The history of multisensory discrepancies affects the ventriloquism effect and the immediate aftereffect differentially

Christoph Kayser [a], Hame Park [b], Herbert Heuer [a,c]

a. Department of Cognitive Neuroscience, Universität Bielefeld, Bielefeld, Germany

b. Department of Neurophysiology & Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20251 Hamburg, Germany

c. Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany

Contact: Christoph Kayser, Universitätsstrasse 25, 33615 Bielefeld, Germany, christoph.kayser@uni-bielefeld.de

Acknowledgements

This study was funded by the Deutsche Forschungsgemeinschaft (DFG KA2661/2-1). The authors declare that they have no conflict of interests. We would like to thank several undergraduate students and research assistants for their help during data collection.

Data and code

The full data analysed here and the respective code are available here: https://github.com/Hame-p/recal_history

Abstract

Multisensory integration and recalibration are two processes by which perception deals with discrepant multisensory signals. In the lab, integration can be probed via the presentation of simultaneous spatially discrepant audio-visual stimuli, i.e. in a spatial ventriloquism paradigm. Recalibration here manifests as an aftereffect bias in unisensory judgements following immediate or long-term exposure to discrepant audio-visual stimuli. Despite many studies exploring the ventriloquism effect (VE) and the immediate aftereffect (VAE), it remains unclear whether the VAE is a direct consequence of the integration of preceding discrepant multisensory signals. We here provide evidence that these two biases are not strictly related. First, we analysed data from ten experiments probing the dependence of each bias on audio-visual discrepancies experienced in multiple preceding trials. This revealed a seven-fold stronger dependence of the VAE on preceding discrepancies compared to the VE. Second, we analysed data from an experiment deviating from the typical trial context in which these biases are probed and found that the VAE can vanish despite the VE being present. We argue that integration may help maintaining a stable percept by reducing immediate sensory discrepancies, whereas recalibration may help maintaining an accurate percept by accounting for consistent discrepancies. Hence, the immediate VAE is not a direct and necessary consequence of the integration of discrepant signals and these two well-studied multisensory response biases can be experimentally dissociated.

Introduction

Our brain often combines multisensory signals received simultaneously to guide behaviour. However, discrepant multisensory signals can also exert lasting influences, even on subsequent unisensory judgments. Well-known examples are the (spatial) ventriloquism effect and the immediate aftereffect (also known as recalibration bias). Here, the discrepant positions of a visual stimulus (the puppet in the case of the ventriloquism act) and an acoustic stimulus (the ventriloquist's voice) are combined when localizing the sound source, and both cues can influence the localization of subsequent unisensory acoustic stimuli, as shown frequently in laboratory experiments (Recanzone, 2009; Wozny and Shams, 2011; Bruns and Roder, 2015; Bosen et al., 2017; Bosen et al., 2018; Park and Kayser, 2019; Badde et al., 2020). Importantly, this aftereffect emerges also in the absence of explicit feedback about task performance, suggesting that it reflects a mechanism by which perception strives for intrinsic consistency between the different senses. While this trial-by-trial recalibration of unisensory perception has been studied frequently, it remains debated whether the immediate aftereffect is a systematic and direct consequence of the directly preceding integration of discrepant multisensory signals (Bruns, 2019; Rohlf et al., 2020; Noppeney, 2021).

In fact, previous work has provided diverging results on the relation between the spatial ventriloquism effect (termed VE in the following) and the immediate aftereffect (termed VAE). One line of studies advocates in favour of a direct causal relation between these. For example, recalibration could be driven by the discrepancy between the integrated multisensory signals and the subsequent unisensory stimulus (Bruns and Roder, 2015; Park and Kayser, 2019; Rohlf et al., 2020; Noppeney, 2021). Support for this view comes from neuroimaging studies demonstrating a spatio-temporal overlap of the neurophysiological processes underlying integration and immediate recalibration (Park and Kayser, 2019, 2021). In addition, behavioural studies have shown that the recalibration bias covaries with the strength of integration in the preceding trial and is reduced when multisensory stimuli are judged as being unrelated (Wozny and Shams, 2011; Park and Kayser, 2020). Furthermore, both are affected by manipulations of stimulus reliability (Rohlf et al., 2021) and when attention is divided across sensory modalities (Badde et al., 2020). Finally, the recalibration bias scales more with the experienced multisensory discrepancy than with the absolute location of the bias-inducing stimulus (e.g. the ventriloquist's puppet) (Wozny and Shams, 2011; Park and Kayser, 2019). If this view was correct, all experimental manipulations affecting integration

should also affect recalibration, and both should for example depend in a similar manner on the history of the recent multisensory experience, such as discrepancies between acoustic and visual signals.

However, an alternative view stipulates that recalibration is independent of whether the preceding multisensory cues are deemed causally related or combined for behaviour. For example, integration could ensure the consistency of sensory estimates in the face of random discrepancies between sensory signals, whereas recalibration could be in the service of accuracy, adjusting sensory estimates based on potential evidence for a modality-specific bias (Di Luca et al., 2009; Block and Bastian, 2011; Burr et al., 2011; Zaidel et al., 2013; Noppeney, 2021). Support comes from studies showing that recalibration, unlike integration, can be independent of stimulus reliability (Zaidel et al., 2011), that immediate recalibration is not tied to specific features of the inducing multisensory stimuli (Bruns and Roder, 2015; Bosen et al., 2018), and that during development integration develops prior to recalibration (Rohlf et al., 2020). If this view was correct, integration and recalibration could be differentially affected by the recent multisensory stimulus history.

In the present study we provide two lines of evidence suggesting that the spatial ventriloquism effect and the immediate aftereffect are not directly (and hence presumably causally) related. In line with many studies on spatial ventriloquism effects we focus on the aftereffect that emerges intrinsically and in the absence of explicit feedback about task performance, noting that the introduction of explicit feedback may engage additional mechanisms of recalibration. First, we analysed multiple datasets compiled across a number of our previous studies to show that the audio-visual VE and VAE depend to different degrees on the short-term history of experienced multisensory discrepancies. Second, based on data obtained following an experimental manipulation affecting the typical trial context in which the VE and VAE are probed, we show that the VAE can vanish despite the VE being present. These data speak against the immediate VAE being a direct consequence of the integration of discrepant multisensory signals in the preceding trial and suggest that these two well-studied multisensory response biases can be experimentally dissociated.

