ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Otsuka T, Yotsumoto Y. 2023 Near-optimal integration of the magnitude information of time and numerosity. *R. Soc. Open Sci.* **10**: 230153.
https://doi.org/10.1098/rsos.230153

Received: 9 February 2023
Accepted: 20 July 2023

Subject Category:
Psychology and cognitive neuroscience

Subject Areas:
psychology/neuroscience

Keywords:
time perception, numerosity, magnitude interaction, multisensory integration, Bayesian, maximum-likelihood estimation

Author for correspondence:
Yuko Yotsumoto
e-mail: cyuko@mail.ecc.u-tokyo.ac.jp

Near-optimal integration of the magnitude information of time and numerosity

Taku Otsuka and Yuko Yotsumoto

Department of Life Sciences, University of Tokyo, Tokyo, Japan

TO, 0000-0002-5375-2033; YY, 0000-0001-7408-0339

Magnitude information is often correlated in the external world, providing complementary information about the environment. As if to reflect this relationship, the perceptions of different magnitudes (e.g. time and numerosity) are known to influence one another. Recent studies suggest that such magnitude interaction is similar to cue integration, such as multisensory integration. Here, we tested whether human observers could integrate the magnitudes of two quantities with distinct physical units (i.e. time and numerosity) as abstract magnitude information. The participants compared the magnitudes of two visual stimuli based on time, numerosity, or both. Consistent with the predictions of the maximum-likelihood estimation model, the participants integrated time and numerosity in a near-optimal manner; the weight of each dimension was proportional to their relative reliability, and the integrated estimate was more reliable than either the time or numerosity estimate. Furthermore, the integration approached a statistical optimum as the temporal discrepancy of the acquisition of each piece of information became smaller. These results suggest that magnitude interaction arises through a similar computational mechanism to cue integration. They are also consistent with the idea that different magnitudes are processed by a generalized magnitude system.

1. Introduction

Magnitude information, such as time, space and numbers, is essential for interacting with the environment. We routinely make conscious and unconscious judgements about how long an event lasted, how large an object is, and how many objects there are. Moreover, different magnitude dimensions are often correlated in the external world and thus provide complementary information about the environment. As if to reflect this relationship in the external world, ample evidence shows that the perception of magnitude in one dimension is influenced by magnitudes in

THE ROYAL SOCIETY PUBLISHING

other dimensions. For example, more numerous stimuli are perceived to last longer than less numerous ones, while longer-lasting stimuli are perceived to be more numerous [1–4]. In addition to these behavioural interactions between magnitudes, neuroimaging studies have shown that the processing of different magnitudes activates partially overlapping brain regions [5–9]. These findings converge into a prevailing theory: 'A theory of Magnitude' (ATOM; [10]). ATOM proposes the presence of a common magnitude-processing system that is shared across different dimensions located mainly in the parietal cortex [10,11].

Although ATOM conceptually accounts for the behavioural interactions between magnitudes (often referred to as 'magnitude interaction'), its predictions remain ambiguous and therefore cannot explain the systematic differences under different experimental conditions. For example, previous studies suggest that the size and directionality of the magnitude interaction depend on the range of the magnitude [2], the time course of the stimulus [3,12,13], and the reliability of each dimension [14–16]. Specifically, the perception of space/numerosity is affected by time when space/numerosity information dynamically accumulates over time [3,12,13,15]; the perception of time is influenced by space/numerosity when the reliability (i.e. precision) of space/numerosity is higher than that of time, but it is less affected when the reliability of the two pieces of information is comparable [14–16]. Therefore, to explain these observations and further elucidate the mechanism of magnitude interaction, a different framework that enables detailed quantitative predictions is required.

Interestingly, a recent study suggests that magnitude interaction reflects an active 'binding' mechanism that integrates different magnitudes of the same stimulus to achieve a unified magnitude representation [17], rather than a contextual interference mechanism as was conventionally assumed [5,10,18,19]. This raises the possibility that magnitude interaction (also referred to as 'magnitude integration' [3,17]) is considered one form of 'cue integration' [20–22]. Cue integration typically occurs when a single physical quantity is specified using multiple features that provide redundant information. For example, in a classical 'ventriloquist effect', the perception of the auditory spatial location is dragged by the nearby visual information [23]. This effect is thought to reflect the brain's function of integrating redundant audiovisual information to obtain a single estimate of a spatial location [24,25].

It has been shown that cue integration, including the ventriloquist effect, can be successfully explained by computational models based on a Bayesian framework [24,26–28]. A well-known model often tested as a benchmark for cue integration is the maximum-likelihood estimation (MLE) model [24,26]. Assuming that the estimates from each cue are unbiased and their noises (i.e. variability) are normally distributed and independent of one another, the MLE model makes two important quantitative predictions: first, the integrated estimate ($\hat{S}_{12}$) is a weighted average of two single estimates ($\hat{S}_1$ and $\hat{S}_2$),

$$\hat{S}_{12} = w_1\hat{S}_1 + w_2\hat{S}_2,$$

with the weights for cues 1 and 2 (w_1 and w_2) given in proportion to their relative reliabilities:

$$w_1 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

and

$$w_2 = 1 - w_1,$$

where σ_1^2 and σ_2^2 indicate the variances (i.e. the inverse of the reliability) associated with cues 1 and 2, respectively. Second, the variance of the integrated estimate (σ_{12}^2) is lower than that of either of the single estimates:

$$\sigma_{12}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \leq \min(\sigma_1^2, \sigma_2^2).$$

Therefore, the integrated estimate has the lowest possible variance. In this sense, the integration following the MLE model is statistically optimal.

As discussed above, magnitude interaction is qualitatively consistent with the MLE prediction that each piece of information is integrated based on its reliability [14–18]. However, no study has directly tested whether the quantitative predictions of the MLE model apply to magnitude integration, where two quantities that have distinct physical units (e.g. time and numerosity) are integrated as abstract magnitude information.

In the present study, we test the hypothesis that the brain integrates the magnitude information of different dimensions according to the MLE model to achieve a unified magnitude representation of the stimulus. We develop a task in which participants are asked to integrate the magnitude information of time and numerosity. In the task, participants compare the magnitude of two visual stimuli based on only time, only numerosity or both time and numerosity. We manipulate the reliability of the numerosity information by changing the contrast of a set of dots. Moreover, based on a previous report that asymmetries in the time course of a stimulus affect the interaction of time and numerosity [3], we examine the effect of the time course of a stimulus in separate experiments. In Experiment 1, we present the dots (i.e. numerosity) at the onset of the stimulus; in Experiment 2, we present the dots near the offset of the stimulus.

2. Methods

2.1. Participants

A total of 42 volunteers (13 females; mean age 20.4, range 18–27 years) participated in the study (21 participants for each of Experiments 1 and 2). All the participants reported normal or corrected-to-normal vision and no history of neurologic or psychiatric conditions. They provided written informed consent before the experiment and received monetary compensation for their participation. The experimental protocol was approved by the Institutional Review Board of the University of Tokyo.

According to a tutorial on cue integration research, 10 participants should be sufficient to test the MLE model [22]. However, because previous studies on magnitude integration have tested approximately 15–20 participants [2,3,17], we recruited 21 participants for each of Experiments 1 and 2.

2.2. Apparatus and stimuli

The experiments were conducted in a dark room. The stimuli were generated using Psychophysics Toolbox-3 [29] and MATLAB (R2021a; The MathWorks, Inc.) and were presented on a 23.6-inch LCD monitor with 1920 × 1080 resolution at a refresh rate of 120 Hz (VIEWPixx 3D; VPixx Technologies, Inc.). Participants were seated in a chair and their heads fixed on a chin rest 75 cm away from the monitor.

All stimuli were presented on a grey background. Stimuli comprised a set of dots presented on a dynamic random-noise patch. A black fixation cross (0.144° radius) was always presented at the centre of the screen except during the response period. The dynamic random-noise patch comprised a circle (8.450° diameter) and was updated every frame by generating random numbers from a normal distribution. The mean luminance was equal to that of the background, and the contrast was constant under all conditions. Each set of dots contained equal proportions of black and white dots randomly placed in a virtual circle. Each dot had a 0.051° radius. The minimum inter-dot and dot-fixation distance was 1.5 times the radius of a single dot, such that each dot did not overlap with any other dot or a fixation cross. The Michelson contrast of the dots was either 100% (high) or 48% (low) in Experiment 1 and either 100% (high) or 55% (low) in Experiment 2 (see 'Conditions'). The mean luminance of the dots was equal to that of the background. The radius of the virtual circle on which the dots were presented was varied in each trial to attenuate the correlation between the numerosity and non-numerical properties, such as the convex hull and density; it was either large (4.200° radius) or small (3.437° radius) for the comparison stimulus, while it was fixed at the average value of large and small for the standard stimulus. An example of stimuli is shown in figure 1.

