by exploring a wider range of average stimulus magnitudes to be sufficiently statistically powered.

The low variation in stimulus magnitude in the earlier studies also makes our analysis of Prediction 5 (Model 3) difficult. We hypothesized that the magnitude of central tendency increases in stimulus magnitude, which should produce a negative coefficient for stimulus magnitude. The results of Model 3 are mixed, with a negative coefficient for only two studies, and neither of them is significant [AB17, $t(12454) = -0.77$, $p = 0.44$; DB19, $t(8564) = -0.31$, $p = 0.75$]. Again, this null result is to be expected given insufficient variation in (average) stimulus magnitude. Our new experiment deliberately remedies this shortcoming.

Response variability decreases with confidence

Finally, we again find strong evidence for confidence-dependent variability. Five studies show a significantly

negative coefficient: AB17 [$t(150) = -2.07$, $p < 0.05$], DB18 [$t(86) = -3.96$, $p < 0.001$], DB19 [$t(466) = -5.62$, $p < 0.0001$], RZ14 [$t(2538) = -10.58$, $p < 0.0001$], and SP17 [$t(4169) = -4.60$, $p < 0.0001$]. These results are very similar to the ones observed in our own experimental data.

General discussion

Using data from earlier studies and a new data set, we have established a relationship between confidence and central tendency that conforms with (but is not necessarily unique to) the generic predictions of Bayesian models. First, we showed that the central tendency effect is lower on high confidence trials. Second, we showed that when sensory noise was exogenously increased via a stimulus duration manipulation, the central tendency effect increased and confidence decreased, demonstrating the causal role of sensory noise. Third, we showed a stronger central tendency effect at higher magnitudes, which is in line with subjective confidence decreasing in stimulus magnitude. Fourth, we showed that across-trial variability in responses decreased in subjective confidence and increased in sensory noise whenever prior uncertainty is relatively large.

Our findings bridge several disparate lines of research on confidence and central tendency effects. Some theories assert that confidence judgments in perceptual decision making tasks reflect the posterior probability of being correct—the *Bayesian confidence hypothesis* (Meyniel et al., 2015; Pouget et al., 2016; Sanders et al., 2016;

Fleming & Daw, 2017; Rahnev et al., 2015). While past experimental work on the Bayesian confidence hypothesis has focused on discrete choice tasks (Aitchison et al., 2015), here we analyzed continuous report tasks, which allowed us to relate confidence judgments to the central tendency effect. Our finding that this relationship held across several different stimulus domains (time, visual numerosity, auditory numerosity, line length, motion direction, and grating orientation) lends support to the generality of our conclusions.

While our findings are specifically consistent with the Bayesian confidence hypothesis, they might also be compatible with alternative models. For example, Adler and Ma (2018) developed several models that map probability representations of uncertainty onto confidence in a non-Bayesian way, and presented experimental evidence that some of these models outperformed the Bayesian model in predicting confidence judgments. Li and Ma (2020) developed a different non-Bayesian model, which determined confidence based on the difference in probability between the top two hypotheses. For our purposes, all of these non-Bayesian models share the key property that confidence is lower when uncertainty (due to sensory noise) is greater.

Some authors have argued that previous data supporting Bayesian models can be explained by simpler heuristic models. For example, Huttenlocher, Hedges, and Vevea (2000) tested a Bayesian model of perceptual judgment similar to the one analyzed here, but their conclusions were questioned by later work showing that the same patterns of behavior could be fit by a model that simply reports an average of recent stimulus magnitudes (Duffy & Smith, 2020). This alternative model has limited explanatory scope for the data we discuss here, because it is silent about the role of confidence in generating judgment and the effects of stimulus duration. Similarly, their account does not explain the patterns of a stronger central tendency effect at higher stimulus magnitudes and of predictable heterogeneity in response variability.

A second line of research bridged by our theory is on the central tendency effect. Recent research has shown that cognitive load (which ostensibly increases sensory noise) strengthens the central tendency effect, broadly consistent with Bayesian models of perception. For example, Allred, Crawford, Duffy, and Smith (2016) found that asking participants to memorize six-digit numbers (high load condition) increased the central tendency effect in the estimation of line length, compared to a low load condition in which participants memorized two-digit numbers. Similarly, Olkkonen, McCarthy, and Allred (2014) found that increasing chromatic noise or the delay between stimulus presentation and estimation increased the central tendency effect in a color estimation task.

Relatedly, there is evidence that the central tendency effect is stronger when sensory information is less reliable or when the magnitude distribution is more concentrated around the center (Ashourian & Loewenstein, 2011; Allred, Crawford, Duffy, & Smith, 2016; Olkkonen, McCarthy, & Allred, 2014; Huttenlocher, Hedges, & Vevea, 2000), again consistent with Bayesian models. Our study took this line of research one step further, showing that confidence judgments respond to an exogenous manipulation of sensory noise, while explaining significant additional response variance not explained by noise alone.

