Multisensory causal inference is feature-specific, not object-based

Stephanie Badde*[1], Michael S Landy[2], & Wendy J Adams[3]

* corresponding author, stephanie.badde@tufts.edu

[1] Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA

[2] Department of Psychology & Center of Neural Science, New York University, 6 Washington Place, New York, NY 10003, USA

[3] Department of Psychology, University of Southampton, 44 Highfield Campus, Southampton SO17 1BJ, United Kingdom

Abstract

Multisensory integration depends on causal inference about the sensory signals. We tested whether implicit causal inference judgments pertain to entire objects or focus on task-relevant object features. Participants in our study judged virtual visual, haptic, and visual-haptic surfaces with respect to two features, slant and roughness, against an internal standard in a two-alternative forced-choice task. Modeling of participants' responses revealed that the degree to which their perceptual judgments were based on integrated visual-haptic information varied unsystematically across features. Thus, for example, a perceived mismatch between visual and haptic roughness would not deter the observer from integrating visual and haptic slant. These results indicate that participants based their perceptual judgments on a feature-specific selection of information, suggesting that multisensory causal inference proceeds not at the object level but at the level of single object features.

Introduction

At every moment in time, we perceive information through our different senses. Yet, these sensory signals do not provide an exact representation of the environment; they are perturbed by noise sources in the environment and in the nervous system. Multisensory integration, the combination of information from different senses, increases the reliability of perceptual estimates. Perceptual estimates based on integration of multiple sources of information are less variable than estimates based on one sensory signal alone (Ernst & Bülthoff, 2004; Fetsch et al., 2013; Trommershauser et al., 2011). However, multisensory integration is only beneficial if both sensory signals originate from the same object. Integrating information from separate sources reduces perceptual variability at the cost of introducing perceptual bias. Hence, multisensory integration should rely on causal inferences about the to-be-integrated sensory signals (Chen & Spence, 2017; Körding et al., 2007; Welch & Warren, 1980).

Consistent with a role of causal inference in multisensory perception, multisensory integration breaks down when the different modalities provide conflicting information. For example, the ventriloquism effect, which describes the mislocalization of sounds (the ventriloquist's utterances) towards a visual object (the puppet), decreases with increasing distance between the cues (Körding et al., 2007; Wallace et al., 2004). As another example, auditory frequency information interferes with tactile frequency perception but only across similar frequencies (Yau et al., 2009). Congruency of the to-be-integrated signals is not the only factor that guides multisensory causal inference: Multisensory integration of almost any feature is conditional on rough spatial and temporal alignment of the sensory signals (Alais et al., 2010; Calvert et al., 2004; Murray & Wallace, 2012). For example, perceptual judgments about visual and haptic stimuli with a large spatial (Gepshtein et al., 2005) or temporal (Parise & Ernst, 2016) offset show no integration effects. Instead, given a large spatiotemporal conflict, observers base their perceptual judgments primarily on only one modality. The influence of temporal and spatial information on multisensory integration of other object features raises the possibility that multisensory causal inference proceeds at the level of objects rather than single object features. Yet, previous studies introduced unmistakable spatial and temporal misalignments, leaving open whether in general task-irrelevant features are considered for multisensory causal inference.

Multisensory causal inference depends on factors beyond the physical properties of the stimuli. Despite perfect cross-modal correspondence between the physical stimuli, integration effects might be small or even absent in some participants (Battaglia et al., 2003; Meijer et al., 2019). Such modulations of cross-modal integration across participants are naturally accounted for by Bayesian causal-inference models (Körding et al., 2007; Sato et al., 2007). According to these models, the brain derives the posterior probability that the sensory signals originated from a common cause, and weighs the outcome of cross-modal integration and unimodal feature estimation accordingly. This posterior is the product of the a priori probability the observer assigns to the common-cause scenario and the likelihood that the sensory signals share a common cause. The common-cause prior is a top-down influence; it varies across stimuli (Odegaard & Shams, 2016), with the observer's previous experiences (Gau & Noppeney, 2016; Hong et al., 2022) as well as their attentional state (Badde et al., 2020). In turn, the likelihood of a common cause is driven by sensory information about the stimuli. Yet, these sensory signals might be biased, and these biases might be specific to one modality. For example, tactile but not visual stimuli on the arm are perceived as closer to the elbow than their actual location, which negatively affects the perceived alignment of physically aligned, visual-tactile stimulus pairs (Badde et al., 2020). Such perceptual cross-modal misalignment reduces the likelihood that the different sensory signals share a common cause and indeed has been identified as major source of reduced or absent cross-modal integration effects (Negen et al., 2022). Finally, integration effects might also seem reduced, if cross-modal information is not integrated in a statistically optimal fashion. But the prevalence of sub-optimal multisensory integration remains unclear as studies drawing this conclusion usually assume that all observers' assign 100% prior probability to the common-cause scenario and have no perceptual biases, which appears implausible (we nevertheless ensured that our conclusions do not critically depend on the assumption that observers behave statistically optimal, see S9). Importantly, if multisensory causal inference proceeds at the object level, the posterior probability that both signals arise from a common cause should be determined by a shared a priori probability of a common cause and a common-cause likelihood based on all sensory information about the encountered object. Thus, an observer who assigns a low a priori probability to the common cause scenario should do so independent of the task. Another observer's perceptual biases would affect the likelihood of a common cause even if the biased feature is currently irrelevant. In contrast, if multisensory causal inference proceeds on the feature level, an observer's common-cause prior might vary across features and perceptual misalignments of a currently irrelevant feature should not affect integration of another feature.