2.3. Conditions

Experiments were conducted in a two-interval forced-choice (2IFC) paradigm. In each trial, participants were sequentially presented with two visual stimuli (i.e. standard and comparison) each of which comprised a set of dots placed on a dynamic random-noise patch. Under all conditions, the presentation duration for the dynamic random-noise patch defined the *time* dimension of the stimulus, while the number of dots defined the *numerosity* dimension. Participants were asked to compare the magnitude of the first and second stimuli based on time alone (T), numerosity alone (N), or both time and numerosity (TN). Specifically, in the time conditions (T), participants indicated whether the noise patch of the first or second stimulus lasted longer while ignoring the number of dots. In the numerosity conditions (N), they indicated whether the first or second stimulus comprised more dots

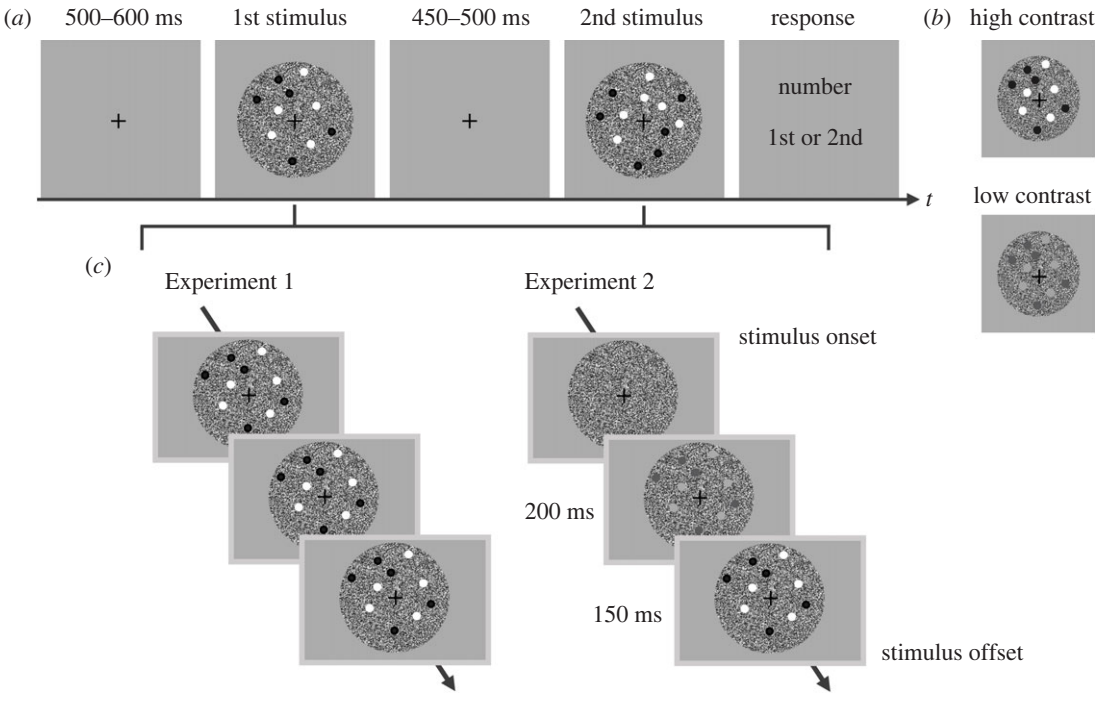

Figure 1. Schematics of the experiments. (*a*) Trial structure of the task. Participants were sequentially presented with two visual stimuli each comprising a set of dots placed on a dynamic random-noise patch, and were asked to compare the magnitudes of the first and second stimuli based on time (i.e. the presentation duration of the dynamic random-noise patch), numerosity (i.e. the number of dots), or both time and numerosity. These three types of trials were presented in separate blocks and participants were informed about what they had to judge before starting each block. The figure depicts an example of a numerosity trial. The order of standard and comparison stimuli was counterbalanced. (*b*) An example of contrast manipulation for numerosity information. The reliability of the numerosity information was manipulated by varying the contrast of the dots in the numerosity and time-numerosity conditions. The contrast was always high in the time conditions. (*c*) The time course of a stimulus in Experiments 1 and 2. In Experiment 1, a set of dots was presented at the stimulus onset. In Experiment 2, a set of dots was presented 350 ms before the stimulus offset, and its contrast was gradually increased.

while ignoring the duration of the noise patch. In the bidimensional conditions (TN), they attended to both the duration of the noise patch and the number of dots and indicated whether the first or second stimulus lasted longer and comprised more dots. Note that in the TN conditions, the *time* and *numerosity* of the comparison stimulus were perfectly correlated ($r = 1$), such that participants could integrate the two dimensions without confusion (see below). Figure 1 shows a visual depiction of a trial.

The standard stimulus lasted 600 ms (*time* dimension) and comprised 24 dots (*numerosity* dimension) in the unidimensional conditions (T, N). By contrast, there were three types of standard stimuli depending on the magnitudes of the *time* and *numerosity* in the TN conditions: $\Delta = -1$, $\Delta = 0$, and $\Delta = 1$, where Δ indicates that *time* was 600 ms multiplied by 1.1^{Δ} while *numerosity* was 24 dots multiplied by $1.1^{-\Delta}$. Therefore, $\Delta = -1$ means that *time* was 600×1.1^{-1} ms while *numerosity* was 24×1.1^{1} dots; $\Delta = 1$ means that *time* was 600×1.1^{1} ms while *numerosity* was 24×1.1^{-1} dots; and $\Delta = 0$ means that *time* was 600 ms while *numerosity* was 24 dots. The reason for setting the standard stimuli with different Δ in the TN conditions was to estimate the empirical weights. Here, Δ can be considered as a 'shift' of *time* and *numerosity* from the standard magnitude of 600 ms and 24 dots, respectively. For example, when $\Delta = -1$, the standard stimulus has *time* of 542 ms and *numerosity* of 26 dots, which can be thought that *time* and *numerosity* are 'shifted' 1.1^{-1} and 1.1^{1} times from the standard magnitude, respectively. These 'shifts' allow the estimation of empirical numerosity weights from the observed PSEs when the standard stimulus is $\Delta = -1$, $\Delta = 0$ and $\Delta = 1$ (see 'Data analysis').

The comparison stimulus took nine possible magnitude values, defined as the 600 ms and 24 dots multiplied by 1.1^{-5}, 1.1^{-3}, 1.1^{-2}, 1.1^{-1}, 1.1^{0}, 1.1^{1}, 1.1^{2}, 1.1^{3} and 1.1^{5}, resulting in durations of 375, 450, 500, 542, 600, 658, 725, 800 and 966 ms (*time* dimension) and numerosities of 15, 18, 20, 22, 24, 26, 29, 32 and 39 dots (*numerosity* dimension). This manipulation was necessary to make the magnitudes of *time* and *numerosity* comparable on a logarithmic scale (these values were rounded off to the nearest integer, and the durations corresponded to the presented durations calculated from the monitor's

refresh rate). In the T conditions, only the *time* dimension of the comparison stimulus could vary while *numerosity* was fixed at the standard magnitude (24 dots); in the N conditions, only *numerosity* could vary while *time* was fixed at the standard magnitude (600 ms). In the TN conditions, both *time* and *numerosity* could vary, but they were perfectly correlated ($r = 1$). That is, *time* and *numerosity* were equivalent in magnitude relative to the standard magnitude of 600 ms and 24 dots, respectively. This manipulation was a prerequisite for the MLE modelling and was based on the assumption that an ideal observer would weight *time* and *numerosity* equally, depending only on their relative reliability, when two magnitudes are perfectly correlated [21,28,30].

To manipulate the reliability of the numerosity information, we used two levels of dot contrast, high and low, in the N and TN conditions (see 'Apparatus and stimuli' and figure 1*b*). The dot contrast was fixed at the high value in the T conditions. Within trials, standard and comparison stimuli had the same contrast.

To examine the effect of the time course of a stimulus on magnitude integration, we conducted two experiments. In Experiment 1, a set of dots was always presented from the stimulus onset to the stimulus offset. In Experiment 2, a set of dots was presented 350 ms before the stimulus offset, gradually increasing its contrast from 0% over 200 ms, reaching the specified contrast (i.e. high or low) 150 ms before the stimulus offset (figure 1*c*). The initial ramp-up period of 200 ms was introduced to prevent the abrupt appearance of the dots because our preliminary experiment suggested that such a short presentation of the dots for 150 ms may mask the perception of time and interfere with the time-numerosity integration. Thus, Experiments 1 and 2 differed in terms of how much of a gap there was between when *time* and *numerosity* information became available; *time* was available only at the offset, *numerosity* was always available from the onset in Experiment 1, while *numerosity* also became available near the offset in Experiment 2. In other words, the gap in the timing for when *time* and *numerosity* information became available was smaller in Experiment 2 than in Experiment 1.