Methods

Experimental design and procedures

Most datasets analysed here have been published previously, though there are three exceptions (see Table 1 for details). Overall the data are from 11 experiments, all of which were designed to probe the ventriloquism effect and the immediate ventriloquism aftereffect following an established design (Wozny and Shams, 2011) that we have used in a number of previous studies (Park and Kayser, 2019, 2020, 2021; Park et al., 2021; Park and Kayser, 2022). Participants were adult volunteers who reported no history of neurological diseases, normal vision and hearing. They were compensated for their time and provided written or electronic informed consent prior to participation. Prior to the actual experiments we performed a screening task probing participants' spatial hearing (Park and Kayser, 2022). The studies were approved by the local ethics committee of Bielefeld University and data were collected anonymously. This makes it impossible to determine whether some participants participated in more than one of the experiments reported here.

In all experiments, the auditory and visual stimuli were presented either simultaneously in multisensory trials (AV) or individually in unisensory trials (A or V; c.f. Figure 1A). The experiments were performed in a dark and anechoic chamber with sounds being presented at 64 dB r.m.s. through one of 5 speakers (Monacor MKS-26/SW, MONACOR International GmbH & Co. KG, Bremen, Germany) located at 5 horizontal locations. The precise spatial spacing of stimulus locations differed between experiments (Table 1). The visual stimuli were projected (Acer Predator Z650, Acer Inc., New Taipei City, Taiwan) onto an acoustically transparent screen placed in front of the speakers (Screen International Modigliani, 2x1 m^2) and they were centred on the same five locations as the acoustic stimuli. For most experiments (except experiment 11, see below), the positions of auditory and visual stimuli were drawn pseudo randomly for individual trials in order to sample a large range of multisensory discrepancies (DVA; the spatial separation of the auditory and visual stimuli on the AV trials) and stimulus positions. The participants were asked to localize the sound (on AV and A trials) or to localize the visual stimulus (on V trials), and they responded by moving a mouse cursor to the respective location and clicking to submit their response. For more details about the procedures see the previous publications. The AV trials served to probe the VE and to induce the aftereffect, the A trials served to probe the VAE and the less frequent V trials mainly served to ensure that participants maintained divided attention across both senses.

Stimuli experiments 1-9: The visual stimulus was a cloud of white dots distributed according to a two-dimensional Gaussian distribution (200 dots, SD of vertical and horizontal spread: 1.6° (Experiments 2-5,7,8) or 2° (Experiments 1,6,9), width of a single dot: 0.12°, duration of stimulus presentation: 50 ms), and the auditory stimulus was a 1300 Hz sine wave tone (50 ms duration). Experiment 9 was originally designed to probe the impact of trial-context and presented the AV, A and V trials in three sub-blocks featuring different proportions of AV trials with a visual stimulus being displaced from the sound towards the center of the visual screen. For the present analysis, however, we combined all trials and ignored this specific manipulation.

Stimuli experiment 10: For this experiment we used naturalistic images and sounds as stimuli. More precisely, the visual stimuli were either the faces of one of five animals (cat, dog, eagle, sheep, tiger) or phase-scrambled versions of these images. The images were faded into the grey background using a Gaussian mask with a horizontal SD of 2° and were equalized for luminance and contrast. The acoustic stimuli where the matching five vocalizations (with 200 ms duration, all r.m.s normalized) or scrambled versions of these. These scrambled sounds were created by filtering the original sounds into 10 equi-spaced band-limited components between 100Hz and 8kHz and randomly interchanging these bands across sounds. AV trials either presented matching naturalistic auditory and visual pairs, scrambled pairs, or pairs featuring one naturalistic and one scrambled stimulus (again these specific manipulations are neglected for the present analyses).

Stimuli experiment 11: This experiment introduced a manipulation of the typical trial sequences intended to break the continuous flow of AV-> A, AV->A trials (interspersed with less frequent V trials) that forms the core of experiments 1-10. Here, trials were presented as triplets: these either consisted of a multisensory 'reset' trial, an AV trial, and a second AV trial to probe the cumulative nature of the VE; or they consisted of a reset trial, an AV trial, and an A trial to probe the VAE (again some of these sequences were followed by V trials; see Figure 3A). Reset trials did not require a response, featured audio-visual stimuli of zero discrepancy at one of the five locations (spacing 11°) and presented visual stimuli with larger spatial uncertainty (dots with a horizontal spread of 5°). The visual stimuli in the AV trials had spread of 2°, sounds were as in experiments 1-9. The experiment was originally designed to probe the influences of the direction of the multisensory discrepancy and the relative locations of the sounds in the AV and A trials. For this reason, the DVA on AV trials was either +11° or -11° and the sound locations in subsequent AV-A trials differed by exactly one position (hence 11°). As a result, the VE and VAE biases in this paradigm can only be probed based on two values of DVA. We used this dataset to probe the history dependence

of the VE (in reset-AV-AV sequences), the VAE (in reset-AV-A sequences) and as sanity check also the VE immediately prior to the VAE (again in reset-AV-A sequences).

Experiment	Publication & DOI	Note	N	Stimulus spacing
1	Park & Kayser. The Journal of Neuroscience 2021, doi: 10.1523/JNEUROSCI.2091-20.2020	Short-term recalibration only	19	11.6°
2	Park, Kayser. Scientific Reports 2020, doi: 10.1038/s41598-020-77730-7	Delay manipulation 1	20	8°
3	Park & Kayser C. Scientific Reports 2020, doi: 10.1038/s41598-020-77730-7.	Delay manipulation 2	21	8°
4	Park & Kayser C. Scientific Reports 2020 , doi: 10.1038/s41598-020-77730-7.	Masking condition	22	8°
5	Park & Kayser C. Scientific Reports 2020 , doi: 10.1038/s41598-020-77730-7.	No-mask condition	22	8°
6	Park, Nannt & Kayser. Cortex 2021, doi: 10.1016/j.cortex.2020.12.001	Young participants only	22	8.5°
7	Park & Kayser C. Cognition 2022, doi: j.cognition.2022.105092	Narrow context	21	6.7°
8	Park & Kayser C. Cognition 2022, doi: j.cognition.2022.105092	Wide context	21	11.6°
9	Previously unpublished	Trial context manipulation	17	11°
10	Previously unpublished	Naturalistic stimuli	20	11°
11	Previously unpublished	Triplets of trials	20	11°

Table 1: Datasets. The datasets analysed here come from 11 different experiments that capitalize on the same overall design to probe the VE and VAE. Experiments 1-8 have been published as part of previous studies (as explained in column 'Note'), experiments 9-11 are analysed here for the first time. Experiments 9 and 10 contained additional manipulations not analysed here. N indicates the number of participants, some of which may have participated in more than one experiment. The last column indicates the spatial separation of the five stimulus locations. The full set of data is available (see 'Data and code' statement).