We tested whether multisensory causal inference is contingent upon all features of an object or alternatively proceeds at the level of single object features. To this aim, we asked participants to judge a series of virtual visual-haptic objects with respect to one of two features, roughness or slant (Fig. 1). Crucially, even though these features were judged independently, any external or sensory factors that might affect participants' causal inferences were identical across tasks. As outlined above, if causal inference pertains to all features of an object, the inferred probability that visual and haptic signals originate from the same source should be independent of the task. Consequently, the degree to which an observer bases perceptual decisions on integrated visual and haptic information, a proxy of the inferred probability that the signals share a common cause, should correspond across roughness and slant. In other words, under the object-based model, an observer who shows reduced integration effects should do so in both tasks. Thus, the degree to which participants rely on integration in the slant and roughness tasks should be correlated across participants. In contrast, if causal inference proceeds at the feature level, a perceived mismatch between visual and haptic roughness would not affect whether visual and haptic slant signals are judged as belonging to the same object, and observers might have feature-specific a priori assumptions about visual and haptic signals sharing a common cause. Therefore, under the feature-specific model, the extent to which an observer relies on integrated visual and haptic information might vary across tasks. To test these predictions, participants judged the roughness or slant of virtual visual, haptic, and visual-haptic surfaces against an internal standard. Performance in unisensory trials was used to predict visual-haptic performance given maximal integration effects, i.e., perceptual decisions based exclusively on optimally integrated visual and haptic sensory signals. This benchmark enabled us to quantify the degree to which participants relied on integration in visual-haptic trials, separately for each feature.

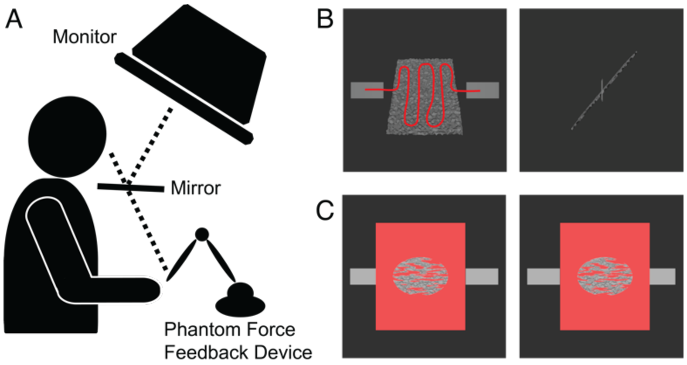

Fig. 1 Setup and Stimuli. (A) Participants viewed stereoscopically presented visual stimuli via a mirror so that they were perceived as co-located with virtual haptic stimuli rendered using a Phantom force-feedback device. (B) The stimuli were rough surfaces, slanted top-back from fronto-parallel. Participants were trained to haptically explore the surfaces following a sinusoidal path illustrated in red. (C) A red occluder was placed in front of the rough surfaces to limit geometric cues for surface slant. In visual and visual-haptic conditions, a peephole in the center of the occluder opened once participants touched the virtual stimulus. Participants wore active shutter glasses so that separate images could be presented to either eye (here, the image presented to the left eye is placed at the right side to enable crossed fusion). [Please set this as a single-column figure.]

Results

Participants varied considerably in the degree to which they relied on integrated visual and haptic information about stimulus slant and roughness. Consistent with feature-specific causal inference, some participants showed maximal integration effects for visual and haptic slant information but not visual and haptic roughness whereas other participants showed the opposite pattern (Fig. 2A; see S1 for all participants' psychometric curves and S2 for the extracted uncertainty).