2.4. Procedure

The conditions (T, N and TN) were presented in separate blocks. Before each block, participants were cued on whether they had to make judgements about *time*, *numerosity*, or both *time* and *numerosity*. In each trial, the standard and comparison stimuli were presented with an inter-stimulus interval (ISI) sampled from a uniform distribution between 450 and 550 ms. The order of the standard and comparison stimuli was counterbalanced across trials. Participants were instructed to press *M* on a conventional keyboard if they perceived the second stimulus as larger in magnitude than the first stimulus, and *C* if they perceived the first stimulus as larger in magnitude than the second stimulus. After the response, the next trial started with an inter-trial interval sampled from a uniform distribution between 500 and 600 ms. In the actual experiment, no feedback was provided regarding the correct or incorrect responses.

Each experiment (i.e. Experiments 1 and 2) was conducted in two sessions on two different days. Each session comprised 21 blocks. These 21 blocks were further divided into three sub-blocks each of which comprised one block of T trials, two blocks of N trials, and four blocks of TN trials. The order of the blocks within the sub-blocks was randomized. Each block comprised either 36 trials (T and N blocks) or 54 trials (TN blocks). The two levels of dot contrast (high and low) and three types of standard stimuli ($\Delta = -1$, $\Delta = 0$ and $\Delta = 1$) were randomly presented within the same block. Specifically, each T block comprised 9 (comparison magnitude) × 4 (repetition) trials, each N block comprised 9 (comparison magnitude) × 2 (dot contrast, high and low) × 2 (repetition) trials and each TN block comprised 9 (comparison magnitude) × 2 (dot contrast, high and low) × 3 (standard stimuli, $\Delta = -1$, $\Delta = 0$, and $\Delta = 1$) trials. Thus, every combination of the different conditions was repeated 24 times. In total, 1944 trials were performed per participant in each experiment.

To familiarize the participants with the task and to allow them to learn that *time* and *numerosity* were perfectly correlated in the TN condition, each experimental session for both Experiments 1 and 2 was preceded by a short practice session. The practice session comprised four blocks of T, N and two TN conditions with the standard stimulus of $\Delta = 0$. After every response, the participants received feedback, i.e. the fixation cross was changed to green/red for 200 ms to indicate a correct/incorrect response.

2.5. Data analysis

For each participant and every 18 conditions, the proportion of trials for which the comparison stimulus was judged as larger was calculated as a function of the comparison magnitude. These data were fitted to a cumulative normal distribution to obtain a psychometric function using Palamedes toolbox 1.10.11 [31].

From the fitted psychometric functions, the point of subjective equality (PSE) and just-noticeable difference (JND) were obtained. The PSE corresponded to the magnitude at which the probability of 'larger' responses reached 50%. The JND was defined as the inverse of the slope parameter (JND = 1/β) of the fitted function. The empirical variance, σ^2_{emp}, was then computed using the following formula for each condition [22]:

$$\sigma^2_{\text{emp}} = \frac{\text{JND}^2}{2}.$$

In the main analysis, we used the empirical variances obtained from the unidimensional conditions (T and N) to determine the predicted weights and variances under the bidimensional conditions (TN). The predicted numerosity weights, $W_{N,\text{pred}}$, were computed as follows:

$$W_{N,\text{ pred}} = \frac{\sigma^2_{T,\text{emp}}}{\sigma^2_{T,\text{emp}} + \sigma^2_{N,\text{emp}}},$$

where $\sigma^2_{T,\text{emp}}$ and $\sigma^2_{N,\text{emp}}$ were empirical variances under the T and N conditions, respectively. Note that we did not analyse the predicted and empirical time weights, W_T, as they are determined by W_N, such that $W_T = 1 - W_N$. The predicted bidimensional variances, $\sigma^2_{TN,\text{pred}}$, were computed according to the following formula:

$$\sigma^2_{TN,\text{pred}} = \frac{\sigma^2_{T,\text{emp}}\sigma^2_{N,\text{emp}}}{\sigma^2_{T,\text{emp}} + \sigma^2_{N,\text{emp}}}.$$

The empirical numerosity weights were calculated from the PSEs obtained from the bidimensional conditions according to the method used in [32]. First, the PSEs were measured for the standard stimuli of $\Delta = -1$, $\Delta = 0$, and $\Delta = 1$, and were plotted as a function of Δ. A regression line was then fitted to the data, and the empirical numerosity weights, $W_{N,\text{emp}}$, were computed as follows:

$$W_{N,\text{emp}} = \frac{1 - \text{slope}}{2},$$

where *slope* was the estimated slope of the regression line [32]. Finally, the empirical bidimensional variances in the TN conditions, $\sigma^2_{\text{TN,emp}}$, were defined as the average of the empirical variances for $\Delta = -1$, $\Delta = 0$ and $\Delta = 1$ for each dot-contrast level [24]. Note that in the subsequent analysis, the standard deviation (σ) rather than the variance was used as an index of the variability parameter.

For statistical testing, a two-way repeated-measures ANOVA was conducted on the weights and variability, with prediction (predicted and empirical) and contrast (high and low) as within-subjects factors. For each dot-contrast level, the predicted and empirical weights/variability were compared using two-sided paired-sample *t*-tests. Moreover, for each dot-contrast level, the empirical bidimensional variability ($\sigma_{TN,\text{emp}}$) was compared with the empirical unidimensional variability of the most reliable dimension ($\sigma_{T,\text{emp}}$ or $\sigma_{N,\text{emp}}$) using two-sided paired-sample *t*-tests. Additionally, we calculated the correlation coefficients between the predicted and empirical weights/variability. These statistical tests were further assessed by evaluating the null hypothesis and the strength of the evidence by computing Bayes factors using JASP [33]. BF_{10} represents evidence in favour of the alternative hypothesis that there is a difference/correlation. $\text{BF}_{10} > 3$ indicates support for the alternative hypothesis, whereas $\text{BF}_{10} < 1/3$ indicates support for the null hypothesis that there is no difference/correlation.

3. Results

3.1. Experiment 1

Figure 2*a* shows the across-participants mean of the psychometric function in the unidimensional (T, N) and bidimensional (TN) conditions. As expected, the slope of the psychometric function for the unidimensional numerosity conditions was flatter when the dot contrast was low than when it was high, indicating that the reliability of the numerosity information indeed varied with the dot contrast. In the low-contrast TN condition, the psychometric functions were almost invariant with Δ, and they overlapped. By contrast, the psychometric functions for the high-contrast TN condition shifted according to the value of Δ. Specifically, the psychometric function shifted to the left when $\Delta = 1$ and