Analysis of response biases

We defined the integration and recalibration biases (VE, VAE) based on typical procedures used in the literature (Wozny and Shams, 2011; Rohe and Noppeney, 2015b; Park et al., 2021). The VE in AV trials was quantified as the difference between the reported sound location and the location at which the sound was presented. The VAE in the A trials was computed as the difference between the reported sound location and the averaged reported location for all A trials with the same stimulus position. Typically, these biases are then investigated in relation to the multisensory discrepancy (DVA) in the respective AV

trial in which the VE is quantified, or the AV trial immediately prior to the A trial in which the VAE is obtained. We here extended this analysis and probed each bias in relation to the DVA in up to three successive AV trials (except for experiment 11, where the flow of trials was interrupted by reset trials). Note that based on the typical design (Figure 1A) these three successive AV trials were each separated by unisensory A (and sometimes also V) trials.

Both biases typically scale with the magnitude of the discrepancy, DVA, with the VE being stronger than the VAE (see Figure 1B for experiment-wise averages). Previous work has shown that they can tamper off at large discrepancies, reflecting the reduced multisensory combination of spatially very discrepant acoustic and visual stimuli. Such a non-linear dependency between bias and DVA is predicted by models of multisensory causal inference, which posit that in typical ventriloquism settings the two multisensory cues are only combined when they likely arise from a common source in the environment (Kording et al., 2007; Odegaard et al., 2015; Rohe and Noppeney, 2015a; Noppeney, 2021). While such a non-linear dependency can reveal interesting aspects of multisensory perception, its relative prevalence varies across experiments, as is evident from Figure 1B. For a consistent analysis across all experiments, and to allow a comparison with previous work focusing only on linear dependencies, we here capitalized on the slope of the linear relation of response bias to DVA. Hence we effectively quantified the proportional response bias per degree of discrepancy.

The dependency of each bias, VE and VAE, on the DVA in up to three AV trials was quantified in three ways. First, we determined for each participant the proportional bias contingent on a specific history of DVAs in preceding trials (Figure 2A). For this analysis we focused on sequences of three AV-A trial pairs (denoted by a,b,c in the following) that all featured a non-zero DVA. We then grouped these according to the sign of DVAs in the respective AV trials and calculated the proportional biases for a number of constellations: for all trial pairs ('c' in Figure 2), for sequences of two AV-A trial pairs with same sign of DVA ('b=c'; hence regardless of the sign of DVA on pair a), for triplets of such trial pairs ('a=b=c'), and for sequences in which the sign changed ('a=b~=c'). For each constellation and participant, the trial-averaged proportional bias was defined as the ratio of VE (VAE) obtained on the AV trial in c and the DVA in the AV trial in pair c. Second, we used partial regression to derive for each participant the dependency of each bias on the three DVAs across all trial pairs, also including those with a zero DVA (Figure 2B). For this analysis we report the partial betas for each participant. Third, we contrasted the predictive power of linear models predicting the response biases from the DVA on the immediate AV trial (c) with models that in addition included one (b) or two preceding AV trials (a and b).

Statistical analysis

Statistical comparisons are based on the 205 datasets analysed, which we treated as independent sample, although it is possible that some participants completed more than one experiment (see above). To describe condition-wise effects across the sample of experiments and participants we relied on population medians. For a statistical assessment of differences between conditions we relied on (percentile-) bootstrap-based confidence intervals obtained using 8000 randomizations each. We emphasize confidence intervals as previous work did not allow specifying an unambiguous null hypothesis about the presence or absence of specific effects. Rather, our goal was to quantify and compare the degree of history dependence of the VE and VAE, and to establish potential differences between these (Fethney, 2010; Greenland et al., 2016). As a measure of effect size of the history dependence we compared the magnitude of these between different conditions in units of percent difference. To compare the predictive power of different linear models we derived the Akaike information criterion (AIC) for individual participant's data and summed the respective AIC values across all 205 datasets for group-level analysis. We then relied on the Akaike weights for a model comparison, which reflect the conditional probability of each model given the data and are ideally suited to compare nested models (Wagenmakers and Farrell, 2004).

Results

History dependence of VE and VAE

We analysed the ventriloquism effect (VE) and the immediate ventriloquism aftereffect (VAE) derived from 205 datasets that were collected across ten variations of the same overall experimental design (see Table 1). This design consisted of a sequence of alternating audio-visual (AV) and auditory (A) trials, interspersed with few visual test trials to maintain attention divided across modalities (Figure 1A). Here, the AV trials allow quantifying the VE as a function of the multisensory discrepancy (DVA) of the respective auditory and visual stimuli, while the A trials allow quantifying the VAE as a function of the DVA in the immediately preceding AV trial. Figure 1B shows the experiment-averaged biases as a function of the multisensory discrepancy (DVA). While this dependency can feature nonlinearities, we here focused on the linear relation of response bias and DVA for sake of consistency across experiments and with previous reports.