We calculated an integration index for each feature and participant (Fig. 2B). This index relates the variability of perceptual estimates in visual-haptic trials to the variability predicted by optimal cue integration based on performance during the unisensory trials (Ernst & Banks, 2002; Landy et al., 1995). By doing so, we related the observed variability to the predicted variability given an inferred probability of 1 that visual and haptic signals share a common cause. If the observer relies exclusively on integrated sensory information, this index will be 1 on average (see S3 for an alternative index). If the observer relies partially on non-integrated, unimodal information, the ratio indicates the factor by which participants' response variability exceeds the benchmark variability. To distinguish between object-based and feature-specific multisensory causal inference, we calculated the product-moment correlation between participants' integration indices for slant and roughness. The posterior distribution of the correlation coefficient was centered at *r*=-0.01 and we obtained a Bayes factor of 3.7 favoring a correlation coefficient equal to zero as predicted by simulations with a model performing feature-specific causal inference (Fig. 3A). In contrast, simulations with a model performing object-based Bayesian causal inference predicted a correlation of *r*=0.48. The Bayes factor for this point hypothesis is 5.9 against the correlation predicted by object-based multisensory causal inference.

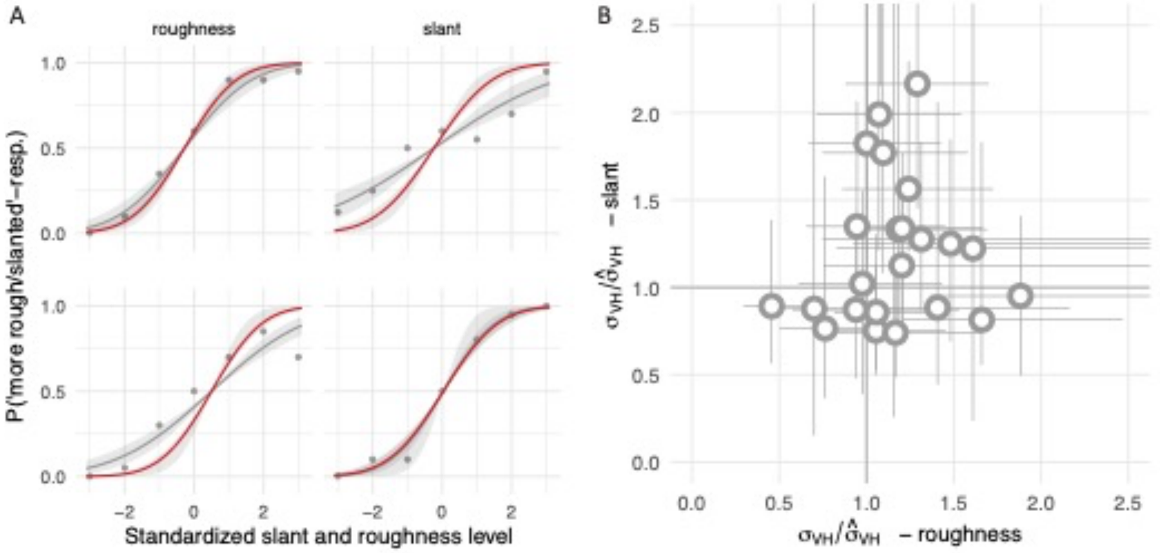

Fig. 2 Results. (A) Psychometric curves for two participants (one per row) in the visual-haptic condition of the roughness (left column) and slant (right column) task. Markers indicate the observed proportion of 'more rough / more slanted than the standard' responses for each feature level shown on a common scale for roughness and slant. Grey curves show psychometric curves fitted to these data; red curves show psychometric curves corresponding to maximal integration effects given the participant's performance in unimodal trials (see S1 for all participants and all conditions). Shaded ribbons indicate 95% confidence intervals for both curves. Top row: sample participant who showed maximal integration effects for roughness but not for slant. Bottom row: sample participant who showed the reversed pattern. (B) Integration indices for both features and all participants. The integration index is the ratio of the standard deviation of the fitted visual-haptic curve and the predicted curve assuming maximal integration effects (see S3 for an alternative index). An index of one indicates maximal integration effects

while larger values suggest less-than-maximal integration effects, indicating perceptual judgments partially based on unimodal information. Error bars indicate95% confidence intervals obtained by bootstrapping the raw data. [Please set this as a two-column figure.]