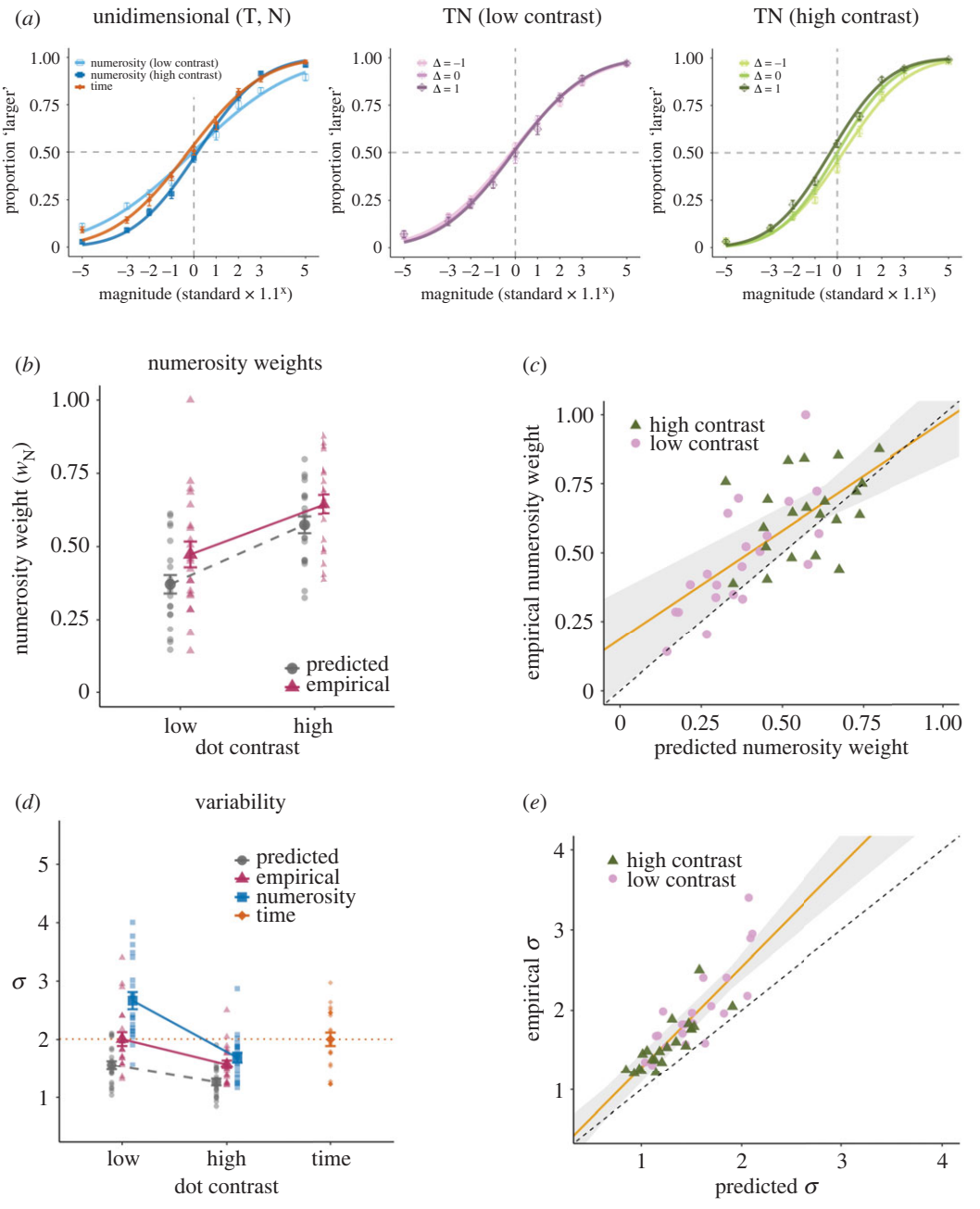

Figure 2. Psychometric functions, numerosity weights, and variability in Experiment 1. (*a*) The across-participants mean of the psychometric function in the unidimensional (time, numerosity) and bidimensional (time-numerosity) conditions separately for each contrast level and Δ. The Δ in the bidimensional conditions indicates that time for the standard stimulus was 600 ms multiplied by 1.1^{Δ} while numerosity was 24 dots multiplied by $1.1^{-\Delta}$ (see the 'Methods' section). The proportion of trials for which the comparison stimulus was judged as larger in magnitude was plotted as a function of the comparison magnitude. (*b*) The across-participants mean of the predicted (circles) and empirical (triangles) numerosity weights. Each data point represents an individual participant. (*c*) Individual predicted and empirical numerosity weights. Symbols indicate contrast levels: high (triangles) and low (circles). The solid line indicates the linear fit of the empirical weights to the predicted weights. The shaded region indicates the 95% confidence interval for the linear fit. (*d*) The across-participants mean of the predicted variability (circles) and empirical variability in the bidimensional (triangles), time (diamonds) and numerosity (squares) conditions. Each data point represents an individual participant. (*e*) Individual predicted and empirical variability. Symbols indicate contrast levels: high (triangles) and low (circles). The solid line indicates the linear fit of the empirical variability to the predicted variability. The shaded region indicates the 95% confidence interval for the linear fit. Error bars indicate the standard error of the mean (s.e.m.).

shifted to the right when $\Delta = -1$ relative to when $\Delta = 0$. This trend suggests that the PSEs were pulled toward the magnitude of the *numerosity* dimension, indicating that the weight of numerosity was greater than that of time. Based on the PSEs and JNDs obtained from these psychometric functions,

we computed the predicted weights and variability according to the MLE model and compared them with the empirical values (see 'Data analysis').

3.1.1. Weights

Figure 2*b* shows the predicted and empirical numerosity weights. A repeated-measures ANOVA revealed significant main effects for prediction and for contrast: $F_{1,\ 20} = 13.28$, $p = 0.002$, $BF_{10} = 71.77$, $\eta_p^2 = 0.399$; and $F_{1,\ 20} = 72.99$, $p < 0.001$, $BF_{10} > 10000$, $\eta_p^2 = 0.785$, respectively. There was no significant interaction: $F_{1,\ 20} = 0.61$, $p = 0.445$, $BF_{10} = 1.48$, $\eta_p^2 = 0.030$. As expected, numerosity weights were larger when the contrast was higher. However, the empirical numerosity weights were larger than the predicted weights when the contrast was low: $t_{20} = 3.39$, $p = 0.003$, $BF_{10} = 14.37$, $d = 0.74$; however, they were not significantly different when the contrast was high: $t_{20} = 2.03$, $p = 0.056$, $BF_{10} = 1.25$, $d = 0.44$. Moreover, there was a significant correlation between the predicted and empirical numerosity weights: $r = 0.692$, $p < 0.001$, $BF_{10} > 10000$ (figure 2*c*).

3.1.2. Variability

Figure 2*d* shows the predicted and empirical variability (σ) in the bidimensional conditions, along with the variability in the unidimensional conditions. A repeated-measures ANOVA revealed significant main effects for prediction and for contrast: $F_{1,\ 20} = 63.26$, $p < 0.001$, $BF_{10} > 10000$, $\eta_p^2 = 0.760$; and $F_{1,\ 20} = 49.99$, $p < 0.001$, $BF_{10} > 10000$, $\eta_p^2 = 0.714$, respectively, and a significant interaction: $F_{1,\ 20} = 6.00$, $p = 0.024$, $BF_{10} = 3.40$, $\eta_p^2 = 0.231$. As expected, the variability in the N and TN conditions was smaller when the contrast was higher. However, the empirical variability was systematically larger than predicted. The deviation was more profound when the contrast was high than when it was low: $t_{20} = 7.73$, $p < 0.001$, $BF_{10} > 10000$, $d = 1.69$ (high contrast), and $t_{20} = 6.58$, $p < 0.001$, $BF_{10} = 9258.65$, $d = 1.44$ (low contrast). Moreover, the bidimensional variability was not smaller than the smallest variability in the unidimensional conditions; when the contrast was low, the bidimensional variability was not significantly different from the variability of time alone: $t_{20} = 0.01$, $p = 0.996$, $BF_{10} = 0.23$, $d = 0.00$; and when the contrast was high, it was not significantly different from the variability of numerosity alone: $t_{20} = 2.03$, $p = 0.055$, $BF_{10} = 1.26$, $d = 0.44$. Nevertheless, there was a significant correlation between the predicted and empirical variabilities: $r = 0.862$, $p < 0.001$, $BF_{10} > 10000$ (figure 2*e*).

These results suggest that the MLE predictions are partially supported for both weights and variability. However, the numerosity weights and bidimensional variability were systematically larger than predicted. Furthermore, we found no evidence to support a variance reduction due to integration, which was the critical prediction of the MLE model.

3.2. Experiment 2

Figure 3*a* illustrates the across-participants average of the psychometric function. As in Experiment 1, the psychometric functions for the unidimensional numerosity conditions indicated that the reliability of the numerosity information varied with the dot contrast. In the low-contrast TN condition, the psychometric functions shifted according to the value of Δ. The psychometric function shifted to the right when $\Delta = 1$ and shifted to the left when $\Delta = -1$ relative to when $\Delta = 0$. This trend suggests that the PSEs were pulled toward the magnitude of the *time* dimension, indicating that the weight of time was greater than that of numerosity. A smaller but similar trend was observed in the high-contrast TN condition. Based on the PSEs and JNDs obtained from these psychometric functions, we computed the predicted weights and variability and compared them with the empirical values.

3.2.1. Weights

Figure 3*b* shows the predicted and empirical numerosity weights. A repeated-measures ANOVA revealed a significant main effect for contrast: $F_{1,\ 20} = 46.20$, $p < 0.001$, $BF_{10} > 10\,000$, $\eta_p^2 = 0.698$. In support of the MLE model, there was no main effect for prediction: $F_{1,\ 20} = 0.01$, $p = 0.918$, $BF_{10} = 0.27$, $\eta_p^2 = 0.001$; neither was there a main effect for interaction: $F1,\ 20 = 3.65$, $p = 0.071$, $BF_{10} = 0.59$, $\eta_p^2 = 0.154$. Confirming these observations, we found no evidence of a difference between the empirical and predicted numerosity weights: $t_{20} = 1.49$, $p = 0.152$, $BF_{10} = 0.59$, $d = 0.36$ (high contrast) and $t_{20} = 0.85$, $p = 0.403$, $BF_{10} = 0.32$, $d = 0.19$ (low contrast). Moreover, there was a significant correlation between the predicted and empirical numerosity weights: $r = 0.476$, $p = 0.001$, $BF_{10} = 25.71$ (figure 3*c*).