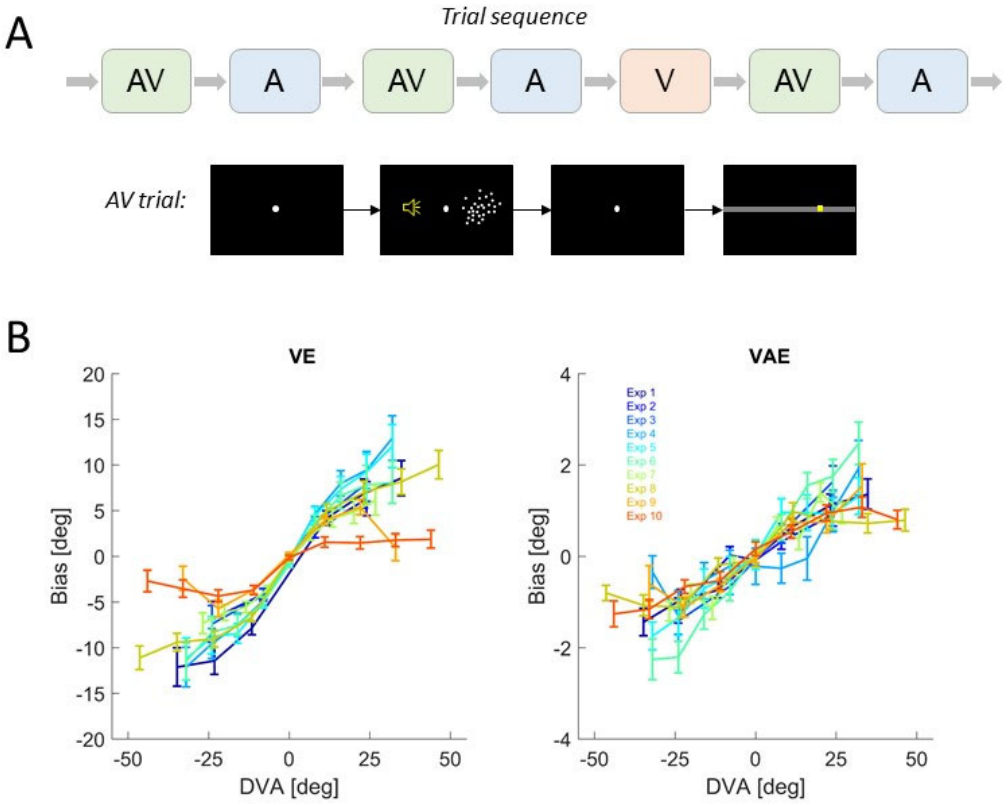

Figure 1: Experimental design and average response biases. Panel A shows the typical sequence of AV and A trials designed to probe the two responses biases (VE, VAE) and the less frequent V trials there to maintain a divided focus on both modalities. The lower panel in A shows a typical sequence of events within an AV trial, including a fixation period, stimulus presentation (here random dots), a post-stimulus period and the response bar on which participants moved a mouse cursor to indicate the perceived location of auditory or visual stimuli. The yellow speaker icon indicates the location of a sound, but was not displayed in the experiment. **Panel B** shows the average (mean and s.e.m.) biases for each of ten experiments (1-10 in Table 1) as a function of the audio-visual discrepancy presented in the AV trial probing the VE or the AV trial immediately prior to the A trial probing the VAE.

Our main question was whether each bias depends only on the multisensory discrepancy (DVA) experienced in the current AV trial (or for the VAE the immediately preceding AV trial), or whether these biases also depend on the DVAs experienced in preceding AV trials. In a first analysis addressing this question we quantified the proportional VE and VAE biases for sequences of trials that were selected contingent on a specific constellation of DVA's (Figure 2A). Specifically, we compare the biases for sequences of AV-A trials featuring the same sign of multisensory discrepancy ('b=c', 'a=b=c'), or for sequences of trials featuring a change in sign ('a=b~=c').

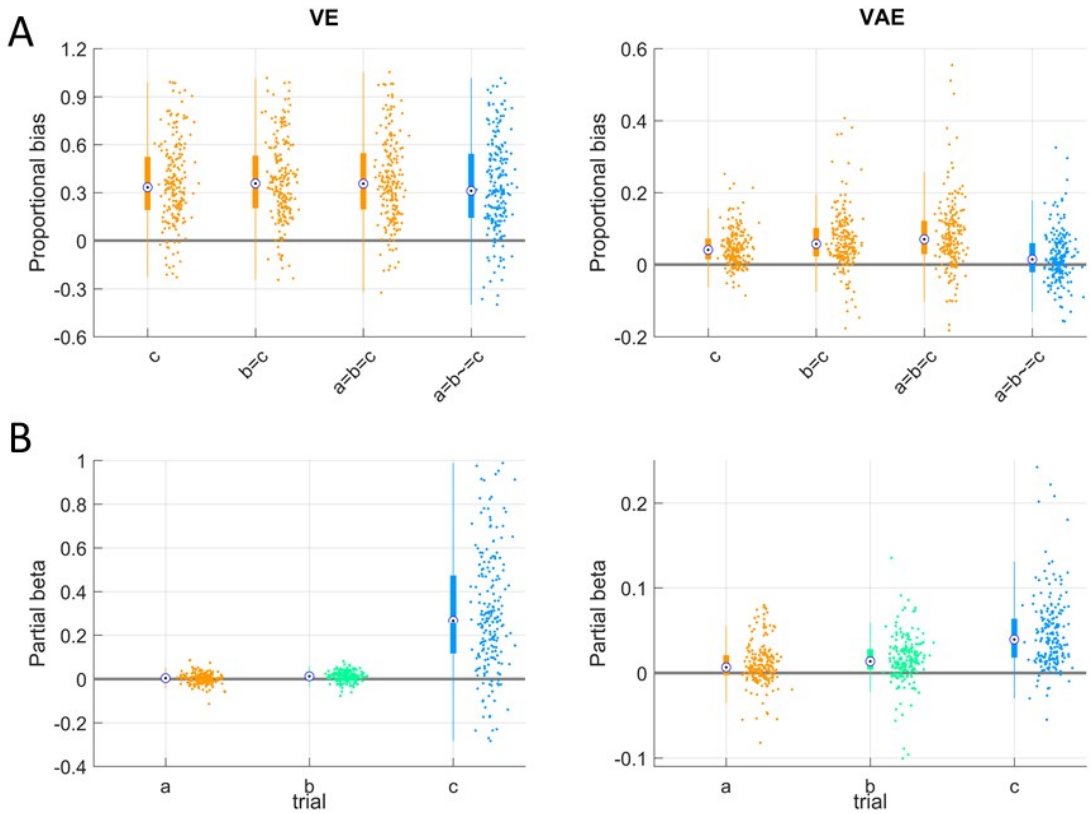

Figure 2: History-dependence is prominent for the VAE but not for the VE. Panel A displays the proportional biases contingent on the history of the sign of the multisensory discrepancy (DVA) experienced during preceding AV trials. Here we focus on sequences of (AV-A) trial pairs featuring a non-zero DVA only and quantified the biases for individual trial pairs (condition c), sequences of two pairs with same sign of DVA in the respective AV trials ('b=c'; regardless of A), triplets featuring the same sign ('a=b=c'), as well as triplets featuring two AV trials with same sign followed by one with an opposing sign of DVA ('a=b~=c'). The criterion of an equal sign of DVA selects trials on which the visual stimulus is always on the same side of the auditory stimulus, regardless of their absolute locations. The bias is derived as the response error relative to the DVA in trial pair c. **Panel B** displays the results of a partial regression analysis quantifying the biases contingent on the DVA in sequences of three trial pairs (a,b,c), including all trials of each experiment. In all panels dots reflect individual datasets (i.e. participants, n=205), while circles and thick bars indicate the median and 25th to 75th percentiles.