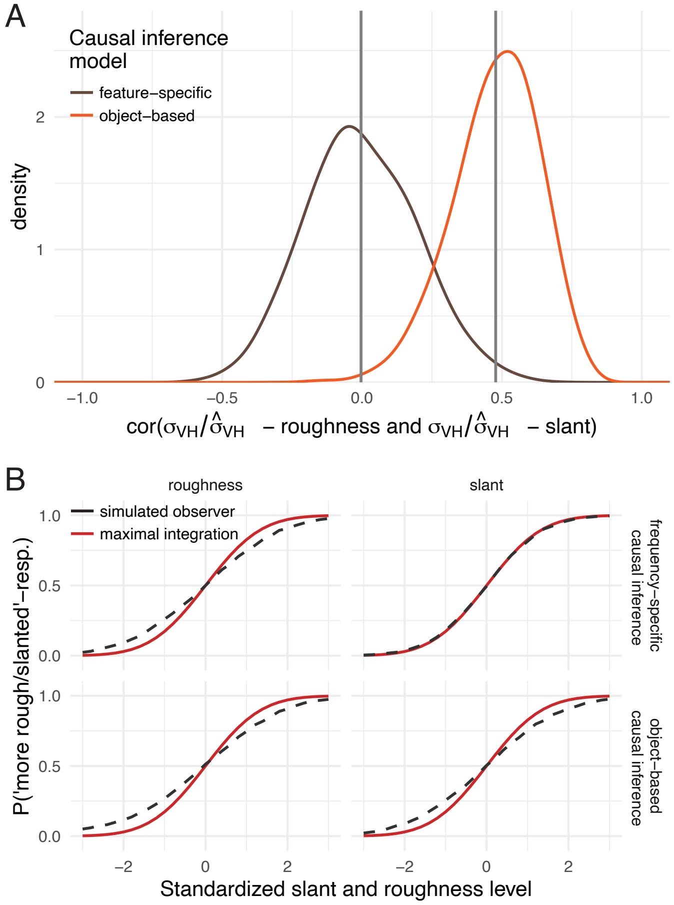

Fig. 3 Model Predictions. (A) Distribution of simulated correlation coefficients between the integration indices for roughness and slant. Correlation coefficients are based on 10,000 simulated datasets of the same size as the original data (26 participants, 20 trials per condition). Data were generated using the feature-specific (brown) and object-based (orange) causal-inference models (see Methods). Vertical lines indicate distribution means. (B) Visual-haptic psychometric curves (black dashed lines) for a single simulated observer with modality-specific biases for roughness but not slant. The feature-specific causal-inference model (top row) predicts a clear deviation from optimal cue integration (red lines), i.e., less than maximal integration effects, for roughness but not for slant whereas the object-based (bottom row) causal-inference model predicts reduced integration effects for both features. [Please set this as a two-column figure.]

Discussion

We tested whether multisensory causal inference is object-based or selectively refers to task-relevant features. To this aim, participants separately judged the slant and roughness of virtual visual, haptic, and visual-haptic surfaces. The degree to which participants relied on integrated visual and haptic information, i.e., treated these signals as originating from the same object, varied unsystematically between surface slant and roughness. These results indicate that multisensory causal inference proceeds not at the object level, but at the level of single object features.

At first glance, our conclusion that multisensory causal inference is feature-specific might seem at odds with the general notion that the perceptual system makes optimal use of all available information. After all, conflict between the senses with respect to a task-irrelevant feature provides strong evidence against a shared origin of the sensory signals. Yet, situations in which such a conflict proves critical for multisensory causal inference might be rare outside of the laboratory. First, sensory signals from different sources will typically also be profoundly misaligned in space or time and multisensory causal inference takes rough location and temporal congruency into account (Körding et al., 2007; McGovern et al., 2016; Wallace et al., 2004). Second, many object features are modality specific. Thus, the number of non-task relevant object features relevant to multisensory causal inference might typically be limited to the location and the onset of the event. Third, supramodal object features and the sensory noise associated with them might be correlated in the real world and thus provide no independent evidence. In sum, multisensory causal inference might not pertain to entire objects because such a potentially costly mechanism rarely provides a perceptual advantage.

The finding that multisensory causal inference does not pertain to entire objects might further seem at odds with reports that multisensory integration is fostered by semantic congruency between cross-modal stimuli (Doehrmann & Naumer, 2008). However, results are mixed regarding the influence of semantic congruency on multisensory integration. Some studies find that congruency between images and sounds of everyday objects (Noppeney et al., 2008) or the gender or emotion of a speaker’s face and voice (Dolan et al., 2001; Vatakis & Spence, 2007) facilitates multisensory integration, whereas other studies report no effects of additional high-level cues towards a common cause of visual and auditory signals (Radeau & Bertelson, 1978; Vatakis & Spence, 2008). Furthermore, semantic congruency refers to the category of the presented object rather than cross-modal congruency of sensory feature information (Chen & Spence, 2017). For example, the sound and sight of a barking dog are semantically congruent, but this congruency does not result from shared features across vision and audition, but rather on modality-specific visual and auditory features of a dog. Thus, taking all cross-modal feature information into account for the sensory-driven likelihood of a common cause would not necessarily lead to the identification of semantically congruent stimuli. Instead, semantic congruency might influence multisensory causal inference by affecting an observer's prior assumptions about a shared origin of cross-modal information. Hence, our result that multisensory causal inference proceeds at the level of single features is not at odds with the notion that semantic congruency affects multisensory causal inference.