Our study is limited in several ways that can be addressed by future research. One is that we did not provide a detailed information processing model of behavior. This was a deliberate choice: our aim was to test generic principles rather than the predictions of specific information processing models, which would inevitably require a number of *ad hoc* assumptions. A second limitation is that we did not explore the effects of magnitude distribution shape on the central tendency effect.

The elementary psychophysical regularities studied here may have broader implications. Economists have begun to consider the significance of noisy cognition for a range of phenomena, including sensitivity to risk and ambiguity, belief updating, and survey responses (Enke & Graeber, 2020; Payzan-LeNestour & Woodford, 2020; Frydman & Jin, 2019; Woodford, 2019; Gabaix, 2019). Models of noisy cognition may therefore hold promise in unifying empirical evidence from disparate fields, explaining not only lower-level perceptual processes but also higher-level cognition.

Acknowledgements We are grateful to Lucy Lai and Henrik Singmann for helpful guidance. This research was supported by the Office of Naval Research (N00014-17-1-2984), the Center for Brains, Minds and Machines (funded by NSF STC award CCF-1231216), and a research fellowship from the Alfred P. Sloan Foundation.

Open Practices Statement All the data and code are available at https://github.com/yyxiang/confidence_central_tendency. The experiments were not preregistered.

References

- Acerbi, L., Wolpert, D. M., & Vijayakumar, S. (2012). Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Computational Biology*, *8*, e1002771.
- Adler, W. T., & Ma, W. J. (2018). Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLoS Computational Biology*, *14*, e1006572.
- Aitchison, L., Bang, D., Bahrami, B., & Latham, P. E. (2015). Doubly Bayesian analysis of confidence in perceptual decision-making. *PLoS Computational Biology*, *11*, e1004519.
- Akdoğan, B., & Balci, F. (2017). Are you early or late?: Temporal error monitoring. *Journal of Experimental Psychology: General*, *146*, 347–361.

- Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review*, *23*, 1825–1831.
- Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One*, *6*, e19551.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Berlin: Springer Science & Business Media.
- Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences*, *116*, 17729–17734.
- Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour*, 1–8.
- Duffy, S., & Smith, J. (2020). Omitted-variable bias and other matters in the defense of the category adjustment model: A comment on Crawford (2019). *Journal of Behavioral and Experimental Economics*, *85*, 101501.
- Duyan, Y. A., & Balci, F. (2018). Numerical error monitoring. *Psychonomic Bulletin & Review*, *25*, 1549–1555.
- Duyan, Y. A., & Balci, F. (2019). Metric error monitoring in the numerical estimates. *Consciousness and Cognition*, *67*, 69–76.
- Duyan, Y. A., & Balci, F. (2020). Monitoring line length reproduction errors. *Consciousness and Cognition*, *77*, 102831.
- Enke, B., & Graeber, T. (2020). Cognitive uncertainty. Technical report, National Bureau of Economic Research.
- Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf u. Härtel.
- Fleming, S., & Daw, N. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*, 91–114.
- Frydman, C., & Jin, L. J. (2019). Efficient coding and risky choice. Available at SSRN 3270773.
- Gabaix, X. (2019). Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, (Vol. 2, pp. 261–343): Elsevier.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*, 279–325.
- Gibbon, J., & Church, R. (1981). Time left: Linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 87–108.
- Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*, 461–469.
- Huttenlocher, J., Hedges, L., & Vevea, J. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220–241.
- Inglis, M., & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition*, *129*, 63–69.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*, 169–231.
- Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications*, *11*, 1–11.
- Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*, 2351–2360.
- Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, *88*, 78–92.
- Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*, 149–157.
- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*, 5–5.
- Payzan-LeNestour, E., & Woodford, M. (2020). 'outlier blindness': Efficient coding generates an inability to represent extreme values. Available at SSRN 3152166.
- Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *Journal of Neuroscience*, *31*, 17220–17229.
- Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*, 285–293.
- Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*, 366.
- Rahnev, D., Desender, K., Lee, A. L., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., . . . , Bang, J. W. (2020). The confidence database. *Nature Human Behaviour*, *4*, 317–325.
- Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*, 1664–1680.
- Ratcliff, R., & McKoon, G. (2018). Modeling numerosity representation with an integrated diffusion model. *Psychological Review*, *125*, 183–217.
- Ratcliff, R., & McKoon, G. (2020). Decision making in numeracy tasks with spatially continuous scales. *Cognitive Psychology*, *116*, 101259.
- Rausch, M., & Zehetleitner, M. (2014). A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion. *Consciousness and Cognition*, *28*, 126–140.
- Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences*, *114*, 412–417.
- Samaha, J., & Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20172035.
- Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. *Neuron*, *90*, 499–506.
- Stevens, S. (1961). To honor Fechner and repeal his law. *Science*, *133*, 80–86.
- Stevens, S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, *1*, 439–446.
- Treisman, M. (1964). Noise and Weber's law: The discrimination of brightness and other dimensions. *Psychological Review*, *71*, 314–330.
- Woodford, M. (2019). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, *12*.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.