This analysis reveals a cumulative influence of DVA on the VAE, whereas such history dependence is at best weak for the VE. Compared to the bias obtained across all AV-A trial pairs (condition 'c'), triplets of trial pairs with the same sign of DVA produce a very comparable VE bias (medians of 0.333 vs. 0.356, for conditions 'c' and 'a=b=c', respectively, an increase by only 6.9%; see Table 2 for medians and confidence intervals for a statistical assessment). In contrast, the VAE increases by 70% (medians 0.041 vs. 0.07). Similarly, when two AV trials featuring the same sign of discrepancy are followed by an AV trial with opposing sign ('a=b~=c') the VE bias differs very little (medians 0.356 vs. 0.311, reduction of 12%), while the VAE is reduced by 80% (medians 0.07 vs. 0.014).

We then extended this explorative analysis to all trials of each experiment (also those with zero DVA) and implemented a partial regression analysis in which we predicted each bias, VE and VAE, by the DVA's sequences of three AV-A trial pair (Figure 2B). The respective partial betas show that each bias is driven the most by the just-experienced discrepancy (trial pair c). The influence of both preceding AV trials was significant for the VAE and only partly so for the VE (c.f. Table 3 for medians and a statistical assessment using confidence intervals). As compared to the influence of the most recent DVA (c), the influence of the preceding DVAs (a, b) was strongly reduced for the VE, but considerably less so for the VAE. More precisely, the partial betas of the preceding AV trial (b) amounted to about a third of the beta for the most recent AV trial (c) for the VAE (medians 0.039 vs 0.014, ratio of medians 0.346 with 99% bootstrap CI [0.279, 0.463]), but for the VE the ratio was about one-twentieth only (medians 0.266 vs 0.013, ratio of medians 0.048 [0.028, 0.068]). To probe whether this indeed reflects a significant difference in history dependence between VE and VAE, we directly compared these 'growth-ratios' (partial beta for trial c divided by partial beta for trial b) between biases. This revealed about a seven-fold difference between the biases: the ratio (VAE divided by VE) of the respective median growth-ratios was significantly different from one (median 7.2, 99% bootstrap CI [4.81, 13.66]), demonstrating that the VAE scales in proportion

significantly more over cumulative trials than the VE (the respective p-value for a difference from a ratio of 1 was $p<10^{-3}$).

In a final analysis we compared the predictive power of linear models predicting each bias based on either only the DVA in the immediate test trials (c) or based on the DVA also in one or two preceding pairs of AV-A trial pairs (b,c; a,b,c). For the VE, a model solely relying on the immediate AV trial (c) was most-likely given the data (AIC weight of nearly 1; relative AICs for the three models depending on trial c only, trials b and c, trials a, b and c: 0, 87, 252). In contrast, for the VAE a model including the multisensory discrepancy also in the immediately preceding AV trial was most likely (AIC weight of nearly one, relative AICs: 156, 0, 97).

Given that for some individual datasets the VE and/or the VAE may deviate from a linear dependency on DVA, we implemented control analyses in which we focused only on the smallest magnitude of the DVA for each experiment. Within this restricted range (DVA's being less than 12°), the VE and VAE are quite well approximated by a linear dependency on DVA. We repeated the above analyses of the proportional biases (Table 2) and the partial regression (Table 3) and obtained basically the same results compared to when including all DVAs. Collectively these findings suggest that the VAE depends on the history of multisensory discrepancies, while the VE does so to a much smaller degree, if at all.

Bias / Trial sequence	c	b=c	a=b=c	a=b~=c
VE	0.333 [0.295, 0.376]	0.357 [0.296, 0.395]	0.356 [0.314, 0.400]	0.311 [0.246, 0.400]
VAE	0.041 [0.029, 0.050]	0.057 [0.044, 0.074]	0.070 [0.054, 0.086]	0.014 [0.004, 0.026]
VE (Control)	0.470 [0.416, 0.531]	0.467 [0.419, 0.547]	0.477 [0.430, 0.542]	0.457 [0.363, 0.560]
VAE (Control)	0.056 [0.034, 0.069]	0.076 [0.048, 0.097]	0.085 [0.046, 0.125]	0.004 [-0.018, 0.027]

Table 2: Proportional biases when quantified for individual trials or contingent on the history of the sign of the multisensory discrepancy in preceding trials. See Figure 2A for a graphical illustration of these results. Here we focus on sequences of (AV-A) trials featuring a non-zero DVA only, and quantified the biases for individual pairs (condition 'c'), sequences of two pairs with same sign of DVA ('b=c'), triplets of such pairs ('a=b=c'), as well as two pairs followed by one of an opposing sign of DVA ('a=b~=c'). The bias is derived as the response error relative to the DVA in the AV trial in c. The table lists the median values and 99% bootstrap confidence intervals across all 205 datasets. The upper two rows are based on trials including the full range of DVAs per experiment, the lower two rows ('Control') are based only on sequences of trials for which the DVA on trial c was of the smallest magnitude for each experiment. This control analysis emphasizes the linear dependency of each bias on DVA. Statistically interesting results pertain

to the comparison of ‘a=b=c’ vs. ‘a=b~=c’: for these conditions the respective VE biases are each closely located within the 99% confidence interval of the other condition, while for the VAE they are not and hence statistically distinct (at $p<0.01$).

Partial beta	Trial a	Trial b	Trial c
VE	0.004 [-0.001, 0.008]	0.013 [0.008, 0.017]	0.266 [0.216, 0.307]
VAE	0.007 [0.004, 0.011]	0.014 [0.011, 0.018]	0.039 [0.033, 0.047]
VE (Control)	0.006 [0.001, 0.011]	0.009 [0.005, 0.014]	0.271 [0.225, 0.325]
VAE (Control)	0.006 [0.002, 0.014]	0.012 [0.007, 0.019]	0.038 [0.029, 0.049]

Table 3: Partial betas for each bias against the discrepancy in sequences of three trials. See Figure 2B for a graphical illustration. Partial betas were computed across all instances of sequences of three trials in each dataset. Shown are the medians and 99% bootstrap confidence intervals across all 205 datasets. The upper two rows are based on trials including the full range of DVAs per experiment, the lower two rows (‘Control’) are based only on sequences of trials for which the DVA on trial c was of the smallest magnitude for each experiment. Statistically interesting results pertain to the insignificant beta for VE in trial a (as based on the confidence interval) and the significant differences between trial c vs. a or b for both biases.

