

RESEARCH ARTICLE | *Sensory Processing*

Supramodal representation of temporal priors calibrates interval timing

Huihui Zhang^{1,2,3} and Xiaolin Zhou^{1,2,4,5,6}

¹Center for Brain and Cognitive Sciences, Peking University, Beijing, China; ²School of Psychological and Cognitive Sciences, Peking University, Beijing, China; ³School of Psychology, The University of Sydney, Sydney, New South Wales, Australia; ⁴Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China; ⁵Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China; and ⁶PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China

Submitted 30 November 2015; accepted in final form 5 June 2017

Zhang H, Zhou X. Supramodal representation of temporal priors calibrates interval timing. *J Neurophysiol* 118: 1244–1256, 2017. First published June 14, 2017; doi:10.1152/jn.01061.2015.—Human timing behaviors are consistent with Bayesian inference, according to which both previous knowledge (prior) and current sensory information determine final responses. However, it is unclear whether the brain represents temporal priors exclusively for individual modalities or in a supramodal manner when temporal information comes from different modalities at different times. Here we asked participants to reproduce time intervals in either a unisensory or a multisensory context. In unisensory tasks, sample intervals drawn from a uniform distribution were presented in a single visual or auditory modality. In multisensory tasks, sample intervals from the two modalities were randomly mixed; visual and auditory intervals were drawn from two adjacent uniform distributions, with the conjunction of the two being equal to the distribution in the unisensory tasks. In the unisensory tasks, participants' reproduced times exhibited classic central-tendency biases: shorter intervals were overestimated and longer intervals were underestimated. In the multisensory tasks, reproduced times were biased toward the mean of the whole distribution rather than the means of intervals in individual modalities. The Bayesian model with a supramodal prior (distribution of time intervals from both modalities) outperformed the model with modality-specific priors in describing participants' performance. With a generalized model assuming the weighted combination of unimodal priors, we further obtained the relative contribution of visual intervals and auditory intervals in forming the prior for each participant. These findings suggest a supramodal mechanism for encoding priors in temporal processing, although the extent of influence of one modality on another differs individually.

NEW & NOTEWORTHY Visual timing and auditory timing influence each other when time intervals in the two modalities are drawn from two adjacent distributions and are randomly intermixed. A Bayesian model with a supramodal prior (distribution of intervals from both modalities) outperforms the model using sensory-specific priors in describing participants' performance. A generalized model further reveals that the prior is represented as a weighted average of the distribution of time intervals from the two modalities, which differ individually.

supramodal; prior; Bayesian modeling; interval timing

TO SURVIVE IN A DYNAMIC environment, humans need to accurately estimate the status of the environment, make decisions, and take action. During this process, not only in the outside world, but also at every stage of neural processing, signals are accompanied with noise. Uncertainty is an inherent part of neural computation (Pouget et al. 2013); yet humans can still make proper responses to the outside world under such uncertainty. How does the human brain achieve meaningful representations of the current state in uncertainty? Recent evidence suggests that many human behaviors, from perception (Alais and Burr 2004; Ernst and Banks 2002; Jacobs 1999; Körding et al. 2007) to sensorimotor control (Körding and Wolpert 2004; Wei and Körding 2009), are consistent with the principles of Bayesian inference. In the Bayesian framework, a representation of the current state of the world is a posterior estimate, which is determined by information from different sources, including previously acquired knowledge (prior) and current sensory information (likelihood; Knill and Pouget 2004).

Interval timing, which is a basis for perception and action, has been shown to be consistent with Bayesian inference in various tasks, such as sensorimotor coincidence timing (Miyazaki et al. 2005), temporal order judgment (Miyazaki et al. 2006), and time estimation (Acerbi et al. 2012; Ahrens and Sahani 2011; Cicchini et al. 2012; Jazayeri and Shadlen 2010). Exposed either to simple uniform distributions (Cicchini et al. 2012; Jazayeri and Shadlen 2010) or to complex temporal contexts (e.g., highly skewed or bimodal distributions; Acerbi et al. 2012), participants can achieve an internal representation of temporal statistics with an approximation and use it to optimize their timing performance. However, these studies focus on situations in which temporal information comes from a single modality, whereas in reality temporal information often comes from multiple modalities.

When temporal information from different modalities is simultaneously presented, cross-modal interaction takes place, with perception dominated by one modality (e.g., auditory dominance over vision; Morein-Zamir et al. 2003; Repp and Penel 2002; Welch et al. 1986). According to Bayesian inference, this occurs because our brain allocates greater weight to the more reliable sensory processing (Hartcher-O'Brien et al. 2014). In these studies, temporal information from different modalities is linked to a single event or presented in the same time window. However, temporal information from different

Address for reprint requests and other correspondence: X. Zhou, School of Psychological and Cognitive Sciences, Peking University, Beijing 100871, China (e-mail: xz104@pku.edu.cn).

modalities can also be separated in time. Results are mixed concerning this kind of cross-modal interaction. Adaptation to auditory rhythm does not affect the perception of visual rhythm (Becker and Rasmussen 2007); in contrast, adaptation to auditory time intervals can influence the perception of the following visual apparent motion (Zhang et al. 2012). Training interval discrimination in the tactile modality can improve performance in a similar task in the auditory modality (Nagarajan et al. 1998), but the interval discrimination learning does not transfer from the auditory to the visual modality (Lapid et al. 2009). Whether finding the interaction or not, these studies only examined the unidirectional cross-modal influence from one modality to another. When temporal information randomly comes from different modalities at different times, will the brain abstract temporal information from all of the modalities to form a supramodal prior or abstract temporal information from individual modalities and encode priors exclusively for each modality?

To address this question, the present study exploited the central-tendency biases (biases toward the mean of the distribution) of timing (Jazayeri and Shadlen 2010; Lejeune and Wearden 2009). We asked participants to reproduce time intervals in different contexts: unisensory and multisensory. In the unisensory tasks, all of the sample intervals drawn from a uniform distribution were presented in a single modality (visual or auditory). The unisensory tasks were used to acquire the Weber fraction (Gibbon 1977; Malapani and Fairhurst 2002) of visual timing or auditory timing through Bayesian modeling (Jazayeri and Shadlen 2010). In the multisensory tasks, sample intervals were randomly presented in the visual or auditory modality in different trials. The auditory and visual time intervals were drawn from two adjacent uniform distributions, respectively, with the conjunction of the two distributions being equal to the distribution in the unisensory tasks. The goal of the multisensory tasks was to directly examine whether the human brain encodes temporal statistics in a supramodal or a modality-specific manner. If the brain achieves a supramodal representation of the priors, participants' reproduced times should be biased toward the mean of the whole distribution of the sample intervals from both modalities; otherwise they should be biased toward the means of intervals from individual modalities. To this end, we built a Bayesian model with parameters (Weber fraction) acquired in the unisensory tasks. We were interested in whether the model with a supramodal prior or the model with modality-specific priors would better describe participants' timing performance in the multisensory context.

MATERIALS AND METHODS

Participants. Six students (three women) from Peking University, aged 18–22 yr, participated in this study. All had normal or corrected-to-normal vision, normal hearing, and were naive to the purpose of the experiment. The participants gave informed consent before the experiment and were paid afterwards. The study was carried out in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the School of Psychological and Cognitive Sciences, Peking University.

Stimuli and experimental design. The experiment was executed in a dark, sound-attenuated room. Visual stimuli were presented on a 22-in. CRT monitor at a resolution of $1,024 \times 768$, driven by a computer at a refresh rate of 100 Hz. Participants viewed all of the stimuli binocularly from a distance of 57 cm, and their heads were

maintained as stationary by using a chin-rest. A headset (AKG K271 MKII) was used to deliver sound stimuli. Computer programs of this study were developed with Matlab 7.1 (MathWorks, Natick, MA) and Psychophysics Toolbox (Brainard 1997; Pelli 1997).

In each trial, participants were first presented a sample interval and were then asked to reproduce it by pressing the spacebar key on the keyboard (Fig. 1A). The sample time interval was presented in the visual or auditory modality. If the sample interval was presented in the visual modality, two successive white disks with a visual angle of 1.5° flashed for 100 ms on the computer screen, demarcating the interval. If the sample interval was presented in the auditory modality, two successive beeps (1,000 Hz, 62 dB, duration of 100 ms) emitted by the headset demarcated the interval.

Participants were required to complete four tasks, two unisensory and two multisensory. We chose 14 sample intervals evenly distributed and ranging from 810 to 1,200 ms. The 14 sample intervals were divided into 2 groups, the shorter 7 intervals in one group, and the longer 7 in the other. In the two multisensory tasks (Fig. 1B), the group of shorter intervals was presented in the auditory modality, and the group of longer intervals in the visual modality in the auditory-visual (AV) task, whereas the mapping between interval groups and modalities was reversed in the visual-auditory (VA) task. For two unisensory tasks, all of the sample intervals were presented in the same modality, either visual (VV task) or auditory (AA task). A previous study has shown that participants can learn the distributions of the intervals in ~500 trials (Jazayeri and Shadlen 2010). Here, to ensure that participants sufficiently learned the distribution of the intervals, each task had three sessions (672 trials per session), and only the last two sessions were analyzed. Participants did one session each day and were allowed to move on to the next task upon completion of the three sessions in the previous task. It required 12 days to finish all four tasks. The order of the four tasks was pseudo-randomly scheduled across participants. Participants completed first the two unisensory tasks and then the two multisensory tasks, or vice versa. They were informed of the modality conveying time intervals when they started each session. But they did not know in advance the modality for the next session, or the next task. They were not told the differences between the AV task and the VA task either.

Procedure. Each session of each task was divided into 12 blocks, with 56 trials in each block. In each block, the sample interval in each trial was randomly chosen from the 14 sample intervals, and each sample interval was repeated 4 times. Participants took a rest for at least 1 min between two successive blocks. In each trial, a sample time interval was presented to participants, and then participants reproduced it (Fig. 1A).