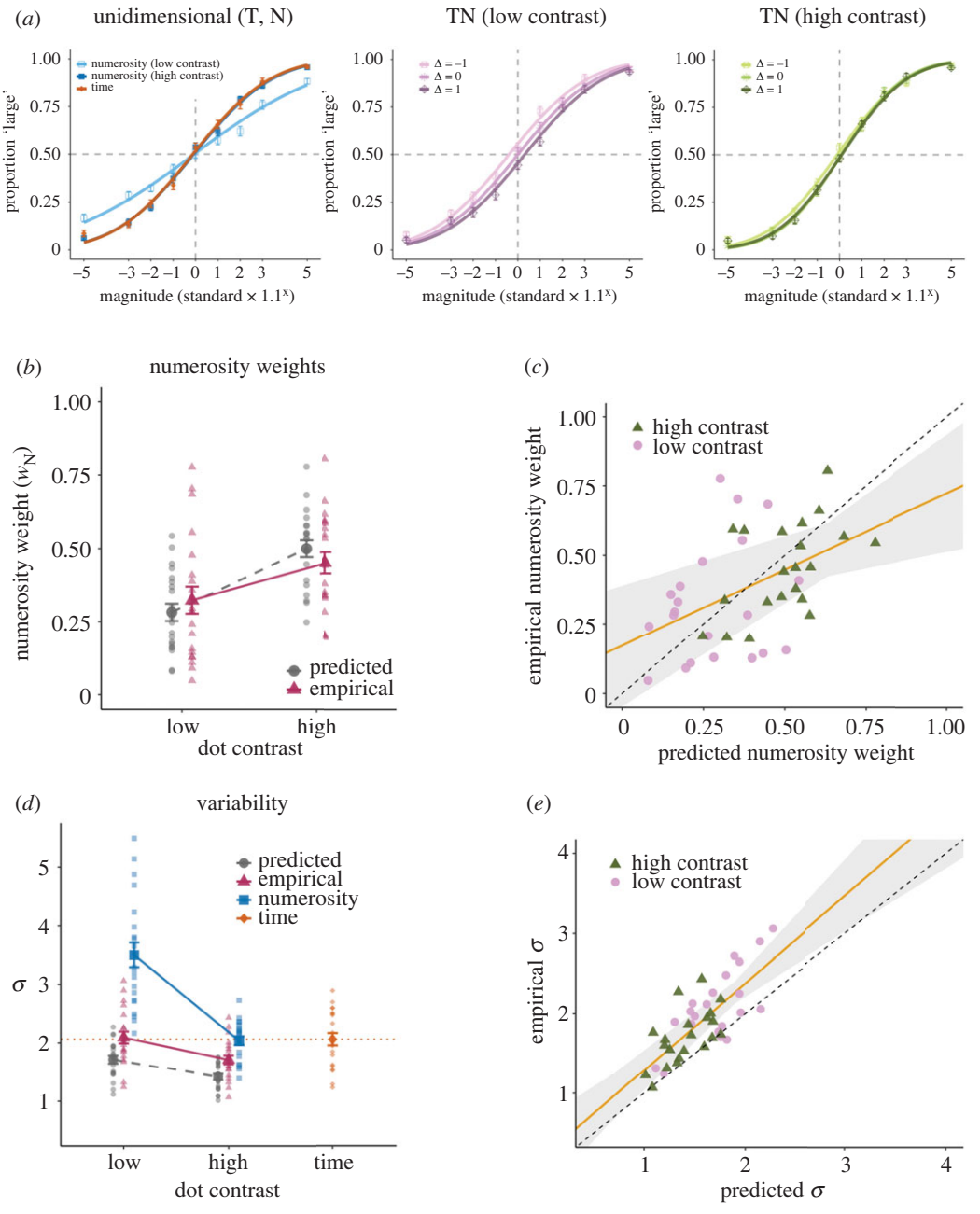

Figure 3. Psychometric functions, numerosity weights, and variability in Experiment 2. (*a*) The across-participants mean of the psychometric function in the unidimensional and bidimensional conditions separately for each contrast level and Δ. (*b*) The across-participants mean of the predicted (circles) and empirical (triangles) numerosity weights. Each data point represents an individual participant. (*c*) Individual predicted and empirical numerosity weights. Symbols indicate contrast levels: high (triangles) and low (circles). The solid line indicates the linear fit of the empirical to the predicted weights. The shaded region indicates the 95% confidence interval for the linear fit. (*d*) The across-participants mean of the predicted variability (circles) and empirical variability in the bidimensional (triangles), time (diamonds) and numerosity (squares) conditions. Each data point represents an individual participant. (*e*) Individual predicted and empirical variabilities. Symbols indicate contrast levels: high (triangles) and low (circles). The solid line indicates the linear fit of the empirical to the predicted variability. The shaded region indicates the 95% confidence interval for the linear fit. Error bars indicate s.e.m.

3.2.2. Variability

Figure 3*d* shows the predicted and empirical variabilities (σ). A repeated-measures ANOVA revealed significant main effects for prediction and for contrast: $F_{1,20} = 34.04$, $p < 0.001$, $BF_{10} > 10000$, $\eta_p^2 = 0.630$; and $F_{1,20} = 66.53$, $p < 0.001$, $BF_{10} > 10000$, $\eta_p^2 = 0.769$, respectively. There was no significant interaction: $F_{1,20} = 1.94$, $p = 0.179$, $BF_{10} = 1.69$, $\eta_p^2 = 0.088$. As in Experiment 1, the empirical variability was

systematically larger than the predicted variability: $t_{20} = 4.98$, $p < 0.001$, $BF_{10} = 366.72$, $d = 1.09$ (high contrast) and $t_{20} = 5.41$, $p < 0.001$, $BF_{10} = 891.48$, $d = 1.18$ (low contrast). Importantly, the bidimensional variability was smaller than either of the unidimensional variabilities when the contrast was high: $t_{20} = 3.82$, $p = 0.001$, $BF_{10} = 33.70$, $d = 0.83$ (versus numerosity) and $t_{20} = 3.69$, $p = 0.001$, $BF_{10} = 26.27$, $d = 0.81$ (versus time); however, it was not different from the variability of time alone when the contrast was low: $t_{20} = 0.30$, $p = 0.769$, $BF_{10} = 0.24$, $d = 0.07$. Moreover, there was a significant correlation between the predicted and empirical variabilities: $r = 0.753$, $p < 0.001$, $BF_{10} > 10000$ (figure 3*e*).

The bidimensional variability was not different from the variability of time alone in the low-contrast condition may be due to a small predicted improvement in variability; the MLE model states that the amount of improvement is determined by the degree to which the two cues differ in reliability. When the two cues are equally reliable, the greatest improvement is predicted. However, when one cue is clearly more reliable than the other, the predicted improvement is small, even if the two cues are optimally integrated. In our data, this explains why the predicted reduction in variability relative to time alone was smaller in the low-contrast condition than in the high-contrast condition (figure 3*d*). It is, therefore, difficult to confirm whether or not the integration occurred in the low-contrast condition. On the other hand, variance reduction during the high-contrast condition indeed suggests integration.

In summary, the results demonstrated that the empirical weights were consistent with the MLE predictions. However, the empirical variability was systematically larger than the predicted variability. Most importantly, the variability in the bidimensional condition was smaller than that in the time and numerosity conditions, supporting the MLE prediction of a variance reduction due to integration.

4. Discussion

This study aimed to test whether the integration of different magnitude information was statistically optimal, as predicted by the MLE model. As a test case, we examined the integration of time and numerosity. We found the following results from two experiments: first, when there was a large discrepancy in the timing of the acquisition of the time and numerosity information, numerosity was afforded more weight than predicted; the bidimensional variability was higher than predicted and comparable to the most reliable unidimensional variability. Second, as the discrepancy in the timing became smaller, the numerosity weights approached the MLE predictions; the bidimensional variability remained higher than predicted, yet smaller than the most reliable unidimensional variability.