The degree to which participants relied on integrated visual and haptic information varied unsystematically across roughness or slant; the integration indices for roughness and slant did not correlate. This absence of a correlation suggests that the inferred probability of a common cause for visual and haptic signals for slant and roughness varied independently across participants, which speaks in favor of the feature-based account of multisensory causal inference. However, it should be noted that, although it is unlikely, such a result is not impossible under the object-based account. Given the stochastic nature of perception in combination with practical limitations on the number of trials per participant, data are bound to be noisy, which decreases the ability to measure an existing correlation. To derive quantitative predictions for the correlation between integration indices given the feature-specific and object-based accounts, we used simulations of a model performing either type of multisensory causal inference (Fig. 3A). The model assumes that observers establish two intermediate estimates of the to-be-judged feature, one based on optimal cue integration of visual and haptic sensory signals, and one based on their favorite modality, the modality they would choose if visual and haptic signals were from different sources. Analogous to previous implementations of Bayesian multisensory causal inference (Badde et al., 2020; Hong et al., 2022, 2022; Körding et al., 2007), these two intermediate estimates are averaged, weighted by the posterior probability of a common cause.

Thus, if the inferred probability that the signals share a common cause is 1, the observer fully bases their perceptual decisions on the integrated estimate and the variance across visual-haptic trials is identical to that predicted by optimal cue integration (the denominator of the integration index). In turn, the lower the inferred probability of a common cause, the more perceptual judgments rely on unimodal information and the larger the variance across visual-haptic trials. The models corresponding to our two hypotheses, object-based and feature-specific causal inference differ with respect to the information that is used to calculate the posterior probability of a common cause. For object-based causal inference the common-cause posterior is derived based a general common cause prior and on all available sensory information about the presented object. Thus, visual and haptic sensory signals indicating the roughness of a presented surface are included in the likelihood that visual and haptic slant signals originated from the same source and vice versa. Consequently, under the object-based model, a perceptual mismatch for roughness will also affect the posterior probability of a common cause and with it the degree of integration in the slant task (Fig. 3B, see Methods and S7 for details). In contrast, our model of feature-specific causal inference relies only on task-relevant sensory information to infer the trial-wise likelihood of a common cause and allows for different a priori assumptions about a common-cause for slant and roughness. Our simulations predict correlations of zero and 0.5 given feature-specific and object-based multisensory causal inference, respectively (Fig. 3A). Based on our data we can accept the former and reject the latter correlation coefficient and with it corresponding account of multisensory causal inference.

In sum, our evidence indicates that multisensory causal inference proceeds at the level of single features rather than entire objects.

Methods

Participants

Twenty-six members of the University of Southampton (16 females, aged 18-34, mean 23 years) participated in the study. All participants reported to have unimpaired or corrected-to-normal vision and to be free of tactile as well as motor impairments. Written informed consent was obtained prior to the experiment. The experiment was approved by the institutional review board of the University of Southampton.

Apparatus and Stimuli

Participants were seated, their head supported by a chin and forehead rest mounted at an angle so that their head was slightly bent forward. The index finger of their dominant hand was placed in a thimble attached to a Phantom force feedback device (GeoMagic, http://www.3dsystems.com). This device measures the fingertip position and exerts a precisely controlled force vector on the fingertip, which allows the user to feel and interact with virtual haptic objects. Participants viewed the display of a CRT monitor via a mirror (Fig. 1A). Position and angle of the mirror were set to evoke the impression that visual and virtual haptic stimuli were in the same plane, located at about table height and 57 cm distance from the participant's eyes. To create the illusion of a three-dimensional visual stimulus, different images (Fig. 1C) were presented to the left and right eye using active shutter glasses (Stereographics Crystal Eyes).

The virtual stimuli were textured rectangles (20 cm wide, 30 cm high; Fig. 1B), which were slanted top-backwards from fronto-parallel (defined with respect to the visual plane, Fig. 1A) by 26 to 38 deg. To create a rough plane, first a 2D grid of 400 x 600 points was created. The initial spacing between points was 5 mm along either axis. To reduce pattern regularity, each grid point's *x*- and *y*-coordinates were randomly and independently shifted by -2 to 2 mm with the shifts being uniformly distributed. Half of the grid points, chosen randomly, were assigned a *z*-coordinate of 0. The *z*-coordinates of the other half of the grid points were drawn from a Gaussian distribution with a standard deviation of 0.1 mm. The mean value of this Gaussian determined the roughness of the stimulus and ranged from 3 to 6 mm. Faces were added to this 3D grid by building triplets of adjacent vertices so that the diagonals were in one direction in the even rows and in the other direction in the odd rows. The textured plane was flanked by two smaller, smooth rectangles located to its left and right (Fig. 1B). These outer bars were placed at the same distance from the observer as the textured rectangle and were aligned with its horizontal midline and the fronto-parallel plane. To prevent the participant from inferring the rough surface's slant from the perspective geometry of the image, view of the rectangle was partly occluded by a larger, red rectangle located in between the stimulus and the observer, at 15 cm distance from the stimulus rectangle. This occluder had a round cutout filled with a lacy, irregular structure to reduce the reliability of visual cues about the roughness of the stimulus with the goal to match the reliability of visual and haptic cues as assessed during piloting (Fig. 1C). Haptic and visual stimuli were coded in Python using the bpy module and (pre-)rendered using Blender (http://www.blender.org). Visual stimuli were rendered for left and right eye viewpoints, assuming an inter-eye distance of 6 cm. The experiments were programmed in C++ interfacing with Open-Haptics to control the haptic device and OpenGL to present the pre-rendered visual stimuli as well as trial information, response buttons, and a visual cursor that indicated the position of the participant's index finger.