At the beginning of each trial, a white central fixation point subtending 0.5° of visual angle appeared on a gray screen. If the fixation point was a disk, it indicated that the sample interval t_s would be presented in the visual modality. If the fixation point was a cross, it indicated that the sample interval t_s would be conveyed using auditory beeps. The fixation point remained on the center of the screen until the participant reproduced the sample interval. For visual trials, after a variable delay ranging from 1,250 to 1,850 ms drawn randomly from a truncated exponential distribution, two successive white disks were presented 5° above the center to demarcate the onset and the offset of the sample interval t_s . For auditory trials, two successive auditory beeps were used to demarcate the sample interval t_s . Regardless of the modality in which the sample interval was presented, participants were instructed to press the space key on the keyboard t_s ms after the presentation of the second marker of the sample interval. The reproduced time t_p was measured from the end of the second marker to the time the key was pressed.

After the reproduction in each trial, participants immediately received feedback regarding their performance (i.e., "correct," "too short," or "too long") (Fig. 1C). "Correct" feedback was given by changing the fixation point from white to green and from 0.5 to 1.5° . Otherwise, words "too short" or "too long" were presented on the

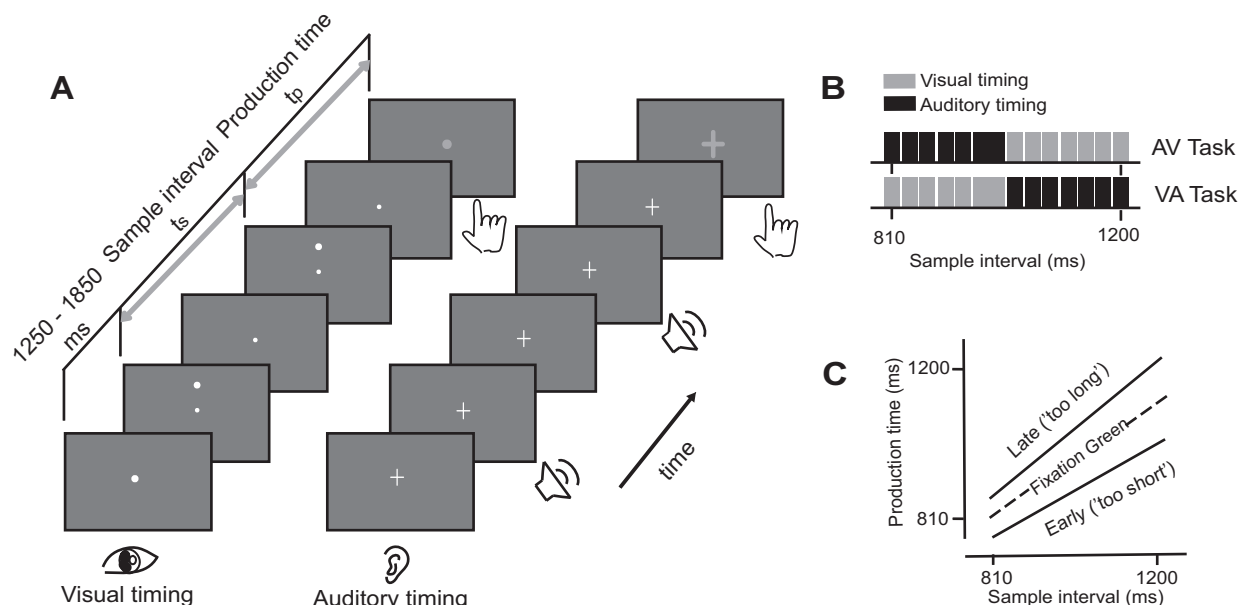


Fig. 1. Experimental design and procedure. *A*: procedure of a visual trial and an auditory trial. At the beginning of each trial, a central fixation point, a disk or a cross, was presented on a gray screen, indicating a visual trial or an auditory trial, respectively. Participants were required to maintain fixation throughout the trial. After a variable delay ranging from 1,250 ms to 1,850 ms, two successive white disks (visual trial) or two beeps (auditory trial) were used to demarcate the onset and the offset of the sample interval t_s ms. Then participants were instructed to press the spacebar key on the keyboard to reproduce t_s . The reproduced time t_p was measured from the end of the second temporal marker to when the key was pressed. After the reproduction, participants immediately received feedback regarding their performance (i.e., “correct,” “too short,” or “too long”). *B*: the illustration of the distribution of sample intervals in the multisensory tasks, AV and VA tasks. The 14 sample intervals were evenly distributed and ranged from 810 ms to 1,200 ms. In the AV task, the seven shorter intervals were presented in the auditory modality, and the seven longer intervals in the visual modality. In the VA task, the seven shorter intervals were presented in the visual modality, and the seven longer intervals in the auditory modality. *C*: the schematic drawing of the feedback schedule. If the reproduced time t_p was in an adjusted window around the sample interval t_s , the fixation point would turn green and increase in size (i.e., “correct”). If t_p was longer than the upper bound of the adjusted time window, the “too long” message was presented, and if t_p was shorter than the lower bound of the adjusted time window, the “too short” message was presented.

center of the screen. Participants achieved a “correct” response if the reproduced time t_p fell into a time window around the sample interval t_s ($t_s \pm k \times t_s$), a “too short” response if the reproduced interval was shorter than the lower bound of the time window (i.e., shorter than $t_s - k \times t_s$), and a “too long” response if the reproduced interval was longer than the upper bound of the time window (i.e., longer than $t_s + k \times t_s$). The parameter k was adaptively adjusted trial by trial based on participants’ performance. If a “correct” feedback was received in the given trial, 0.015 was subtracted from k . Otherwise, 0.015 was added to k . By controlling the width of the specified time window, participants received “correct” feedback in ~50% of trials for all four tasks so that participants’ performance was comparable across tasks.

RESULTS

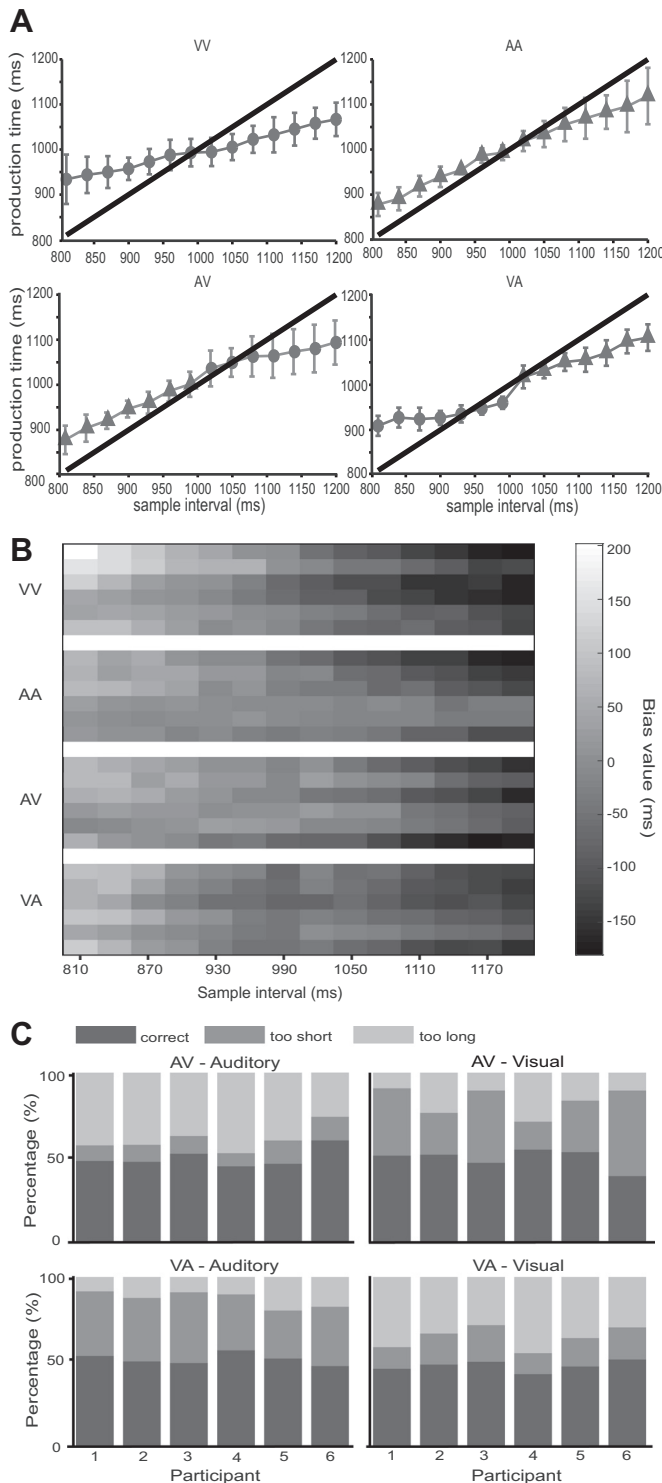
Behavioral results. Average production times across six participants are shown in Fig. 2A. For the unisensory tasks (VV and AA tasks), the production times showed the classic central-tendency biases: they were systematically biased toward the mean of the distribution of sample intervals (Jazayeri and Shadlen 2010). For the multisensory tasks (AV and VA tasks), the production times exhibited cross-modal central-tendency biases: they appeared to be biased toward the mean of the distribution of all of the 14 intervals from both the visual and auditory modalities.