Concerning experimental manipulation, Experiments 1 and 2 differed only in the time course of the stimulus. Therefore, we inferred that the lack of variance reduction from the unidimensional variability in Experiment 1 was likely due to a large discrepancy in the timing at which each piece of information became available. However, one might think that this does not explain why the numerosity weights were systematically larger than predicted in Experiment 1. Two possible explanations may accommodate the overweighting of numerosity: First, a bias toward numerosity may have occurred because numerosity was more salient than time due to the difference in the available duration (i.e. numerosity was available earlier and for a longer period than time). Second, observers may have assumed a prior over the reliability of numerosity information, based on a lifelong experience, that numerosity is a more reliable dimension than time when judging magnitude information [34]. Although we cannot rule out the latter possibility, the former seems more plausible as it parsimoniously accounts for the difference between Experiments 1 and 2 in terms of the bidimensional variability as well as the numerosity weights. In other words, the overweighting of numerosity and the lack of variance reduction in Experiment 1 may both be attributed to the earlier availability of numerosity information compared to time information. Relatedly, in Experiment 2, we included a 200 ms ramp-up period before the 150 ms presentation of the dots to prevent the abrupt appearance of the dots. While it is unlikely that the numerosity information was processed throughout this entire 350 ms, it might be the case that the numerosity processing had already completed before the stimulus offset even in Experiment 2, given that the numerosity processing completes around 150–250 ms after stimulus onset [35,36]. This could explain why the empirical variability remained higher than the predicted variability in Experiment 2. Further work is needed to assess whether limiting the presentation duration more would alter the efficacy of the magnitude integration. For example, restricting the presentation duration to less than 150 ms could hinder the processing of numerical magnitude [36] and discourage its integration with the other dimension.

Collectively, the present results suggest that human observers can integrate different magnitude information in a near-optimal manner. Previously, the MLE model has accounted for the integration of

multiple cues that provide redundant information about a single physical quantity, such as the visual-haptic integration of length [26], audiovisual integration of spatial location [24], and integration of various visual cues to depth [37,38]. By contrast, the current study demonstrates that the two quantities with different physical units, time and numerosity, are near-optimally integrated as abstract magnitude information. Furthermore, the contrasting results of our two experiments suggest that matching the availability of each piece of information is crucial for optimal magnitude integration. Interestingly, spatial–temporal discrepancies in the acquisition of information have been shown to affect the optimal integration of single physical quantities such as spatial location and orientation [39–41]. For example, Plaisier *et al*. [41] reported that, in the visual-haptic integration of surface orientation, optimal integration breaks down when there is a discrepancy between the visual and haptic exploration modes (i.e. the instantaneous perception of the surface orientation by touching/seeing two spots versus sequential perception by tracing the surface). This suggests that the time-numerosity interaction may result from a similar computational mechanism to the multisensory integration of single physical quantities.

In contrast to the multisensory integration of a single physical quantity, one might consider the integration of different magnitude dimensions to be outside the scope of the application of the MLE model. However, even in the audiovisual integration of spatial location, visual and auditory signals are assumed to be transformed into a common body-coordinate system [21]. This is conceptually similar to assuming that time and numerosity are transformed into a common magnitude representation [10]. Therefore, we believe it would be reasonable to apply the MLE model to the integration of multiple magnitude dimensions that have distinct physical units.

However, it should be noted that the near-optimal integration observed in our study does not necessarily mean that the magnitude integration shares a common mechanism with the cue integration of other features. In particular, the present study cannot determine whether the magnitude integration takes place at the perceptual or at the post-perceptual decision level. Still, our results indicate that the magnitude integration is affected by the timing at which each piece of information is acquired, suggesting that it could not be explained solely by the decision-level mechanisms, such as pure cognitive reasoning. Moreover, considering that magnitude interaction may occur not only at the decision level but also at the perceptual level [17,42,43], it is possible that magnitude integration partly involves perception-level mechanisms.

The present study employed a distinct experimental paradigm compared to those typically used to test magnitude interaction. In most previous studies, participants compared the magnitude of a specific dimension while the other dimension was independently varied. In the present study, however, the magnitudes of the two dimensions were perfectly correlated, and the participants were explicitly asked to integrate them. Although essential for the MLE modelling, these manipulations preclude us from assessing the cross-dimensional magnitude interaction bias (i.e. a bias on one dimension according to the other). Moreover, these manipulations were somewhat artificial in light of the relationship between magnitudes in the real world; in reality, different magnitudes do not necessarily correlate, and there are situations where integration may not be appropriate. Therefore, the generalizability of our results may be limited to scenarios where two magnitudes are redundant and must be combined.

Nevertheless, our findings provide further insight into the mechanism of the cross-dimensional magnitude interaction. Separate studies have hypothesized that the direction and the size of magnitude interaction are affected by the processing time course [3,12] and reliability [14,15,34] of each piece of information. However, the processing time course and reliability have not been independently manipulated within a single study, leaving uncertain the contribution of each factor to the interaction. For example, Togoli *et al*. [3] reported that a bidirectional interaction between time and numerosity occurred when numerosity information dynamically unfolded over time. While this result suggests that the time course of the stimulus and the neural processing dynamics it entails are important factors, dynamically unfolding numerosity information was accompanied by reduced reliability, and thus the results could also be explained by the reliability of the numerosity information. Meanwhile, Cai & Wang [15] proposed that the crucial factor for magnitude interaction is the relative representational noise of each dimension (i.e. reliability). However, their experiments used dynamic stimuli to manipulate the reliability of length information, confounded by the time course of the stimulus processing. In contrast to these studies, our study independently manipulated the stimulus's reliability and time course by varying the dots' contrast and the timing at which numerosity information became available. We found that numerosity weights flexibly varied with reliability (i.e. dot contrast), in accord with studies claiming the influence of reliability on magnitude interaction [14,15,34]. Moreover, integration was affected by the timing at which each piece of information was acquired, showing the influence of the time course of the stimulus processing [3,12].

Taken together, our findings suggest that both the reliability of each dimension and the time course of the stimulus processing potentially contribute to the cross-dimensional magnitude interaction. Further work is needed to test this prediction by examining magnitude interaction bias while independently manipulating the reliability and stimulus time course.

Our results for the near-optimal integration seem to align with the recent hypothesis that magnitude interaction reflects active magnitude binding rather than passive contextual interference between dimensions [17]. Togoli *et al*. [17] reported that a congruent integration effect was observed when time and numerosity belonged to the same stimulus, whereas an opposite, 'repulsive' effect was observed when they were conveyed by distinct stimuli. Regarding the latter repulsive effect, the authors postulate that it may reflect the brain's ability to disambiguate the magnitude of time and numerosity that are in strict spatial proximity but belong to distinct stimuli. Interestingly, a similar phenomenon has been reported in multisensory cue integration. Wallace *et al*. [44] reported that in the audiovisual perception of spatial location, the localization of an auditory target 'repulses away' from the location of the task-irrelevant visual stimulus when the spatial disparity between the audiovisual cues is sufficiently small but they are perceived as distinct events. This finding was later replicated and described well by the Bayesian causal inference model of multisensory cue integration, an extension of the MLE model used here [27,45,46]. Therefore, a promising direction for future work may be leveraging these advances in computational models of cue integration to delineate the mechanism of the magnitude interaction effect. Indeed, similar Bayesian models have already been proposed to account for spatio-temporal interferences [18,47,48]. Connecting these Bayesian models with our findings of the near-optimal magnitude integration would also be an interesting future direction.

Evidence of magnitude interaction has been accumulated not only between perceptions of different magnitudes but also between perception and action [49–51]. For example, Lindemann *et al*. [50] reported that perceived numerical magnitude affects the size of reaching movements (i.e. grip aperture). It is unclear whether these cross-dimensional magnitude interactions between perception and action can be explained by the reliability-based integration model, such as the MLE model. Meanwhile, sensorimotor integration in time perception has been well reported [52–55]. In particular, it has been suggested that time perception is closely related to the motor system, and some studies reported that sensorimotor integration of temporal information follows the reliability-based integration model [52,54]. Given the close link between time and other magnitude dimensions, it is conceivable that the cross-dimensional magnitude interaction between perception and action could also follow the reliability-based integration. This possibility awaits further testing.