The cumulative increase of the VE can be explained by a superposition of two response biases

The above data show that numerically both response biases increase following a series of trials with a multisensory discrepancy pointing in the same direction, though the increase was proportionally much stronger and reliable for the VAE. Although the VAE is measured in unisensory trials, it should also manifest during multisensory trials. Under the assumption that recalibration serves to ensure an accurate percept given dynamical discrepancies between sensory signals, it should be present independent of whether the recalibrated signal is accompanied by other signals or not. However, in multisensory trials the VAE would usually be shadowed by the much stronger integration bias. We implemented an analysis to directly probe whether the modest cumulative increase of the VE seen above could be explained by an underlying increase of a recalibration bias that also manifests during participant’s judgements in the AV trials.

Specifically, we reasoned that the VE bias following three AV-A trial pairs featuring the same sign of DVA (i.e. ‘a=b=c’ in Figure 2) should comprise the integration bias arising in trial c plus the superimposed VAE bias arising from two trial pairs of the same DVA sign (b and c in Figure 2) as they precede each third one. Similarly, the VE following two trial pairs (‘a=b’) should comprise the VAE bias arising from one preceding

trial pair. With this in mind, we compared the cumulative increase in VE from pairs to triplets of AV trials (i.e. 'a=b=c' minus 'b=c') with the expected increase after discounting the respective VAE biases. The median observed for cumulative growth itself (i.e. the median of the condition-difference) was positive but not significant (VE 'a=b=c' minus 'b=c': median and 99%-bootstrap CI 0.010 [-0.001, 0.017]), while the corrected growth was negative (-0.015 [-0.024, -0.001]). This shows that the cumulative increase seen in the VE may even be fully attributed to a superposition of integration and recalibration biases, adding to the argument that the integration response bias is essentially immediate.

Context sensitivity of VE and VAE

The above results show a clear cumulative influence of preceding multisensory stimuli on the VAE but hardly, if at all, on the VE. This speaks against the hypothesis that the VE and VAE are directly related, with the VAE being an immediate and direct consequence of the VE. We observed further evidence against such a relation in a second set of data that was obtained in a slightly different variant of the typical experimental paradigm (experiment 11; n=20 participants). That experiment was originally designed to probe the mutual influence of two design features on the VAE: the multisensory discrepancy in the preceding AV trial and the relative position of the auditory stimuli across the AV and A trials. This experiment featured specific sequences of trials defined by their sound locations, their DVAs, and by interruptions of the typical AV-A pairs by multisensory 'reset' trials (Figure 3A, and see Methods). Thus, the VE and the VAE were probed in a different context of other trials than in experiments 1-10.

The data obtained in this experiment again provide little or no evidence for a history dependence of the VE (Figure 3B). While the partial beta for the current AV trial was significant (trial 'b': median and 99% bootstrap CI 0.48 [0.40, 0.65]), the partial beta for the preceding AV trial was not significant (trial 'a': 0.05 [0, 0.16]). Importantly, in this experiment the VAE was not significant – numerically it was even in the counter-adaptive direction (median beta: -0.02 [-0.07, 0.02]; Figure 3B), despite there being a clear VE also in the AV trials immediately prior to the A trials (median VE: 0.49 [0.41, 0.64]). This data demonstrates that the emergence of a VE in an AV trial is not a sufficient condition for the VAE to emerge in an immediately following A trial.

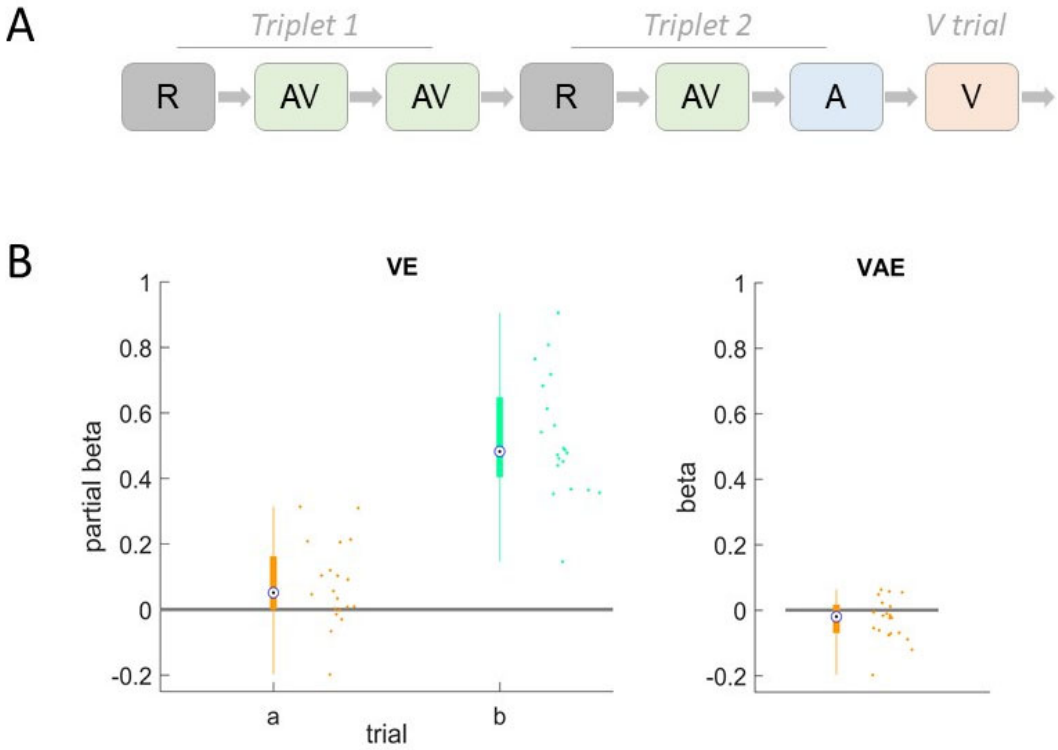

Figure 3: Results from an experiment with different trial context. Panel A displays the two types of trial sequences in this experiment. These were either designed to probe the (two-trial) history dependence of the VE (Triplet 1) or to probe the VAE (Triplet 2). These triplets alternated, started each with a multisensory reset trial (R) not requiring a response and were interrupted by infrequent V trials. **Panel B** shows the VE bias for sequences of two immediately subsequent AV trials obtained using the partial regression analysis (trials a, b; probed in triplets of type 1). Right panel shows the VAE (probed for Triplets of type 2). N=20 participants.