To confirm this observation, we computed the biases for each participant using the mean production time for a given sample interval minus the given sample interval. As illustrated in Fig. 2B, in the four tasks, as the sample intervals became longer, the value of bias decreased gradually from positive to

negative. Moreover, for unisensory tasks, the mean absolute value of bias in the auditory task (AA task) was smaller than that in the visual task (VV task) (Wilcoxon signed-rank test, $P < 0.05$). We also computed the percentage of each type of feedback (“too short,” “too long,” and “correct”) in the AV and VA tasks (Fig. 2C). For all participants, the percentage of “correct” was around 50% for both visual trials and auditory trials in the two tasks. For auditory trials in the AV task and visual trials in the VA task, the percentage of “too short” feedback was less than the percentage of “too long” feedback (Wilcoxon signed-rank test, $P < 0.05$), indicating that participants tended to overestimate sample intervals that were of shorter durations. For auditory trials in the VA task, the percentage of “too short” feedback was greater than the percentage of “too long” feedback (Wilcoxon signed-rank test, $P < 0.05$), indicating that participants tended to underestimate these seven intervals that were of longer length. For visual trials in the AV task, however, the percentage of “too short” was not significantly different from the percentage of “too long” (Wilcoxon signed-rank test, $P = 0.075$).

We also computed the mean of the overall error ($\sqrt{\text{bias}^2 + \text{variance}}$) for each participant to confirm the completion of prior learning. For all of the participants, the overall error decreased $>10\%$ in the second session compared with the first session at least in one task. Their performance was relatively stable in the last two sessions. The difference in mean overall error between the last two sessions was $<10\%$ for participants 1–3 in all of the tasks. But for participant 4 in the

AV task, *participant 5* in both AV and VA tasks, and *participant 6* in VA tasks, the mean overall error decreased 34.5, 44.8, 11.3, and 23.6%, respectively in the third session compared with the second session, which may indicate that they were still at the stage of active learning. As shown in Fig. 3, their performance seems to change across sessions, from being biased toward the mean of the distribution of time intervals from both modalities to being biased toward the mean of the distribution of time intervals from individual modalities.



Bayesian observer model. To explore whether the human brain encodes temporal statistics through a supramodal mechanism, we built the same three-stage (measurement, estimation, and reproduction) Bayes least squares observer model as Jazayeri and Shadlen (2010) did. In the measurement stage, for a given interval t_s , the measured interval t_m by the brain differed from it because of measurement noise. As timing has a characteristic of scalar variability (i.e., the standard deviation of the estimation of given time increases linearly as a function of given time) (Gibbon 1977; Malapani and Fairhurst 2002), we assumed that conditional probability distribution $p(t_m|t_s)$ is Gaussian and centered at t_s with the standard deviation of $w_m t_s$. The coefficient w_m is the Weber fraction (ratio of the standard deviation to the mean), quantifying the amplitude of measurement noise. In the estimation stage, the observer uses t_m to get the estimated time, t_e , for a given sample interval t_s . The prior distribution of sample interval is discrete uniform, but we simplify it as continuous uniform. According to Bayes' rule, the posterior, $\pi(t_s|t_m)$, is the product of the prior multiplied by the likelihood function and then is normalized. For the Bayes least squares observer model, the cost function is the squared error function, $(t_e - t_s)^2$. The optimal estimated interval t_e can be derived by minimizing posterior expected loss, i.e., the integral of the cost function for each t_s , weighted by its posterior probability, $\pi(t_s|t_m)$. Similar to the measurement of t_s , in the production stage, the produced time t_p is also accompanied with motor-related noise. We assumed that the conditional $p(t_p|t_e)$ is Gaussian centered on t_e with standard deviation $w_p t_e$. The parameter w_p indicates the degree of noise for reproduction. Same as what Jazayeri and Shadlen (2010) did, we integrated out the two hidden variables, t_e and t_m , by marginalization. The relationship between the two psychophysically measurable variables, t_s and t_p , can be described by a Bayesian model with only two parameters, w_m and w_p . Details about the Bayesian observer model are provided in the APPENDIX.

Fig. 2. Behavioral results. *A*: average production times for participants in unisensory tasks (AA and VV) and multisensory tasks (AV and VA). For unisensory tasks, all of the sample intervals were presented in a single modality, visual (VV) or auditory (AA). In the AV task, seven shorter intervals were presented in the auditory modality, and the seven longer intervals were presented in the visual modality, whereas the mapping between interval groups and modalities was reversed in the VA task. The circles and triangles indicate the average reproduced time for a given visual and auditory sample interval, respectively. The solid diagonal line exhibits the unbiased condition with the reproduced time equal to the given time interval. Error bars represent SD. *B*: bias of time production for each participant in all four tasks. The rows are divided into four parts according to tasks (VV, AA, AV, and VA), in which each row represents one participant. Each column represents a given sample interval. The value of the bias is exhibited using gray values. *C*: the percentage of each kind of feedback participants received in the multisensory tasks. The "AV-Auditory" and "AV-Visual" conditions, respectively, indicate auditory trials and visual trials in the AV task. The "VA-Auditory" and "VA-Visual" conditions, respectively, indicate auditory trials and visual trials in the VA task. For the seven shorter sample intervals (i.e., "AV-Auditory" and "VA-Visual" conditions), participants received the "too short" feedback less than the "too long" feedback (Wilcoxon signed-rank test, $P < 0.05$ for both conditions). For the seven longer sample intervals, for "VA-Auditory" condition, participants received the "too short" feedbacks more than the "too long" feedbacks (Wilcoxon signed-rank test, $P < 0.05$); for "AV-Visual" condition, participants received the "too short" feedback the same amount of times as the "too long" feedback (Wilcoxon signed-rank test, $P = 0.075$).

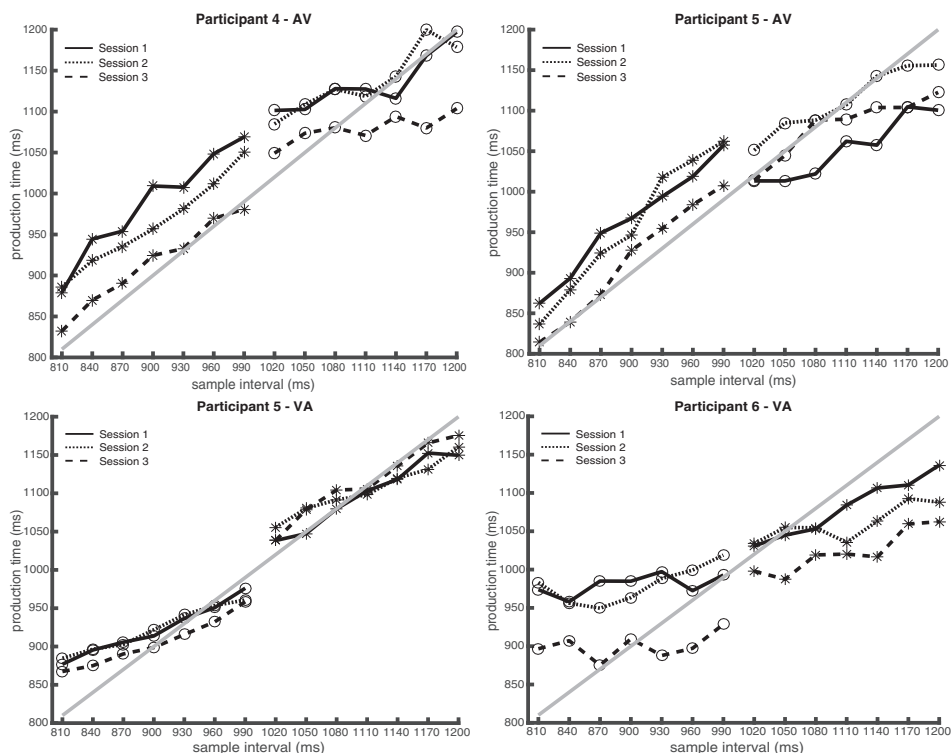


Fig. 3. Individual participants' average production times for all sample time intervals across sessions. *Top left:* participant 4's performance in the AV task. *Top right:* participant 5's performance in the AV task. *Bottom left:* participant 5's performance in the VA task. *Bottom right:* participant 6's performance in the VA task. The circles and asterisks indicate the average reproduced time for a given visual and auditory sample interval, respectively. The solid line, the dotted line, and the dashed line represent the first session, the second session, and the third session, respectively. The shaded diagonal line exhibits the unbiased condition with the reproduced time equal to the given time interval.

Estimating parameter w_m and w_p . In the unisensory VV and AA tasks, the prior distribution of sample intervals was uniform between 810 and 1,200 ms. We used participants' performance in tasks VV and AA to compute each participant's Weber fraction, w_m and w_p , through the Bayes least squares model (for details see APPENDIX). The values of w_m and w_p are shown in Fig. 4A. For the measurement Weber fraction, parameter w_m in visual timing was larger than that in auditory timing for all the participants (Wilcoxon signed-rank test, $P < 0.05$). This indicates that the auditory modality was more reliable than the visual modality in temporal processing. For the production Weber fraction w_p , w_p in the visual task (VV) was equal to that in the auditory task (AA) (Wilcoxon signed-rank test, $P = 0.075$), which indicates that uncertainty of reproduction was the same for visual and auditory interval timing.

Predicting participants' performance in AV and VA tasks. Since parameters w_m and w_p indicate the neural noise related to sensory measurement and time production, respectively, we assumed that they were relatively stable for each participant. That is, in the sensory measurement stage, the likelihood function $p(t_m|t_s)$ for the measurement t_m of the same value was assumed to be the same in the multisensory or unisensory tasks; in the production stage, the likelihood function $p(t_p|t_e)$ for reproduction t_p of the same value was assumed to be the same in the multisensory or unisensory tasks. We tested how the prior was encoded in the audiovisual context using parameter w_m and w_p acquired in the unisensory tasks (for details see APPENDIX).