Regarding the existing theory, the present results support the basic idea of ATOM [10] that different magnitude information is processed by a common magnitude system, which subsequently contributes to efficient motor output [11]. However, while the MLE model assumes independence of the noise associated with each piece of information, ATOM suggests that the encoding mechanism is partially shared across different magnitude dimensions [10,11]. Since these two assumptions are unlikely to be simultaneously satisfied, further theoretical development may require a modification of either model. Recently, it has been proposed that rather than absolute magnitudes, such as a single length and duration, relative magnitudes (i.e. the magnitude of the relationship between two absolute magnitudes, known as ratios) are processed by a common ratio processing mechanism [56]. For example, a recent behavioural study suggests that temporal and spatial ratios are processed by a domain-general ratio processing system [57]. Single-cell recordings in non-human primates and fMRI studies in humans also indicate the presence of ratio-selective neurons in the prefrontal and parietal cortex, regions often associated with the processing of absolute magnitudes [58,59]. By contrast, behavioural studies on cross-dimensional adaptation between space, time and numerosity found null or asymmetrical effects, indicating (partially) separate encoding mechanisms for different magnitudes [60–62]. Therefore, a modified view of ATOM might be that absolute magnitudes are initially encoded by (partially) separate mechanisms but are transformed into abstract relative magnitudes (i.e. ratios) by a generalized ratio processing system in later processing stages [63,64]. Assuming this view, the integration of magnitudes between dimensions with distinct physical units would occur after they are transformed into abstract ratios. Finally, the present results seem incompatible with Conceptual Metaphor Theory (CMT), a competing theory of ATOM [65]. CMT states that magnitudes of abstract dimensions, such as time, are represented by mapping them onto more concrete dimensions, such as space and number [65–68]. Unlike this prediction, however, the present results show that time and numerosity are equally weighted during integration according to their relative reliability if the experimental conditions are strictly controlled (but see [67,68]).

In the present experiments, the empirical variability was consistently higher than that predicted by the MLE, although there was a variance reduction apparently due to integration. This suggests that

magnitude integration may not satisfy some of the MLE assumptions [22]. Potential reasons include the possibilities that (i) there was a mix of subjects/trials that properly integrated the two magnitudes and subjects/trials that relied only on the reliable dimension without integrating the two magnitudes, (ii) the MLE model assumption that noises are normally distributed does not hold true for time and numerosity representations, and (iii) there is a correlation between the noises associated with the time and numerosity representations. In particular, the third possibility seems consistent with the prediction of ATOM that there is a partially common encoding mechanism for time and numerosity. Future research should consider each of these possibilities and further elaborate on the relationship between the MLE model and ATOM.

Ethics. The experimental protocol was approved by the Institutional Review Board of the University of Tokyo (#258-8) and was in accordance with the Declaration of Helsinki. All the participants provided written informed consent before the experiment.

Data accessibility. The data, scripts and details of the statistical results are available on Open Science Framework: https://osf.io/bx4qt/ [69].

Authors' contributions. T.O.: conceptualization, data curation, formal analysis, methodology, visualization, writing—original draft, writing—review and editing; Y.Y.: conceptualization, funding acquisition, methodology, project administration, resources, supervision, validation, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was supported by Grants-in-Aid for Scientific Research (grant nos. 22H01101 and 19H01771) to Y.Y., and JSPS Grant-in-Aid for JSPS fellows (grant no. 23KJ0681) to T.O.

Acknowledgements. We thank Editage (www.editage.com) for English language editing.

References

1. Hayashi MJ, Valli A, Carlson S. 2013 Numerical quantity affects time estimation in the suprasecond range. *Neurosci. Lett.* **543**, 7–11. (doi:10.1016/j.neulet.2013.02.054)
2. Javadi AH, Aichelburg C. 2012 When time and numerosity interfere: the longer the more, and the more the longer. *PLoS ONE.* **7**, e41496. (doi:10.1371/journal.pone.0041496)
3. Togoli I, Fornaciai M, Bueti D. 2021 The specious interaction of time and numerosity perception. *Proc. R. Soc. B* **288**, 20211577. (doi:10.1098/rspb.2021.1577)
4. Xuan B, Zhang D, He S, Chen X. 2007 Larger stimuli are judged to last longer. *J. Vis.* **7**, 1–5. (doi:10.1167/7.10.2)
5. Borghesani V, de Hevia MD, Viarouge A, Pinheiro-Chagas P, Eger E, Piazza M. 2019 Processing number and length in the parietal cortex: sharing resources, not a common code. *Cortex.* **114**, 17–27. (doi:10.1016/j.cortex.2018.07.017)
6. Dormal V, Dormal G, Joassin F, Pesenti M. 2012 A common right fronto-parietal network for numerosity and duration processing: an fMRI study. *Hum. Brain Mapp.* **33**, 1490–1501. (doi:10.1002/hbm.21300)
7. Harvey BM, Fracasso A, Petridou N, Dumoulin SO. 2015 Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. *Proc. Natl Acad. Sci. USA* **112**, 13 525–13 530. (doi:10.1073/pnas.1515414112)
8. Pinel P, Piazza M, Le Bihan D, Dehaene S. 2004 Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* **41**, 983–993. (doi:10.1016/S0896-6273(04)00107-2)
9. Skagerlund K, Karlsson T, Träff U. 2016 Magnitude processing in the brain: an fMRI study of time, space, and numerosity as a shared cortical system. *Front. Hum. Neurosci.* **10**, 500. (doi:10.3389/fnhum.2016.00500)
10. Walsh V. 2003 A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* **11**, 483–488. (doi:10.1016/j.tics.2003.09.002)
11. Bueti D, Walsh V. 2009 The parietal cortex and the representation of time, space, number and other magnitudes. *Phil. Trans. R. Soc. B* **364**, 1831–1840. (doi:10.1098/rstb.2009.0028)
12. Lambrechts A, Walsh V, Wassenhove Vv. 2013 Evidence accumulation in the magnitude system. *PLoS ONE* **8**, e82122. (doi:10.1371/journal.pone.0082122)
13. Martin B, Wiener M, van Wassenhove V. 2017 A Bayesian perspective on accumulation in the magnitude system. *Sci. Rep.* **7**, 1–14. (doi:10.1038/s41598-017-00680-0)
14. Cai ZG, Connell L. 2015 Space–time interdependence: evidence against asymmetric mapping between time and space. *Cognition* **136**, 268–281. (doi:10.1016/j.cognition.2014.11.039)
15. Cai ZG, Wang R. 2022 Cross-dimensional magnitude interaction is modulated by representational noise: evidence from space–time interaction. *Psychol. Res.* **86**, 196–208. (doi:10.1007/s00426-020-01472-4)
16. Schlichting N, de Jong R, van Rijn H. 2018 Robustness of individual differences in temporal interference effects. *PLoS ONE* **13**, e0202345. (doi:10.1371/journal.pone.0202345)
17. Togoli I, Bueti D, Fornaciai M. 2022 The nature of magnitude integration: contextual interference versus active magnitude binding. *J. Vis.* **22**, 1–15. (doi:10.1167/jov.22.11.11)
18. Cai ZG, Wang R, Shen M, Speekenbrink M. 2018 Cross-dimensional magnitude interactions arise from memory interference. *Cognit. Psychol.* **106**, 21–42. (doi:10.1016/j.cogpsych.2018.08.001)
19. Schlichting N, de Jong R, van Rijn H. 2020 Performance-informed EEG analysis reveals mixed evidence for EEG signatures unique to the processing of time. *Psychol. Res.* **84**, 352–369. (doi:10.1007/s00426-018-1039-y)
20. Alais D, Newell F, Mamassian P. 2010 Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving* **23**, 3–38. (doi:10.1163/187847510X488603)
21. Ernst MO, Bülthoff HH. 2004 Merging the senses into a robust percept. *Trends Cogn. Sci.* **4**, 162–169. (doi:10.1016/j.tics.2004.02.002)
22. Rohde M, van Dam LCJ, Ernst MO. 2016 Statistically optimal multisensory cue integration: a practical tutorial. *Multisensory Res.* **29**, 279–317. (doi:10.1163/22134808-00002510)
23. Warren DH, Welch RB, McCarthy TJ. 1981 The role of visual-auditory 'compellingness' in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept. Psychophys.* **30**, 557–564. (doi:10.3758/BF03202010)
24. Alais D, Burr D. 2004 The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262. (doi:10.1016/j.cub.2004.01.029)
25. Rohe T, Noppeney U. 2015 Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol.* **13**, e1002073. (doi:10.1371/journal.pbio.1002073)