Task and Design

Participants compared the roughness or slant of a visual, haptic, or visual-haptic test stimulus to that of a remembered standard stimulus presented in the same modality (one-interval, two-alternative, forced choice). The standard stimulus was presented at the beginning and at regular time points throughout each block of trials, and it was identical across experiments. Roughness and slant of the standard were equal to the average over the test stimuli: a roughness with extrusions of on average 4.5 mm and slanted top-back by 32 deg relative to fronto-parallel. Test stimuli in the roughness experiment had a roughness of 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, or 6.0 mm and were slanted top-back by 32 deg; test stimuli in the slant experiment were slanted top-back by -38, -36, -34, -32, -30, -28, or -26 deg and had a roughness of 4.5 mm.

Procedure

At the beginning of a trial, the stimulus was hidden behind a solid red occluder to ensure comparable exploration times for visual and haptic stimulus features in visual-haptic trials. Participants were instructed to move their finger to the left bar flanking the textured stimulus. In haptic and visual-haptic trials, participants then moved their finger to the left outer edge of the textured stimulus and explored it following a sinusoidal path (Fig. 1B, left panel). In visual trials, the textured plane was not haptically rendered, participants kept their finger on the left outer bar until they were ready to make a response. In visual trials, the lacy peephole at the center of the occluder (Fig. 1C) would open once participants touched the left bar; in visual-haptic trials it would open once they touched the rough texture. In visual-haptic trials the peephole closed once participants moved their finger away from the stimulus and reopened as soon as they touched the stimulus again. Virtual buttons located above the stimulus were visually and haptically rendered 500 ms after the beginning of the trial. When the standard stimulus was presented, only one button, labelled "Standard" was rendered; when a test stimulus was presented, two buttons were rendered. These buttons were labelled "Less Rough" and "More Rough" when stimulus roughness was judged and "Forwards" and "Backwards" for slant judgments. The trial ended once participants pressed one of the virtual buttons. No feedback was provided, and stimulus exploration time was not limited.

Six trials in which the standard stimulus was presented occurred at the beginning of each block and the standard stimulus was presented again after every three test-stimulus presentations. Visual, haptic, and visual-haptic trials were blocked. A block consisted of five repetitions of the seven test-stimulus levels, presented in randomized order, and each participant completed four blocks per modality resulting in 20 repetitions per stimulus. The three modality conditions were alternated across blocks; order was varied across participants but held constant within participants across the roughness and slant-discrimination tasks. Participants completed the two tasks in random order. Each task took 2-3 hours to complete. Testing was spread across several sessions.

Data Analysis

Test stimulus levels were described using a common scale for both slant and roughness, ranging from -3 to 3. Cumulative Gaussian distribution functions Φ with a lapse rate λ, were fit to the proportion of more rough / more top-back responses as a function of stimulus level s, $p(s) = \frac{\lambda}{2} + (1 - \lambda)\Phi(s; \mu, \sigma^2)$ using maximum likelihood. (We did not fix μ at 0, the feature level of the standard stimulus, as participants might have formed a biased internal representation of the standard stimulus. Doing so, as well as adding the lapse rate does not affect the outcome of our main analysis, see S4.) Six separate psychometric functions were fitted, one for each combination of task (roughness and slant discrimination) and stimulus modality (visual, haptic, and visual-haptic). 95% confidence intervals for the parameter estimates were derived by bootstrapping the data stratified by feature level and repeating the fitting procedure for each bootstrap.

If participants optimally combine visual and haptic information in visual-haptic trials and rely exclusively on the outcome of this integration, the variance of the psychometric function in this condition is a function of the unimodal variances, $\widehat{\sigma_{vh}^2} = \frac{\sigma_v^2 \sigma_h^2}{\sigma_v^2 + \sigma_h^2}$ (Ernst & Banks, 2002; Landy et al., 1995). We quantified the degree to which participants relied on visual-haptic integration by calculating the ratio of the estimated and predicted visual-haptic variances, $\frac{\sigma_{vh}^2}{\widehat{\sigma_{vh}^2}}$. If participants base their perceptual decisions exclusively on the optimally integrated estimate, this index will be 1 on average. If not, the ratio indicates the factor by which participants' response variability exceeds the variability given full integration (see S3 for an alternative index). Two integration indices were calculated for each participant, one for each task. We used the estimated variance parameters of the three psychometric functions as estimates of visual, haptic, and visual-haptic variances. Thus, we implicitly assumed that the internal standard stimulus did not contribute to the slope of the psychometric function. This simplifying assumption had only negligible influence on the integration index (S5).