For the multisensory tasks, we used two models to simulate each trial based on each participant's actual sequence of sample interval t_s . For each model, the simulation was conducted 10 times and averaged to reduce variability. *Model 1* assumed a supramodal prior, which means that participants

took the distribution of all of the 14 sample intervals from both modalities as the prior. *Model 2* assumed that participants used the distribution of seven sample intervals from one modality as a prior for interval timing in this modality, and the distribution of seven sample intervals from the other modality as a prior for interval timing in that modality. Figure 4B illustrates the mean of the produced time t_p given each sample interval t_s by participants and two models. As can be seen from this figure, the simulated performance of *model 1* with a supramodal prior (*middle column*) was more similar to participants' real behavior (*left column*) than the performance of *model 2* (*right column*). To confirm this observation, we compared the models' goodness of fit. A model's goodness of fit, F , was defined as the log likelihood of t_p given t_s for all of the trials, because both models have the same parameters. We used the value of w_m and w_p obtained from the unisensory task to compute the goodness of fit for each participant in each task. Figure 4C illustrates the relative goodness of fit of the two models, i.e., log likelihood ratio. As expected, *model 2*'s goodness of fit was worse than that of *model 1* for all of the participants in both VA and AV tasks, suggesting that participants encode temporal priors in a supramodal manner (*model 1*) rather than in a unisensory manner (*model 2*).

Individual priors in AV and VA tasks. Although *model 1* with a supramodal prior outperformed *model 2* with sensory-specific priors for all of the participants, there could be two potential problems. First, the extent of influence from the distribution of time intervals in one modality on the timing in another modality might be different for different participants. As shown in Fig. 2C, for visual trials in the AV task, the percentage of "too short" was not significantly different from the percentage of "too long" (Wilcoxon signed-rank test, $P = 0.075$). *Participant 2* and *participant 4* contributed most to the nonsignificance between "too short" and "too long" feedbacks.

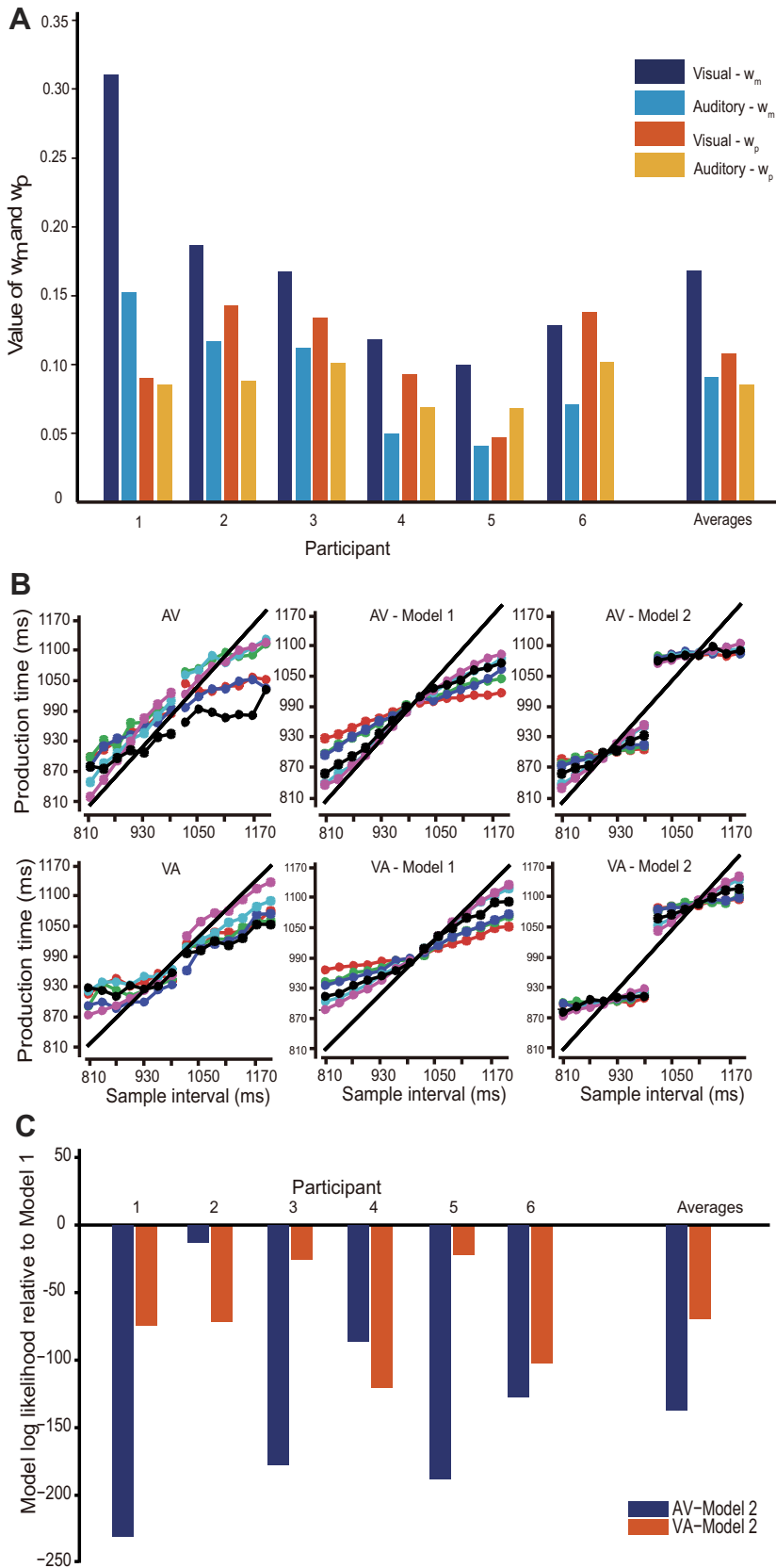


Fig. 4. Modeling results. *A*: parameter w_m and w_p estimated from the VV and AA tasks for each participant and for averages across participants. From *left to right*, the four bars in each group represent the sensory noise parameter w_m in the visual modality, sensory noise parameter w_m in the auditory modality, production noise parameter w_p in the visual task, and production noise parameter w_p in the auditory task. The parameter w_m of the visual modality was larger than that of the auditory modality for all of the participants (Wilcoxon signed-rank test, $P < 0.05$). For the production noise parameter w_p , w_p of the visual task (VV) was not significantly different from that of the auditory task (AA) (Wilcoxon signed-rank test, $P = 0.075$). *B*: the real and simulated performance for all six participants. Data patterns from *left to right* were the real data, simulated data by *model 1*, and simulated data by *model 2*. The *top row* illustrates the results in the AV task, and the *bottom row* illustrates the results in the VA task. Each circle indicates the mean value of t_p for a given sample interval t_s . Each color represents each participant. The black diagonal line represents the ideal situation, in which the produced time t_p is equal to the sample interval t_s . *C*: the model comparison for each participant and averages across participants in the AV and VA tasks. The vertical axis represents the log likelihood for *model 2* relative to *model 1*.

It conflicted with supramodal representation of priors, according to which participants should underestimate these longer visual time intervals, resulting in receiving more “too short” feedback. Second, we assumed that the noise parameters, w_m and w_p , remained the same across the unisensory and multisensory tasks, which may not be true.

To resolve the two problems, we generalized our model to capture individual differences in priors by employing weights and by acquiring the w_m and w_p directly for the multisensory tasks. In the new model (*model 3*), the prior is a weighted average of the prior distribution of visual time intervals and the prior distribution of auditory time intervals. The weight of the distribution of the sample intervals of shorter length (auditory intervals in AV task and visual intervals in VA task) is w , and the weight of the distribution of the sample intervals of longer length (visual intervals in AV task and auditory intervals in VA task) is $1 - w$. To help understand the relationship between *model 3* and the other two models, we used w_A and w_V to indicate the weight of distribution of sample intervals from the same modality for auditory timing and visual timing, respectively. For auditory timing, w_A (w in the AV task; $1 - w$ in the VA task) indicated the weight of the distribution of sample intervals from the auditory modality, where $1 - w_A$ indicated the weight of the distribution of sample intervals from visual modality. Similarly, for visual timing, w_V (w in the VA task; $1 - w$ in the AV task) indicated the weight of distribution of sample intervals from the visual modality, where $1 - w_V$ indicated the weight of the distribution of sample intervals from the auditory modality. If w_A and w_V equal to 0.5, the new model is the same as the model with a supramodal prior (*model 1*). If w_A and w_V equal to 1, the new model is the same as the model with sensory-specific priors (*model 2*). Thus, instead of qualitatively comparing whether the supramodal prior model (*model 1*) or the sensory-specific priors model (*model 2*) better explained participants' performance, we employed weights to quantify the relative contribution of sample intervals from two modalities. Results in the unisensory tasks showed that the parameter w_p quantifying motor noise was not different for visual timing and auditory timing. Hence we used one w_p for both visual and auditory timing. We used participants' performance in the AV and VA tasks to directly get each participant's Weber fraction, w_m and w_p , and the weight w through the new generalized Bayes least-squares model (see APPENDIX).

Figure 5A illustrates individual participants' weights of the prior distribution of time intervals from the same modality (w_A for auditory timing and w_V for visual timing) in the AV and VA tasks. For visual trials in the VA task, all the participants' weights were between 0.5 and 1, indicating that the distribution of visual sample intervals contributed more to the prior, but the distribution of auditory sample intervals also influenced the computed prior. For visual trials in the AV task, 5 out of 6 participants showed a similar pattern. For auditory timing in the AV task and VA task, weights varied across participants; on average, they were close to 0.5. Interestingly, for *participant 5*'s auditory timing in the AV task, the weight of auditory time intervals was close to 0. This happened also for *participant 3*'s auditory timing in the VA task. It suggests that their visual timing had large influence on their auditory timing in the audiovisual context.