26. Ernst MO, Banks MS. 2002 Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433. (doi:10.1038/415429a)
27. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. 2007 Causal inference in multisensory perception. *PLoS ONE* **2**, e943. (doi:10.1371/journal.pone.0000943)
28. Roach NW, Heron J, McGraw PV. 2006 Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc. R. Soc. B* **273**, 2159–2168. (doi:10.1098/rspb.2006.3578)
29. Brainard DH. 1997 The psychophysics toolbox. *Spat. Vis.* **10**, 433–436. (doi:10.1163/156856897(00357)
30. Cicchini GM, Mikellidou K, Burr DC. 2018 The functional role of serial dependence. *Proc. R. Soc. B* **285**, 20181722. (doi:10.1098/rspb.2018.1722)
31. Prins N, Kingdom FAA. 2018 Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the palamedes toolbox. *Front. Psychol.* **9**, 1250. (doi:10.3389/fpsyg.2018.01250)
32. Hartcher-O'Brien J, Alais D. 2011 Temporal ventriloquism in a purely temporal context. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1383–1395. (doi:10.1037/a0024234)
33. JASP Team. 2022 JASP (Version 0.16.4)[Computer software]. See https://jasp-stats.org/.
34. Dormal V, Pesenti M. 2013 Processing numerosity, length and duration in a three-dimensional Stroop-like task: towards a gradient of processing automaticity? *Psychol. Res.* **77**, 116–127. (doi:10.1007/s00426-012-0414-3)
35. Fornaciai M, Brannon EM, Woldorff MG, Park J. 2017 Numerosity processing in early visual cortex. *Neuroimage.* **157**, 429–438. (doi:10.1016/j.neuroimage.2017.05.069)
36. Fornaciai M, Park J. 2018 Early numerosity encoding in visual cortex is not sufficient for the representation of numerical magnitude. *J. Cogn. Neurosci.* **30**, 1788–1802. (doi:10.1162/jocn_a_01320)
37. Jacobs RA. 1999 Optimal integration of texture and motion cues to depth. *Vision Res.* **39**, 3621–3629. (doi:10.1016/S0042-6989(99)00088-7)
38. Landy MS, Maloney LT, Johnston EB, Young M. 1995 Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res.* **35**, 389–412. (doi:10.1016/0042-6989(94)00176-M)
39. Hong F, Badde S, Landy MS. 2022 Repeated exposure to either consistently spatiotemporally congruent or consistently incongruent audiovisual stimuli modulates the audiovisual common-cause prior. *Sci. Rep.* **12**, 1–19. (doi:10.1038/s41598-022-19041-7)
40. Odegaard B, Wozny DR, Shams L. 2017 A simple and efficient method to enhance audiovisual binding tendencies. *PeerJ.* **5**, e3143. (doi:10.7717/peerj.3143)
41. Plaisier MA, van Dam LCJ, Glowania C, Ernst MO. 2014 Exploration mode affects visuohaptic integration of surface orientation. *J. Vis.* **14**, 22, 1–12. (doi:10.1167/14.13.22)
42. Hayashi MJ, Kanai R, Tanabe HC, Yoshida Y, Carlson S, Walsh V, Sadato N. 2013 Interaction of numerosity and time in prefrontal and parietal cortex. *J. Neurosci.* **33**, 883–893. (doi:10.1523/JNEUROSCI.6257-11.2013)
43. Rammsayer TH, Verner M. 2014 The effect of nontemporal stimulus size on perceived duration as assessed by the method of reproduction. *J. Vis.* **14**, 17, 1–10. (doi:10.1167/14.5.17)
44. Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. 2004 Unifying multisensory signals across time and space. *Exp. Brain Res.* **158**, 252–258. (doi:10.1007/s00221-004-1899-9)
45. Noppeney U, Jones SA, Rohe T, Ferrari A. 2018 See what you hear – How the brain forms representations across the senses. *Neuroforum.* **24**, A169–A181. (doi:10.1515/nf-2017-A066)
46. Rohe T, Ehlis AC, Noppeney U. 2019 The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nat. Commun.* **10**, 1–17. (doi:10.1038/s41467-019-09664-2)
47. Chen Y, Zhang B, Kording KP. 2016 Speed constancy or only slowness: what drives the kappa effect. *PLoS ONE.* **11**, e0154013. (doi:10.1371/journal.pone.0154013)
48. Chen Y, Peng C, Avitt A. 2021 A unifying Bayesian framework accounting for spatiotemporal interferences with a deceleration tendency. *Vision Res.* **187**, 66–74. (doi:10.1016/j.visres.2021.06.005)
49. Ranzini M, Lugli L, Anelli F, Carbone R, Nicoletti R, Borghi A. 2011 Graspable objects shape number processing. *Front. Hum. Neurosci.* **5**, 147. (doi:10.3389/fnhum.2011.00147)
50. Lindemann O, Abolafia JM, Girardi G, Bekkering H. 2007 Getting a grip on numbers: numerical magnitude priming in object grasping. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 1400–1409. (doi:10.1037/0096-1523.33.6.1400)
51. Badets A, Pesenti M. 2010 Creating number semantics through finger movement perception. *Cognition* **115**, 46–53. (doi:10.1016/j.cognition.2009.11.007)
52. De Kock R, Zhou W, Datta P, Mychal Joiner W, Wiener M. 2023 The role of consciously timed movements in shaping and improving auditory timing. *Proc. R. Soc. B* **290**, 20222060. (doi:10.1098/rspb.2022.2060)
53. Iordanescu L, Grabowecky M, Suzuki S. 2013 Action enhances auditory but not visual temporal sensitivity. *Psychon. Bull. Rev.* **20**, 108–114. (doi:10.3758/s13423-012-0330-y)
54. Shi Z, Ganzenmüller S, Müller HJ. 2013 Reducing bias in auditory duration reproduction by integrating the reproduced signal. *PLoS ONE.* **8**, e62065. (doi:10.1371/journal.pone.0062065)
55. Zatorre RJ, Chen JL, Penhune VB. 2007 When the brain plays music: auditory–motor interactions in music perception and production. *Nat. Rev. Neurosci.* **8**, 547–558. (doi:10.1038/nrn2152)
56. Lewis MR, Matthews PG, Hubbard EM. 2016 Chapter 6 - neurocognitive architectures and the nonsymbolic foundations of fractions understanding. In *Development of mathematical cognition* (eds DB Berch, DC Geary, KM Koepke), pp. 141–164. San Diego, CA: Academic Press.
57. Lagacé-Cusiac R, Tremblay PF, Ansari D, Grahn JA. 2023 Investigating the relationships between temporal and spatial ratio estimation and magnitude discrimination using structural equation modeling: evidence for a common ratio processing system. *J. Exp. Psychol. Hum. Percept. Perform.* **49**, 108–128. (doi:10.1037/xhp0001062)
58. Vallentin D, Nieder A. 2010 Representations of visual proportions in the primate posterior parietal and prefrontal cortices. *Eur. J. Neurosci.* **32**, 1380–1387. (doi:10.1111/j.1460-9568.2010.07427.x)
59. Jacob SN, Nieder A. 2009 Tuning to non-symbolic proportions in the human frontoparietal cortex. *Eur. J. Neurosci.* **30**, 1432–1442. (doi:10.1111/j.1460-9568.2009.06932.x)
60. Anobile G, Burr DC, Iaia M, Marinelli CV, Angelelli P, Turi M. 2018 Independent adaptation mechanisms for numerosity and size perception provide evidence against a common sense of magnitude. *Sci. Rep.* **8**, 13571. (doi:10.1038/s41598-018-31893-6)
61. Otsuka T, Yotsumoto Y. 2022 Partially separable aspects of spatial and temporal estimations in virtual navigation as revealed by adaptation. *i-Percept.* **13**, 20416695221078880. (doi:10.1177/20416695221078878)
62. Tsouli A, Dumoulin SO, te Pas SF, van der Smagt MJ. 2019 Adaptation reveals unbalanced interaction between numerosity and time. *Cortex.* **114**, 5–16. (doi:10.1016/j.cortex.2018.02.013)
63. Cona G, Wiener M, Scarpazza C. 2021 From ATOM to GradiATOM: cortical gradients support time and space processing as revealed by a meta-analysis of neuroimaging studies. *Neuroimage.* **224**, 117407. (doi:10.1016/j.neuroimage.2020.117407)
64. Van Opstal F, Verguts T. 2013 Is there a generalized magnitude system in the brain? Behavioral, neuroimaging, and computational evidence. *Front. Psychol.* **4**, 435. (doi:10.3389/fpsyg.2013.00435)
65. Winter B, Marghetis T, Matlock T. 2015 Of magnitudes and metaphors: explaining cognitive interactions between space, time, and number. *Cortex.* **64**, 209–224. (doi:10.1016/j.cortex.2014.10.015)
66. Casasanto D, Boroditsky L. 2008 Time in the mind: using space to think about time. *Cognition* **106**, 579–593. (doi:10.1016/j.cognition.2007.03.004)
67. Pitt B, Casasanto D. 2022 The order of magnitude: why SNARC-like tasks (still) cannot support a generalized magnitude system. *Cogn Sci.* **46**, e13108. (doi:10.1111/cogs.13108)
68. Casasanto D, Pitt B. 2019 The faulty magnitude detector: why SNARC-like tasks cannot support a generalized magnitude system. *Cogn. Sci.* **43**, e12794. (doi:10.1111/cogs.12794)
69. Otsuka T, Yotsumoto Y. 2023 Data from: Near-optimal integration of the magnitude information of time and numerosity. *Open Science Framework*. (https://osf.io/bx4qt/)