Discussion

The collective evidence obtained here shows that the immediate aftereffect in the spatial ventriloquism paradigm (i.e. the VAE) is not an obligatory and direct consequence of the integration of discrepant multisensory signals (the VE) and both scale differentially with the history of experienced multisensory discrepancies. As we discuss in the following, this speaks against the integration of discrepant multisensory signals as direct driver of trial-wise recalibration and fuels the debate about their mechanistic and neural relation.

Adaptive perception in an unstable environment

Multisensory integration and recalibration are two phenomena by which perception deals with discrepant signals in our environment. Such discrepancies can be dynamic and may feature predictable and unpredictable components (Noppeney, 2021). Some studies have speculated that multisensory integration helps maintaining a stable percept by reducing immediate discrepancies between seemingly redundant signals, while recalibration protects perception against consistent discrepancies and helps to ensure the accuracy of individual sensory estimates despite systematic discrepancies between sensory signals (Ernst and Bulthoff, 2004; Zaidel et al., 2013; Bruns, 2019; Noppeney, 2021). Importantly, especially given the cumulative nature of recalibration, both response biases may effectively superimpose during individual trials.

Our data support the notion that sensory integration is a process that reduces discrepant sensory estimates when these supposedly originate from a common source based on the immediately observed discrepancy (Shams and Beierholm, 2010; Odegaard et al., 2015; Rohe and Noppeney, 2015b). Importantly, in the large data set analysed here, the VE was mostly strongly driven by the audio-visual discrepancy experienced in the current trial, while the discrepancies experienced in the preceding multisensory trials had only a much weaker influence. A control analysis suggests that this weak influence may be even accounted for by a superposition of a cumulative recalibration bias that also exerts an influence on judgments during multisensory trials. This finding corroborates that integration adjusts perceptual estimates to remove immediate sensory discrepancies. At the same time, it does not rule out that individual components feeding into integration depend also on the more extensive history of sensory signals. For example, the integration bias scales with the relative sensory reliability of the probed modality, in line with the notion of Bayesian optimal reliability weighted cue combination (Ernst and

Bulthoff, 2004; Angelaki et al., 2009). A recent study has shown that the brain keeps track of the individual unisensory reliabilities and discounts previous experience in an exponential manner (Beierholm et al., 2020). Hence, our estimates of unisensory reliabilities are shaped by the collective past, which renders the weights assigned to individual sensory inputs during integration history dependent. In contrast to this, our data show that the dependency of the integration bias on the experienced multisensory discrepancies declines super-exponentially and is essentially immediate. Collectively, these findings suggest that integration per se is an adaptive process that takes the recent sensory experience into account, though the influence of the discrepancy between the individual sensory estimates seems rather instantaneous.

Multisensory recalibration has many facets

Multisensory recalibration has been shown to emerge at multiple time scales, and is often studied on the trial-by-trial level, as done here, or following minutes of discrepant multisensory exposure (Radeau and Bertelson, 1974; Recanzone, 1998; Bruns and Roder, 2015; Bosen et al., 2018; Bruns, 2019; Bruns and Röder, 2019). Our data are compatible with the idea that the immediate aftereffect is a precursor of a long-term effect, given that the VAE bias grows considerably across several trials (Park and Kayser, 2021). However, the collective evidence suggests that the transition from immediate to long-term aftereffect involves a number of functional and physiological changes. For example, compared to immediate recalibration, long-term recalibration exhibits higher specificity to the adapted sensory features, such as tone frequencies or the location within the visual field (Bruns and Roder, 2015, 2017; Bosen et al., 2018). This higher specificity comes with a more extensive activation of short-latency neurophysiological processes that seem to arise from early sensory cortices, pointing to partially distinct cerebral origins of the immediate and long-term aftereffects (Recanzone et al., 2000; Bruns et al., 2011; Bruns and Roder, 2015; Zierul et al., 2017; Park and Kayser, 2021). These studies suggest that the immediate aftereffect arises in a top-down manner, possibly controlled by prefrontal cortex, and guides the formation of a long-term effect that then also involves low-level sensory cortices.

Furthermore, recalibration emerges both in the presence and the absence of explicit feedback about the accuracy of individual judgements (Adams et al., 2010; Ernst and Luca, 2011; Zaidel et al., 2011; Zaidel et al., 2013). Importantly, the presence or absence of feedback may also differentiate functionally distinct forms of recalibration. In studies on visual-vestibular perception long-term recalibration following feedback was found to depend on cue reliability similar to sensory integration (Zaidel et al., 2013) while long-term recalibration without feedback was independent of reliability (Zaidel et al., 2011). Again this

points to a functional disconnection of recalibration in general from multisensory integration. Our results support this notion in the sense that also trial-wise recalibration can be functionally disconnected from integration.

Integration does not mandate recalibration

Assuming that the immediate VAE and the VE are not strictly linked, why do they correlate so frequently across participants or experimental manipulations? One view stipulates that recalibration is driven by the discrepancy between the integrated multisensory signals and the subsequent unisensory test stimulus, rendering recalibration directly dependent on the outcome of integration (Zaidel et al., 2013; Bruns, 2019; Noppeney, 2021). This view predicts that if two cues are deemed sufficiently discrepant to unlikely originate from a common source, then the outcome of multisensory causal inference should emphasize the task-relevant unisensory estimate over the integrated one, leaving essentially no multisensory bias to drive recalibration. Hence, an integration bias becomes a prerequisite for recalibration, a view that may also pertain to development where recent data show that integration develops prior to recalibration (Rohlf et al., 2020). An ideal scenario to refute this would be an experimental design abolishing an integration bias such as the VE in the presence of a clear recalibration bias such as the VAE. While we did not obtain such a differentiation, our findings show that the VE does not mandate a significant VAE, that is, immediate recalibration is not a necessary outcome of integration, which also speaks against a direct link.