Next, we simulated participants' performance again using the new model (*model 3*) with individual participants'

weighted priors and new w_m and w_p . As shown in Fig. 5B, the new model described the experimental data well. We used Akaike Information Criterion (AIC) to formally compare *model 1* (supramodal prior, three parameters: visual w_m , auditory w_m , and w_p), *model 2* (unisensory prior, three parameters: visual w_m , auditory w_m , and w_p), and *model 3* (weighted prior, five parameters: weight w_A for auditory timing, weight w_V for visual timing, visual w_m , auditory w_m , and w_p). As shown in Table 1, *model 3* was the best (lowest value for AIC) in describing each participant's performance in both AV and VA tasks. *Model 1* was better than *model 2* except for *participant 5* in the VA task; this pattern was consistent with the result of model comparison between *model 1* and *model 2* that we carried out before (Fig. 4C). For the previous model comparison (Fig. 4C), we used w_m and w_p acquired from unisensory tasks to calculate the goodness of fit for each model; here we acquired w_m and w_p directly from the AV and VA tasks to calculate the AIC. For *participant 5* in VA task, *model 2* was better than *model 1*, suggesting that *participant 5* acquired unisensory priors in VA task.

Nevertheless, *model 3* explained the data best. With *model 3*, we acquired w_m and w_p in AV and VA tasks. Then we compared the values of the w_m and w_p in different tasks (Table 2). The motor-related noise parameter, w_p , was stable across tasks, but the sensory measurement noise parameter, w_m , did vary across different tasks.

DISCUSSION

To study whether priors in interval timing are encoded in a supramodal manner or exclusively for individual modalities, we manipulated sensory contexts (unisensory vs. multisensory) in a time interval reproduction task. Our central finding was that, in the multisensory tasks (i.e., when visual intervals and auditory intervals drawn from two distinct but adjacent distributions were intermixed), participants' reproduced times showed cross-modal central-tendency biases: they were biased toward the mean of the distribution of sample intervals from both modalities. The Bayesian model with a supramodal prior better explained participants' timing performance in the multisensory context than the model with modality-specific priors. Using a Bayesian model with weighted priors, we further showed that the representation of priors was a weighted average of the distribution of visual sample intervals and the distribution of auditory sample intervals, although there were large individual differences in the weights. The mutual influences between visual and auditory interval timing suggest a centralized, supramodal mechanism for encoding priors in interval timing.

The finding of supramodal encoding of priors provides new insights on understanding how our brain processes temporal information from different modalities. The human brain can integrate simultaneously presented multisensory temporal information according to Bayesian inference by allocating greater weight to more reliable sensory processing (Burr et al. 2009; Hartcher-O'Brien et al. 2014). Here we showed that, when temporal information from different modalities are separated in time, outside the traditional integration window of several hundreds of milliseconds (Meredith et al. 1987; Spence and Squire 2003), participants can statistically learn them as a supramodal prior and use it to optimize their timing behavior.

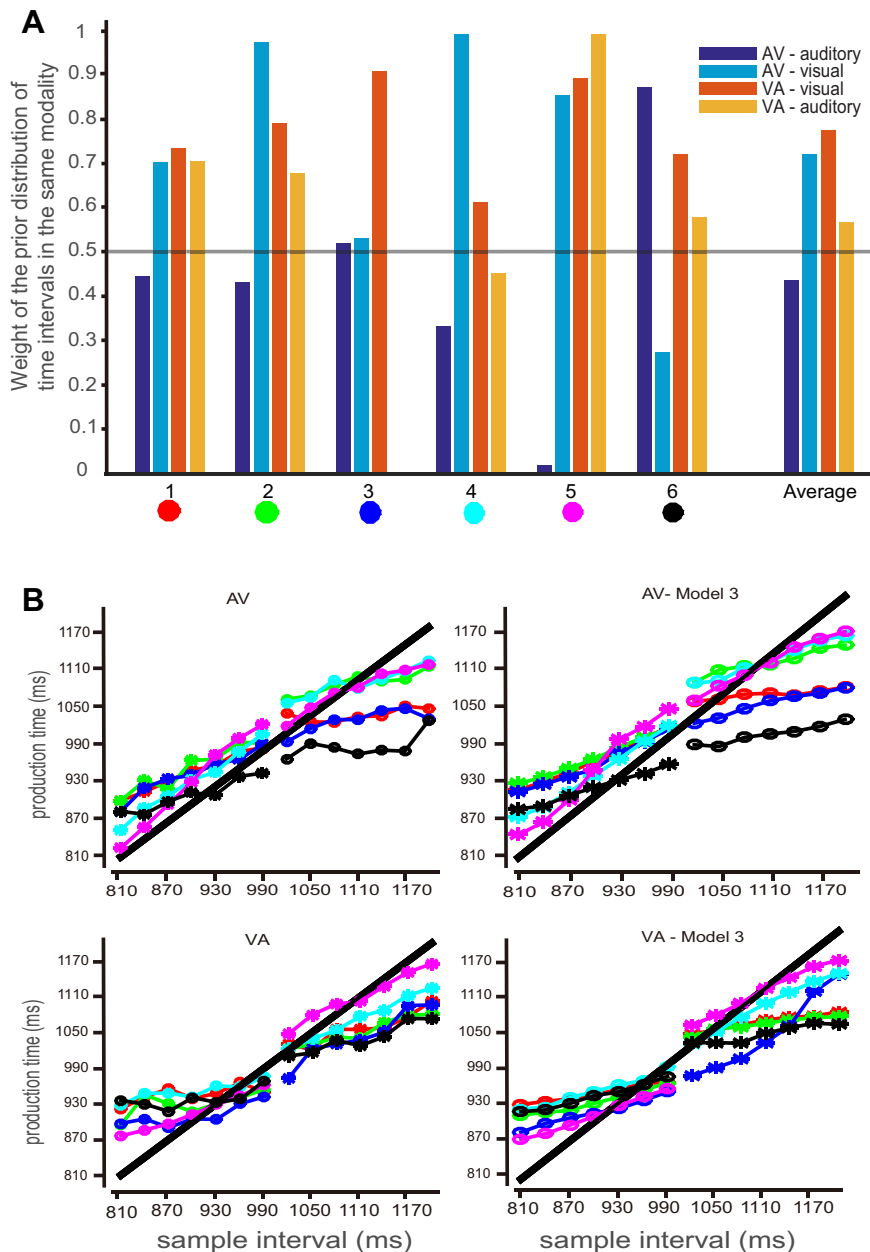


Fig. 5. Individual priors. *A*: the weight of the prior distribution of time intervals in the same modality for each participant and for averages across participants. From *left to right*, the four bars in each group represent the auditory timing in the AV task, the visual timing in the AV task, the visual timing in the VA task, and the auditory timing in the VA task. The color of circles below each group of bars represents individual participant. *B*: the experimental data and simulated data using the model with individual priors (*model 3*) for all six participants. *Left*: the experimental data. *Right*: the simulated data. The *top row* illustrates the results in the AV task, and the *bottom* illustrates the results in the VA task. Each circle (for visual timing) or asterisk (for auditory timing) indicates the mean value of t_p for a given sample interval t_s . Each color represents each participant, which is consistent with the color of circles below the bar graph in *A*. The black diagonal line represents the ideal situation, in which the produced time t_p is equal to the sample interval t_s .

This finding of multisensory contextual calibration in interval timing is in line with the modality effect (i.e., participants tend to overestimate auditory stimuli and underestimate visual stimuli of equivalent durations when auditory and visual stimuli are

intermixed within a session; Gu and Meck 2011; Penney et al. 2000; Wearden et al. 1998). This modality effect shows that there are mutual influences between duration perception of visual and auditory stimuli. The current finding expands the modality effect by showing that, when visual and auditory time intervals are of different durations, they can still affect each other. Recently, Shi and colleagues (2013) proposed a possible mapping from Bayesian timing to the information-processing model of interval timing with three stages: clock, memory, and decision (Church 1984; Meck 1983; Treisman 1963). They linked the likelihood to the clock stage and before the reference memory. The current results of Bayesian timing in the multisensory context can be interpreted in this framework: it is plausible that sample intervals from both sensory modalities are mixed and stored in memory to form the temporal priors. As shown in Fig. 5, the relative weights of the distribution of time intervals from two modalities were different for visual

Table 1. Model comparison: AIC for three models in multisensory tasks for each participant

Task	Model No.	AIC for Each Participant					
		1	2	3	4	5	6
AV	1	-2,765	-1,997	-2,159	-2,119	-3,112	-2,010
	2	-2,403	-1,855	-1,856	-2,038	-2,983	-1,634
	3	-2,819	-2,028	-2,155	-2,152	-3,211	-2,072
VA	1	-2,816	-2,095	-2,278	-2,917	-3,402	-1,467
	2	-2,557	-1,970	-2,173	-2,638	-3,467	-1,287
	3	-2,863	-2,130	-2,377	-2,921	-3,480	-1,480

AIC, Akaike Information Criterion.

Table 2. The w_m and w_p in different tasks

Parameter	Task	Participant No.						Mean \pm SD
		1	2	3	4	5	6	
w_m -Visual	VV	0.309	0.185	0.166	0.117	0.098	0.127	0.167 \pm 0.077
	AV	0.270	0.083	0.178	0.059	0.050	0.207	0.141 \pm 0.090
	VA	0.208	0.146	0.114	0.141	0.082	0.150	0.140 \pm 0.042
w_m -Auditory	AA	0.151	0.116	0.111	0.049	0.040	0.070	0.089 \pm 0.043
	AV	0.110	0.119	0.112	0.070	0.041	0.103	0.092 \pm 0.031
	VA	0.206	0.222	0.040	0.066	0.039	0.230	0.134 \pm 0.094
w_p	VV/AA	0.087	0.115	0.116	0.080	0.057	0.119	0.096 \pm 0.025
	AV	0.070	0.101	0.093	0.087	0.063	0.102	0.086 \pm 0.016
	VA	0.071	0.099	0.091	0.065	0.058	0.122	0.084 \pm 0.024

For w_p in the unisensory tasks, it was the mean of w_p acquired in the VV task and w_p acquired in the AA task.

timing and auditory timing in a certain task, suggesting that participants took on separate priors for visual timing and auditory timing. It is possible that participants knew there should be two separate priors for visual timing and auditory timing, but they could not accurately extract the distribution of sample intervals from an individual modality to form the prior for timing in that modality, because sample intervals from both modalities were mixed in memory. The representation of prior for visual timing is partially contributed by the distribution of auditory time intervals and vice versa, suggesting a supramodal mechanism underlying time perception.