The two alternative accounts of multisensory causal inference make predictions about the correlation between the integration indices for roughness and slant. We approximated the posterior distribution of the correlation coefficient $\rho_{\mathrm{idx}_r \mathrm{idx}_s}$using Markov chain Monte Carlo sampling. Specifically, a two-dimensional Gaussian with covariances parametrized as $\{\sigma_{\mathrm{idx}_r}, \sigma_{\mathrm{idx}_s}, \rho_{\mathrm{idx}_r \mathrm{idx}_s}\}$ was fit to the pairs of participants' ($i = 1, \ldots, n$) integration indices $\mathrm{idx}_{r,i}$ and $\mathrm{idx}_{s,i}$ using Stan's (Stan Development Team, 2022) leapfrog algorithm (see S6 for details). Bayes factors for point hypothesis $H0{:}\, \rho = \rho_0$ and $H1{:}\, \rho \neq \rho_0$ were calculated based on the ratio of the posterior and prior densities at ρ_0 (Wagenmakers et al., 2010). We used a log-spline function to estimate the densities from the distribution of the samples.

A correlation coefficient of zero would indicate that causal inference proceeded at the feature level and a positive correlation would indicate object-level causal inference. The range of correlation coefficients we can expect in the latter scenario is not self-evident. Given the probabilistic nature of perceptual decisions and natural restrictions on the number of administered trials per participant, the estimated variances are associated with an error that is independent across features and thus should reduce the measurable correlation. We established that with 26 participants and 20 trials per stimulus level and condition we would be able to find a correlation by running simulations of our experiment, under the assumptions that participants perform object-level causal inference. In more detail, we used either a feature-specific or an object-based Bayesian causal-inference model (see below) to generate 10,000 datasets of the same size as our original data. Model parameters for each simulated participant were sampled from the range of parameter estimates we obtained for our real participants. Each artificial dataset was analyzed in the same way as the original data and the correlation between the integration indices for roughness and slant was stored (see S8 for representative examples).

Data analyses were performed in R (Version 4.2.2), causal-inference models were implemented in Python (Version 3.8.8). Code and raw data are publicly available (DOI will be provided after acceptance of the manuscript).

Models of Object-Based and Feature-Specific Causal Inference

We assumed that observers solved either task by comparing an estimate $\hat{s}_i$ of the relevant stimulus feature, e.g., the roughness of the surface presented in trial *i*, to their internal representation of the standard stimulus (see S7 for the full set of equations specifying each model). We allowed this internal representation of the standard stimulus to be biased. We further assumed that, to derive the estimate of the feature, observers relied on a weighted average of an optimally integrated visual-haptic estimate, $\hat{x}_{vh,i,C=1}$, and an estimate $\hat{x}_{v \text{ or } h,i,C=2}$ based on the sensory signal in their ‘favorite’ modality. Their ‘favorite’ modality refers to the modality they relied on if they had to choose between vision and haptics because they perceived the signals as originating from different sources. We assumed that this modality preference was constant across trials. The weighting of the integrated and unimodal estimates depends on the posterior probability that both sensory signals $(m_{v,i}, m_{h,i})$ originated from the same source ($C = 1$), i.e., $\hat{x}_i = P\left(C = 1 \middle| m_{v,i}, m_{h,i}\right) \hat{x}_{vh,i,C=1} + P\left(C = 2 \middle| m_{v,i}, m_{h,i}\right) \hat{x}_{v \text{ or } h,i,C=2}$. This is equivalent to model averaging in Bayesian causal inference (Körding et al., 2007) with the only difference being that typically one of the modalities is the 'favorite' modality by instruction.

We further assumed that the sensory signals, also called measurements, $m_{v,i}, m_{h,i}$ were corrupted by Gaussian-distributed noise with variance σ_v^2 and σ_h^2. We additionally allowed the sensory signals to be biased (Badde et al., 2020; Hong et al., 2021, 2022), as modality-specific biases in the sensory signals are a root cause of reduced cross-modal integration effects (Negen et al., 2022).