Still, both processes could be driven by an assessment of the causal relation of multisensory signals, a process that also depends on their respective reliabilities. But additionally, recalibration may emerge only when there is statistical evidence for the potential emergence of a systematic discrepancy between the sensory evidence and the respective judgements, also in the absence of explicit feedback (Zaidel et al., 2011). Manipulations breaking the sequence of multisensory and unisensory trials typically used to probe the VE and VAE may render an assessment of systematic discrepancies between sensory evidence and judgements more difficult and hence reduce the recalibration effect. This is precisely what we found in experiment 11. Still, it remains unclear whether such implicit judgements about causal relations between sensory signals or some other feature of the specific stimulus sequence used in this experiment caused the absence of the VAE. For example, the more stereotyped relative locations of the sounds in the AV and A trials in experiment 11, compared to the more unpredictable design used in the other experiments, may also have affected recalibration. More studies are required to disentangle the influence of multisensory

and unisensory stimulus properties on immediate recalibration and to probe whether recalibration can emerge in the absence of integration.

References

Adams WJ, Kerrigan IS, Graf EW (2010) Efficient visual recalibration from either visual or haptic feedback: the importance of being wrong. J Neurosci 30:14745-14749.

Angelaki DE, Gu Y, DeAngelis GC (2009) Multisensory integration: psychophysics, neurophysiology, and computation. Curr Opin Neurobiol 19:452-458.

Badde S, Navarro KT, Landy MS (2020) Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. Cognition 197:104170.

Beierholm U, Rohe T, Ferrari A, Stegle O, Noppeney U (2020) Using the past to estimate sensory uncertainty. eLife 9.

Block HJ, Bastian AJ (2011) Sensory weighting and realignment: independent compensatory processes. J Neurophysiol 106:59-70.

Bosen AK, Fleming JT, Allen PD, O'Neill WE, Paige GD (2017) Accumulation and decay of visual capture and the ventriloquism aftereffect caused by brief audio-visual disparities. Exp Brain Res 235:585-595.

Bosen AK, Fleming JT, Allen PD, O'Neill WE, Paige GD, Bosen AK, Chatterjee M (2018) Multiple time scales of the ventriloquism aftereffect. PLoS One 13:e0200930.

Bruns P (2019) The Ventriloquist Illusion as a Tool to Study Multisensory Processing: An Update. Front Integr Neurosci 13:51.

Bruns P, Roder B (2015) Sensory recalibration integrates information from the immediate and the cumulative past. Sci Rep 5:12739.

Bruns P, Roder B (2017) Spatial and frequency specificity of the ventriloquism aftereffect revisited. Psychol Res.

Bruns P, Röder B (2019) Repeated but not incremental training enhances cross-modal recalibration. J Exp Psychol Hum Percept Perform 45:435-440.

Bruns P, Liebnau R, Roder B (2011) Cross-modal training induces changes in spatial representations early in the auditory processing pathway. Psychol Sci 22:1120-1126.

Burr D, Binda P, Gori M (2011) 173Multisensory Integration and Calibration in Adults and in Children. In: Sensory Cue Integration (Trommershäuser J, Kording K, Landy MS, eds), p 0: Oxford University Press.

Di Luca M, Machulla TK, Ernst MO (2009) Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. J Vis 9:7.1-16.

Ernst MO, Bulthoff HH (2004) Merging the senses into a robust percept. Trends Cogn Sci 8:162-169.

Ernst MO, Luca MD (2011) 224Multisensory Perception: From Integration to Remapping. In: Sensory Cue Integration (Trommershäuser J, Kording K, Landy MS, eds), p 0: Oxford University Press.

Fethney J (2010) Statistical and clinical significance, and how to use confidence intervals to help interpret both. Australian critical care : official journal of the Confederation of Australian Critical Care Nurses 23:93-97.

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG (2016) Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European journal of epidemiology 31:337-350.

Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS ONE 2:e943.
Noppeney U (2021) Perceptual Inference, Learning, and Attention in a Multisensory World. Annu Rev Neurosci 44:449-473.
Odegaard B, Wozny DR, Shams L (2015) Biases in Visual, Auditory, and Audiovisual Perception of Space. PLoS Comput Biol 11:e1004649.
Park H, Kayser C (2019) Shared neural underpinnings of multisensory integration and trial-by-trial perceptual recalibration in humans. eLife 8.
Park H, Kayser C (2020) Robust spatial ventriloquism effect and trial-by-trial aftereffect under memory interference. Sci Rep 10:20826.
Park H, Kayser C (2021) The Neurophysiological Basis of the Trial-Wise and Cumulative Ventriloquism Aftereffects. J Neurosci 41:1068-1079.
Park H, Kayser C (2022) The context of experienced sensory discrepancies shapes multisensory integration and recalibration differently. Cognition 225:105092.
Park H, Nannt J, Kayser C (2021) Sensory- and memory-related drivers for altered ventriloquism effects and aftereffects in older adults. Cortex 135:298-310.
Radeau M, Bertelson P (1974) The after-effects of ventriloquism. The Quarterly journal of experimental psychology 26:63-71.
Recanzone GH (1998) Rapidly induced auditory plasticity: the ventriloquism aftereffect. Proc Natl Acad Sci U S A 95:869-875.
Recanzone GH (2009) Interactions of auditory and visual stimuli in space and time. Hear Res 258:89-99.
Recanzone GH, Guard DC, Phan ML, Su TK (2000) Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. J Neurophysiol 83:2723-2739.
Rohe T, Noppeney U (2015a) Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. PLoS Biol 13: e1002073
Rohe T, Noppeney U (2015b) Sensory reliability shapes perceptual inference via two mechanisms. J Vis 15:22.
Rohlf S, Bruns P, Röder B (2021) The Effects of Cue Reliability on Crossmodal Recalibration in Adults and Children. Multisens Res:1-19.
Rohlf S, Li L, Bruns P, Röder B (2020) Multisensory Integration Develops Prior to Crossmodal Recalibration. Curr Biol 30:1726-1732.e1727.
Shams L, Beierholm UR (2010) Causal inference in perception. Trends Cogn Sci 14:425-432.
Wagenmakers EJ, Farrell S (2004) AIC model selection using Akaike weights. Psychon Bull Rev 11:192-196.
Wozny DR, Shams L (2011) Recalibration of auditory space following milliseconds of cross-modal discrepancy. J Neurosci 31:4607-4612.
Zaidel A, Turner AH, Angelaki DE (2011) Multisensory calibration is independent of cue reliability. J Neurosci 31:13949-13962.
Zaidel A, Ma WJ, Angelaki DE (2013) Supervised Calibration Relies on the Multisensory Percept. Neuron.
Zierul B, Roder B, Tempelmann C, Bruns P, Noesselt T (2017) The role of auditory cortex in the spatial ventriloquism aftereffect. Neuroimage 162:257-268.