Findings in the present study challenge the nonexistence of cross-modal influence in timing tasks (Becker and Rasmussen 2007; Lapid et al. 2009) and the proposal of modality-specific mechanisms for time perception. Using an adaptation task, Becker and Rasmussen (2007) showed that adaptation to auditory rhythm did not affect the perception of visual rhythm. With a learning paradigm, Lapid et al. (2009) found that interval discrimination learning did not transfer from the auditory to the visual domain. However, the paradigms used in these studies may not be sensitive enough to reveal the cross-modal influence. Indeed, using an implicit timing paradigm, we demonstrated in a previous study that adaptation to auditory time intervals can influence the perception of the subsequent visual apparent motion in which the time interval between the consecutive visual stimuli is crucial (Zhang et al. 2012). In the present study, the distribution of visual time intervals and the distribution of auditory time intervals were adjacent and in narrow temporal windows, and visual trials and auditory trials were intermixed. Both manipulations could be important for detecting the mutual influences between visual timing and auditory timing. If the two distributions were not adjacent, or they were adjacent but wider (e.g., 500–1,500 ms instead of current 810–1,200 ms), participants might be able to realize the modality difference and separate the distribution of visual time intervals and the distribution of auditory time intervals. Grahn and colleagues (2011) reported that auditory timing influenced visual timing but not vice versa in a block design. For the present setup, if a sensory-specific mechanism underlies time perception, we should have observed no mutual influence. The mutual influences between visual timing and auditory timing we observed here provides new evidence for a supramodal mechanism underlying time perception.

Participants' timing performance in the multisensory context was suboptimal. They could not extract the precise prior distribution of intervals in a certain modality, which resulted in

greater errors. As shown in Fig. 5A, participants' ability of extracting priors differed between individuals. The sequence of four tasks (AA-VV-AV-VA, VA-AV-AA-VV, AV-VA-VV-AA, AV-VA-AA-VV, VA-AV-VV-AA, and VV-AA-VA-AV, respectively, for the six participants) could not explain the variation. For example, while both *participant 3* and *participant 4* did the AV task first and then the VA tasks, *participant 3* showed a sensory-specific prior in visual timing in the VA task, while *participant 4* showed a sensory-specific prior in visual timing in the AV task. Overall, in both AV and VA tasks, five of six participants showed that the weight of the prior distribution of visual sample intervals in visual timing was greater than the weight of the prior distribution of auditory sample intervals in auditory timing. The weight was even close to zero for auditory timing in the AV task for *participant 5* and for auditory timing in the VA task for *participant 3*, suggesting that the distribution of visual sample intervals dominated the prior for auditory timing. Note that, the stronger influence of visual sample intervals in the present study might be due to the experimental setup. At the beginning of each trial, a visual disk or cross indicated whether it was a visual or auditory trial; at the end of each trial, feedback was also given in visual. The visual modality provided more task-related information, leading the participants to rely more on visual sample intervals in computing the supramodal prior.

The suboptimality of extracting sensory-specific priors was possibly due to insufficient learning in our tasks because we tested only three sessions for each task. Although most of participants' performance (*participants 1, 2, 3, and 6* in the AV task; *participants 1, 2, 3, and 4* in the VA task) was stabilized after the first session, it might change again with more sessions. It is plausible that the suboptimality of extracting priors in timing in the multisensory context occurs at the early stage of learning, and participants can acquire the sensory-specific priors with more learning sessions. As shown in Fig. 3, participants whose performance was not stabilized in the last two sessions (*participant 4* in the AV task, *participant 5* in both the AV and VA tasks, and *participant 6* in the VA task) seem to show a sign of changing priors from supramodal to sensory-specific over sessions. However, even if participants can acquire modality-specific priors in the future, the data still suggest that time is processed by a dedicated, supramodal mechanism, rather than a sensory-specific mechanism. With extensive learning, the supramodal mechanism underlying time perception may separate the visual sample intervals and auditory time intervals. But a sensory-specific mechanism cannot explain the present findings of mutual influences between visual timing and auditory timing.

Although in the present study visual timing and auditory timing influenced each other in a bidirectional manner, modality differences did exist. For the unisensory tasks, Weber fraction acquired from the Bayes least squares model was greater for visual timing than that for auditory timing, suggesting that temporal processing in the auditory modality is more precise than that in the visual modality (Grondin 2010; Welch et al. 1986; Westheimer 1999). Moreover, although both auditory timing and visual timing exhibited central-tendency biases in unisensory tasks, auditory reproduction biases were smaller than visual ones. These findings are consistent with previous studies showing that constant error (i.e., the difference between the estimated/produced and target intervals) and temporal vari-

ability in reproduction tasks are smaller using auditory than visual cues (Kolers and Brewster 1985). This phenomenon can be explained by Bayesian inference: for more uncertain sensory measurements, participants rely more on prior expectations, resulting in greater bias toward the mean of the prior distribution. Our results in the AA task were somewhat different from those of Cicchini et al. (2012), which showed that participants did not exhibit the central-tendency biases in the auditory interval timing tasks. Compared with the distribution ranging from 847 to 1,200 ms in their studies, we used a wider range of distribution (810–1,200 ms). According to Bayesian inference, a wider prior distribution results in greater biases, and the wider prior may have contributed to the observed auditory biases in the present study. Further studies are needed to discern whether the difference between the two studies was due to the difference in temporal distribution or due to other factors, e.g., the difference in task background (unisensory vs. multisensory).

The variability of both visual and auditory timing changed across different tasks (Table 1). For unisensory timing, the Weber fraction of visual timing in the VV task was greater than that of auditory timing in the AA task for all of the participants. However, for multisensory timing, the auditory timing was no longer more precise than visual timing in AV and VA tasks. The discrepancies of Weber fraction seem to be related to participants' ability of extracting sensory-specific priors. For example, for *participants 2, 4, and 5*, the weight of the distribution of visual time intervals for visual timing in AV task was >0.8 (Fig. 5A), and their Weber fraction of visual timing in AV task decreased ~50% compared with that in the VV task. For *participant 6*, the weight of the distribution of visual time intervals for visual timing in AV task was <0.5 (Fig. 5A), and the Weber fraction of visual timing in the AV task increased >50% compared with that in the VV task. Attention allocation might play a role in both the variability of timing and the extracting of sensory-specific priors. Previous studies showed that time perception was more variable when attention was divided by using multiple temporal targets (Brown and West 1990) or a nontemporal concurrent task (Brown 1997). It is possible that, in our multisensory tasks, when participants allocated more attentional resources to a certain modality, the timing in that modality was more accurate, and, meanwhile, they could extract more precise (sensory-specific) priors for the timing in that modality.

The supramodal encoding of priors may generalize to other tasks beyond interval timing. As priors represent the statistical information of the environment, intuitively, they should be extracted from previous sensory inputs and updated over time by a high-level mechanism. Indeed, previous studies have shown that likelihood and priors are represented independently (Beierholm et al. 2009; Vilares et al. 2012). Evidence from a variety of tasks such as size judgment (Morgan 1992) and spatial frequency discrimination (Lages and Treisman 1998) revealed that the central-tendency bias exists extensively across tasks (Hollingworth 1910; Petzschner et al. 2015). It is possible that the cross-modal central-tendency bias observed in the present study occurs for these tasks when information from different modalities is presented in an intermixed manner. This calls for future research in line with the present study.

In conclusion, by using an interval reproduction paradigm and by presenting temporal information either through a single

modality or through two modalities, we examined how priors are encoded and to what extent they affect participants' performance in reproduction tasks. In both the unisensory and multisensory tasks, we observed similar central-tendency biases in which the reproduced times were systematically biased toward the mean of distributions of all of the sample intervals. The Bayesian modeling further indicated that participants encoded priors in interval timing through a supramodal, centralized mechanism. The present study illustrates a new approach toward multisensory interaction.

APPENDIX

Bayesian observer model. For a given interval t_s , the measured interval t_m differs from t_s because of measurement noise. As timing has a characteristic of scalar variability (i.e., the standard deviation of the estimation of given time increases linearly as a function of given time), we assumed that conditional probability distribution $p(t_m|t_s)$ is Gaussian and centered at t_s with the standard deviation of $w_m t_s$. Parameter w_m is the Weber fraction associated with measurement, quantifying the amplitude of measurement noise. The likelihood function $\lambda_{t_m}(t_s)$ could be written as:

$$\lambda_{t_m}(t_s) = p(t_m|t_s) = \frac{1}{\sqrt{2\pi(w_m t_s)^2}} e^{-\frac{(t_s - t_m)^2}{2(w_m t_s)^2}} \tag{A1}$$

The prior distribution of sample interval is discrete uniform with the minimum value t_s^{\min} and the maximum value t_s^{\max} . For simplification, we modeled the prior distribution $\pi(t_s)$ as continuous uniform.

$$\pi(t_s) = \begin{cases} \frac{1}{t_s^{\max} - t_s^{\min}} & \text{for } t_s^{\min} \leq t_s \leq t_s^{\max} \\ 0 & \text{otherwise} \end{cases} \tag{A2}$$

According to Bayes' rule, the posterior, $\pi(t_s|t_m)$, is the product of the prior multiplied by the likelihood function and is then normalized.

$$\pi(t_s|t_m) = \frac{\pi(t_s)p(t_m|t_s)}{\int \pi(t_s)p(t_m|t_s)dt_s} = \begin{cases} \frac{p(t_m|t_s)}{\int_{t_s^{\min}}^{t_s^{\max}} p(t_m|t_s)dt_s} & \text{for } t_s^{\min} \leq t_s \leq t_s^{\max} \\ 0 & \text{otherwise} \end{cases} \tag{A3}$$

Combined with the cost function $l(t_e, t_s)$, the estimated interval t_e could be derived from the posterior $\pi(t_s|t_m)$. Cost function $l(t_e, t_s)$ represents the cost of erroneously estimating t_s as t_e . The best estimated t_e is achieved by minimizing the posterior expected loss.

$$t_e = f_l(t_m) = \arg \min_{t_e} \left[\int l(t_e, t_s) \pi(t_s|t_m) dt_s \right] \tag{A4}$$

Note that the optimal estimate, t_e , is a deterministic function of the measured interval t_m . In this deterministic function $f_l(t_m)$, the subscript l reflects a particular cost function. Different Bayesian models can be built with different decision rules (cost functions). Jazayeri and Shadlen (2010) found that the Bayes least squares model with cost function, $(t_e - t_s)^2$, can successfully describe participants' performance in interval timing. In this study, we used the Bayes least squares model to describe participants' performance in audiovisual interval timing. The estimator function $f_{\text{BLS}}(t_m)$ corresponds to the mean of the posterior.

$$f_{\text{BLS}}(t_m) = \frac{\int_{t_s^{\min}}^{t_s^{\max}} t_s p(t_m|t_s) dt_s}{\int_{t_s^{\min}}^{t_s^{\max}} p(t_m|t_s) dt_s} \tag{A5}$$

Similar to the measurement of t_s , the produced time t_p is also accompanied with noise. We assumed that the conditional $p(t_p|t_e)$ is Gaussian centered on t_e with standard deviation $w_p t_e$. The parameter w_p indicates the degree of noise for reproduction.

$$p(t_p|t_e) = \frac{1}{\sqrt{2\pi(w_p t_e)^2}} e^{-\frac{(t_p - t_e)^2}{2(w_p t_e)^2}} \quad (A6)$$

We then built the mathematical model to describe how the Bayesian observer produces t_p from the given sample interval t_s . We canceled out variable t_m and t_e by several calculations. First, we applied the chain rule to split the joint conditional distribution of variables t_m , t_e , and t_p to three intervening conditional probabilities:

$$p(t_p, t_e, t_m|t_s, w_m, w_p) = p(t_p|t_e, t_m, t_s, w_m, w_p)p(t_e|t_m, t_s, w_m, w_p)p(t_m|t_s, w_m, w_p) \quad (A7)$$

Second, we simplified the above equation by using the relationship between these variables. For the first term on the *right-hand* side of the above Eq. A7, t_p is determined by t_e and w_p , which allows us to omit the other conditional variables (t_m , t_s , and w_m). For the second term, t_e is determined by t_m , so that t_s , w_m , and w_p can be canceled out. For the third term, w_p is not relevant to t_m . So Eq. A7 could be simplified as:

$$p(t_p, t_e, t_m|t_s, w_m, w_p) = p(t_p|t_e, w_p)p(t_e|t_m)p(t_m|t_s, w_m) \quad (A8)$$

Moreover, t_e is a determinate function of t_m , $t_e = f(t_m)$. So the conditional probability $p(t_e|t_m)$ could be written as a Dirac delta function.

$$p(t_p, t_e, t_m|t_s, w_m, w_p) = p(t_p|t_e, w_p)\delta[t_e - f(t_m)]p(t_m|t_s, w_m) \quad (A9)$$

Next, we omitted the dependence on the two hidden variables t_e and t_m by marginalization.

$$\begin{aligned} p(t_p|t_s, w_m, w_p) &= \iint p(t_p, t_e, t_m|t_s, w_m, w_p) dt_m dt_e \\ &= \iint p(t_p|t_e, w_p)\delta[t_e - f(t_m)]p(t_m|t_s, w_m) dt_m dt_e \\ &= \int p(t_p|f(t_m), w_p)p(t_m|t_s, w_m) dt_m \end{aligned} \quad (A10)$$

Using substitutions from Eqs. A1, A5, and A6, Eq. A10 describes the conditional probability of t_p , given the sample interval t_s and the model parameter w_m and w_p . So far, the relationship between the two psychophysically measurable variables, t_s and t_p , can be described by a Bayesian model with only two parameters, w_m and w_p .

Estimating parameter w_m and w_p . From Eq. A10 we acquired the conditional probability $p(t_p|t_s, w_m, w_p)$ given the sample interval t_s and model parameter w_m and w_p . For simplification, we assumed that the value of t_p is independent across trials, although these values may be correlated because of the adaptive feedbacks. So the conditional probability of all the N trials could be written as:

$$p(t_p^1, t_p^2, \dots, t_p^N|t_s, w_m, w_p) = \prod_{i=1}^N p(t_p^i|t_s, w_m, w_p) \quad (A11)$$

By taking logarithm of both sides, we could change the products into sums.

$$\log p(t_p^1, t_p^2, \dots, t_p^N|t_s, w_m, w_p) = \sum_{i=1}^N \log p(t_p^i|t_s, w_m, w_p) \quad (A12)$$

In the unisensory tasks (VV and AA task), we measured the reproduced time t_p given certain sample interval t_s using the psychophysical method. We gained the parameter w_m and w_p by maximizing the Eq. A12, which could be achieved with `fminsearch` command in MATLAB (MathWorks). Integrals of Eqs. A5 and A10 are not

analytically solvable and were thus approximated numerically using the trapezoidal rule.

Predicting participants' performance in AV and VA tasks. For the multisensory tasks (AV and VA tasks), we built two Bayesian models with different priors to explore whether or not the brain encodes priors in a supramodal manner. *Model 1* assumes a supramodal prior, which means that participants take the distribution of all of the 14 sample intervals from both modalities as the prior. *Model 2* assumes that participants take the sample interval from a single modality as a prior. For example, in the AV task, seven intervals presented in the auditory modality range from 810 to 990 ms, and seven intervals presented in the visual modality range from 1,020 to 1,200 ms. According to *model 1*, the prior for auditory and visual timing is the same, a uniform distribution from 810 to 1,200 ms, whereas in *model 2*, the prior for auditory timing is uniform from 810 to 990, and the prior for visual timing is from 1,020 to 1,200 ms. First, we simulated each trial based on each participant's real sequence of sample interval t_s using *model 1* and *model 2*. Then we quantitatively compared the goodness of fit of *model 1* and *model 2*. For a model's goodness of fit, F , we used the log likelihood of t_p given t_s in all of the N trials. We used the value of w_m and w_p obtained from the unisensory task to compute the goodness of fit, F , for each participant in each task.

$$F = \log p(t_p^1, t_p^2, \dots, t_p^N|t_s, w_m, w_p) = \sum_{i=1}^N \log p(t_p^i|t_s, w_m, w_p) \quad (A13)$$

Rebuilding individual priors. We generalized our model to capture individual differences in priors by employing weights. That is, in the multisensory AV and VA tasks, the prior was a weighted average of the prior distribution of visual time intervals and the prior distribution of auditory time intervals. Instead of Eq. A2, we modeled the weighted prior distribution for visual timing or auditory timing in Eq. A14, where w is the weight, ranging from 0 to 1. The weight of the distribution of the sample intervals of shorter length (auditory intervals in AV task and visual intervals in VA task) is w , and the weight of the distribution of the sample intervals of longer length (visual intervals in AV task and auditory intervals in VA task) is $1 - w$. The t_s^{\min} , t_s^{medium} , and t_s^{\max} are the minimum, the mean, and the maximum of the distribution of time intervals from both modalities, i.e., 810 ms, 1,005 ms and 1,200 ms, respectively.

$$\pi(t_s) = \begin{cases} \frac{w}{t_s^{\text{medium}} - t_s^{\min}} & \text{for } t_s^{\min} \leq t_s < t_s^{\text{medium}} \\ \frac{1-w}{t_s^{\max} - t_s^{\text{medium}}} & \text{for } t_s^{\text{medium}} \leq t_s \leq t_s^{\max} \\ 0 & \text{otherwise} \end{cases} \quad (A14)$$

Thus we got the new posterior, $\pi(t_s|t_m)$ in Eq. A15.

$$\pi(t_s|t_m) = \begin{cases} \frac{wp(t_m|t_s)}{w \int_{t_s^{\min}}^{t_s^{\text{medium}}} p(t_m|t_s) + (1-w)} & \text{for } t_s^{\min} \leq t_s < t_s^{\text{medium}} \\ \frac{\int_{t_s^{\text{medium}}}^{t_s^{\max}} p(t_m|t_s) + (1-w)p(t_m|t_s)}{w \int_{t_s^{\min}}^{t_s^{\text{medium}}} p(t_m|t_s) + (1-w)} & \text{for } t_s^{\text{medium}} \leq t_s < t_s^{\max} \\ 0 & \text{otherwise} \end{cases} \quad (A15)$$

The new estimator function $f_{\text{BLS}}(t_m)$ is described by Eq. A16.

$$f_{BLS}(t_m) = \frac{w \int_{t_s^{\min}}^{t_s^{\text{medium}}} t_s p(t_m | t_s) dt_s + (1-w) \int_{t_s^{\text{medium}}}^{t_s^{\max}} t_s p(t_m | t_s) dt_s}{w \int_{t_s^{\min}}^{t_s^{\text{medium}}} p(t_m | t_s) dt_s + (1-w) \int_{t_s^{\text{medium}}}^{t_s^{\max}} p(t_m | t_s) dt_s} \quad (A16)$$

Similar to how we got w_m and w_p in the unisensory VV and AA tasks, we obtained the five parameters, w for visual timing, w for auditory timing, w_m for visual timing, w_m for auditory timing, and w_p , by maximizing Eq. A12, which could be achieved with `fmincon` command in MATLAB.

ACKNOWLEDGMENTS

We thank Dr. Philip R. Blue, Dr. Kunlin Wei, Dr. Lihan Chen, and Dr. Hang Zhang for constructive comments and suggestions concerning the early versions of the manuscript.

GRANTS

This study was supported by the National Basic Research Program of China (973 Program 2015CB856400).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

H.Z. and X.Z. conceived and designed research; H.Z. performed experiments; H.Z. analyzed data; H.Z. and X.Z. interpreted results of experiments; H.Z. prepared figures; H.Z. drafted manuscript; H.Z. and X.Z. edited and revised manuscript; H.Z. and X.Z. approved final version of manuscript.

REFERENCES

- Acerbi L, Wolpert DM, Vijayakumar S. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput Biol* 8: e1002771, 2012. doi:10.1371/journal.pcbi.1002771.
- Ahrens MB, Sahani M. Observers exploit stochastic models of sensory change to help judge the passage of time. *Curr Biol* 21: 200–206, 2011. doi:10.1016/j.cub.2010.12.043.
- Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14: 257–262, 2004. doi:10.1016/j.cub.2004.01.029.
- Becker MW, Rasmussen IP. The rhythm aftereffect: support for time sensitive neurons with broad overlapping tuning curves. *Brain Cogn* 64: 274–281, 2007. doi:10.1016/j.bandc.2007.03.009.
- Beierholm UR, Quartz SR, Shams L. Bayesian priors are encoded independently from likelihoods in human multisensory perception. *J Vis* 9: 1–9, 2009. doi:10.1167/9.5.23.
- Brainard DH. The psychophysics toolbox. *Spat Vis* 10: 433–436, 1997. doi:10.1163/156856897X00357.
- Brown SW. Attentional resources in timing: interference effects in concurrent temporal and nontemporal working memory tasks. *Percept Psychophys* 59: 1118–1140, 1997. doi:10.3758/BF03205526.
- Brown SW, West AN. Multiple timing and the allocation of attention. *Acta Psychol (Amst)* 75: 103–121, 1990. doi:10.1016/0001-6918(90)90081-P.
- Burr D, Banks MS, Morrone MC. Auditory dominance over vision in the perception of interval duration. *Exp Brain Res* 198: 49–57, 2009. doi:10.1007/s00221-009-1933-z.
- Church RM. Properties of the internal clock. *Ann N Y Acad Sci* 423: 566–582, 1984. doi:10.1111/j.1749-6632.1984.tb23459.x.
- Cicchini GM, Arrighi R, Cecchetti L, Giusti M, Burr DC. Optimal encoding of interval timing in expert percussionists. *J Neurosci* 32: 1056–1060, 2012. doi:10.1523/JNEUROSCI.3411-11.2012.
- Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429–433, 2002. doi:10.1038/415429a.
- Gibbon J. Scalar expectancy theory and Weber's law in animal timing. *Psychol Rev* 84: 279–325, 1977. doi:10.1037/0033-295X.84.3.279.
- Grahn JA, Henry MJ, McAuley JD. fMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. *Neuroimage* 54: 1231–1243, 2011. doi:10.1016/j.neuroimage.2010.09.033.
- Gronin S. Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Atten Percept Psychophys* 72: 561–582, 2010. doi:10.3758/APP.72.3.561.
- Gu BM, Meck WH. New perspectives on Vierordt's law: memory-mixing in ordinal temporal comparison tasks. In: *Multidisciplinary Aspects of Time and Time Perception. Lecture Notes in Computer Science*, edited by Vatakis A, Esposito A, Giagkou M, Cummins F, Papadelis G. Berlin: Springer, 2011, vol. 6789, p. 67–78. doi:10.1007/978-3-642-21478-3_6.
- Hartcher-O'Brien J, Di Luca M, Ernst MO. The duration of uncertain times: audiovisual information about intervals is integrated in a statistically optimal fashion. *PLoS One* 9: e89339, 2014. [Erratum. *PLoS One* 9: e96134, 2014.] doi:10.1371/journal.pone.0089339.
- Hollingworth HL. The central tendency of judgment. *J Philos Psychol Sci Methods* 7: 461–469, 1910.
- Jacobs RA. Optimal integration of texture and motion cues to depth. *Vision Res* 39: 3621–3629, 1999. doi:10.1016/S0042-6989(99)00088-7.
- Jazayeri M, Shadlen MN. Temporal context calibrates interval timing. *Nat Neurosci* 13: 1020–1026, 2010. doi:10.1038/nn.2590.
- Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27: 712–719, 2004. doi:10.1016/j.tins.2004.10.007.
- Kolers PA, Brewster JM. Rhythms and responses. *J Exp Psychol Hum Percept Perform* 11: 150–167, 1985. doi:10.1037/0096-1523.11.2.150.
- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS One* 2: e943, 2007. doi:10.1371/journal.pone.0000943.
- Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. *Nature* 427: 244–247, 2004. doi:10.1038/nature02169.
- Lages M, Treisman M. Spatial frequency discrimination: visual long-term memory or criterion setting? *Vision Res* 38: 557–572, 1998. doi:10.1016/S0042-6989(97)88333-2.
- Lapid E, Ulrich R, Rammsayer T. Perceptual learning in auditory temporal discrimination: no evidence for a cross-modal transfer to the visual modality. *Psychon Bull Rev* 16: 382–389, 2009. doi:10.3758/PBR.16.2.382.
- Lejeune H, Wearden J. Vierordt's The Experimental Study of the Time Sense (1868) and its legacy. *Eur J Cogn Psychol* 21: 941–960, 2009. doi:10.1080/09541440802453006.
- Malapani C, Fairhurst S. Scalar timing in animals and humans. *Learn Motiv* 33: 156–176, 2002. doi:10.1006/lmot.2001.1105.
- Meck WH. Selective adjustment of the speed of internal clock and memory processes. *J Exp Psychol Anim Behav Process* 9: 171–201, 1983. doi:10.1037/0097-7403.9.2.171.
- Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7: 3215–3229, 1987.
- Miyazaki M, Nozaki D, Nakajima Y. Testing Bayesian models of human coincidence timing. *J Neurophysiol* 94: 395–399, 2005. doi:10.1152/jn.01168.2004.
- Miyazaki M, Yamamoto S, Uchida S, Kitazawa S. Bayesian calibration of simultaneity in tactile temporal order judgment. *Nat Neurosci* 9: 875–877, 2006. doi:10.1038/nn1712.
- Morein-Zamir S, Soto-Faraco S, Kingstone A. Auditory capture of vision: examining temporal ventriloquism. *Brain Res Cogn Brain Res* 17: 154–163, 2003. doi:10.1016/S0926-6410(03)00089-2.
- Morgan MJ. On the scaling of size judgements by orientational cues. *Vision Res* 32: 1433–1445, 1992. doi:10.1016/0042-6989(92)90200-3.
- Nagarajan SS, Blake DT, Wright BA, Byl N, Merzenich MM. Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *J Neurosci* 18: 1559–1570, 1998.
- Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10: 437–442, 1997. doi:10.1163/156856897X00366.
- Penney TB, Gibbon J, Meck WH. Differential effects of auditory and visual signals on clock speed and temporal memory. *J Exp Psychol Hum Percept Perform* 26: 1770–1787, 2000. doi:10.1037/0096-1523.26.6.1770.
- Petzschnner FH, Glasauer S, Stephan KE. A Bayesian perspective on magnitude estimation. *Trends Cogn Sci* 19: 285–293, 2015. doi:10.1016/j.tics.2015.03.002.
- Pouget A, Beck JM, Ma WJ, Latham PE. Probabilistic brains: knowns and unknowns. *Nat Neurosci* 16: 1170–1178, 2013. doi:10.1038/nn.3495.
- Repp BH, Penel A. Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J Exp Psychol Hum Percept Perform* 28: 1085–1099, 2002. doi:10.1037/0096-1523.28.5.1085.

- Shi Z, Church RM, Meck WH.** Bayesian optimization of time perception. *Trends Cogn Sci* 17: 556–564, 2013. doi:[10.1016/j.tics.2013.09.009](https://doi.org/10.1016/j.tics.2013.09.009).
- Spence C, Squire S.** Multisensory integration: maintaining the perception of synchrony. *Curr Biol* 13: R519–R521, 2003. doi:[10.1016/S0960-9822\(03\)00445-7](https://doi.org/10.1016/S0960-9822(03)00445-7).
- Treisman M.** Temporal discrimination and the indifference interval. Implications for a model of the “internal clock”. *Psychol Monogr* 77: 1–31, 1963. doi:[10.1037/h0093864](https://doi.org/10.1037/h0093864).
- Vilares I, Howard JD, Fernandes HL, Gottfried JA, Kording KP.** Differential representations of prior and likelihood uncertainty in the human brain. *Curr Biol* 22: 1641–1648, 2012. doi:[10.1016/j.cub.2012.07.010](https://doi.org/10.1016/j.cub.2012.07.010).
- Wearden JH, Edwards H, Fakhri M, Percival A.** Why “sounds are judged longer than lights”: application of a model of the internal clock in humans. *Q J Exp Psychol B* 51: 97–120, 1998. doi:[10.1080/713932672](https://doi.org/10.1080/713932672).
- Wei K, Körding K.** Relevance of error: what drives motor adaptation? *J Neurophysiol* 101: 655–664, 2009. doi:[10.1152/jn.90545.2008](https://doi.org/10.1152/jn.90545.2008).
- Welch RB, DuttonHurt LD, Warren DH.** Contributions of audition and vision to temporal rate perception. *Percept Psychophys* 39: 294–300, 1986. doi:[10.3758/BF03204939](https://doi.org/10.3758/BF03204939).
- Westheimer G.** Discrimination of short time intervals by the human observer. *Exp Brain Res* 129: 121–126, 1999. doi:[10.1007/s002210050942](https://doi.org/10.1007/s002210050942).
- Zhang H, Chen L, Zhou X.** Adaptation to visual or auditory time intervals modulates the perception of visual apparent motion. *Front Integr Neurosci* 6: 100, 2012. doi:[10.3389/fnint.2012.00100](https://doi.org/10.3389/fnint.2012.00100).