According to our object-based causal-inference model, the posterior probability of a common cause in trial *i* is derived based on the likelihoods of a common and separate causes given all available sensory information about the object presented in that trial (i.e., the visual and haptic measurements of slant and roughness, $m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}}, m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}}$) and a task-independent prior probability that visual and haptic signals originate from a common cause $p_{\mathrm{c=1}}$,

$$P\left(C_{\mathrm{s(lant)\ or\ r(oughness)}} = 1 \middle| m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}}, m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}}\right) =$$

$$\frac{p_{\mathrm{c=1}}\, P(m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}}, m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}} | C=1)}{p_{c=1}\, P(m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}}, m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}} | C=1) + (1-p_{c=1})\, P(m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}}, m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}} | C=2)}$$

(see S7 for a full specification of the likelihoods). Thus, the posterior probability of a common cause is derived identically across tasks. In contrast, the feature-specific multisensory causal-inference model assumes that the likelihoods of common and separate causes refer only to task-relevant sensory information (e.g., the visual and haptic measurements of the slant of the surface presented in trial *i*, $m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}}$) and allow for separate common-cause priors, one for slant, $p_{\mathrm{c=1},s}$, and one for roughness, $p_{\mathrm{c=1,r}}$. Thus, the posterior probability for a common cause in the slant task,

$P\left(C_{\mathrm{s(lant)}} = 1 \middle| m_{v,\mathrm{s}}, m_{h,\mathrm{s}}\right) = \frac{p_{\mathrm{c=1,s}}\, P(m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}} | C=1)}{p_{\mathrm{c=1,s}}\, P(m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}} | C = 1) + (1-p_{\mathrm{c=1,s}})\, P(m_{v,\mathrm{s,i}}, m_{h,\mathrm{s,i}} | C=2)}$, differs from the one in the roughness task $\left(C_{\mathrm{r(oughness)}} = 1 \middle| m_{v,\mathrm{r}}, m_{h,\mathrm{r}}\right) = \frac{p_{\mathrm{c=1,r}}\, P(m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}} | C=1)}{p_{\mathrm{c=1,r}}\, P(m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}} | C = 1) + (1-p_{\mathrm{c=1,r}})\, P(m_{v,\mathrm{r,i}}, m_{h,\mathrm{r,i}} | C=2)}$.

Acknowledgements

This work was supported by the National Institute of Health, grant NIH EY08266 to MSL and the German Research Foundation, grant BA 5600 1/1 to SB. We thank Walter Wang and Davide Deflorio for help with data acquisition and Erich Graf for technical and logistical support.

References

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory Processing in Review: From Physiology to Behaviour. *Seeing and Perceiving*, *23*(1), 3–38. https://doi.org/10.1163/187847510X488603

Badde, S., Navarro, K. T., & Landy, M. S. (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. https://doi.org/10.1016/j.cognition.2019.104170

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391. https://doi.org/10.1364/JOSAA.20.001391

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. MIT Press.

Chen, Y.-C., & Spence, C. (2017). Assessing the Role of the “Unity Assumption” on Multisensory Integration: A Review. *Frontiers in Psychology*, *8*, 445. https://doi.org/10.3389/fpsyg.2017.00445

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, *1242*, 136–150. https://doi.org/10.1016/j.brainres.2008.03.071

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(17), 10006–10010. https://doi.org/10.1073/pnas.171288598

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442. https://doi.org/10.1038/nrn3503

Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, *124*, 876–886. https://doi.org/10.1016/j.neuroimage.2015.09.045

Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, *5*(11), 7. https://doi.org/10.1167/5.11.7

Hong, F., Badde, S., & Landy, M. S. (2021). Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception. *PLOS Computational Biology*, *17*(11), e1008877. https://doi.org/10.1371/journal.pcbi.1008877

Hong, F., Badde, S., & Landy, M. S. (2022). Repeated exposure to either consistently spatiotemporally congruent or consistently incongruent audiovisual stimuli modulates the audiovisual common-cause prior. *Scientific Reports*, *12*(1), 15532. https://doi.org/10.1038/s41598-022-19041-7

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*(3), 389–412. https://doi.org/10.1016/0042-6989(94)00176-M

McGovern, D. P., Roudaia, E., Newell, F. N., & Roach, N. W. (2016). Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Scientific Reports*, *6*(1), 24673. https://doi.org/10.1038/srep24673

Meijer, D., Veselič, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex*, *119*, 74–88. https://doi.org/10.1016/j.cortex.2019.03.026

Murray, M. M., & Wallace, M. T. (Eds.). (2012). *The Neural Bases of Multisensory Processes*. CRC Press/Taylor & Francis. http://www.ncbi.nlm.nih.gov/books/NBK92848/

Negen, J., Slater, H., Bird, L.-A., & Nardini, M. (2022). Internal biases are linked to disrupted cue combination in children and adults. *Journal of Vision*, *22*(12), 14. https://doi.org/10.1167/jov.22.12.14

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The Effect of Prior Visual Information on Recognition of Speech and Sounds. *Cerebral Cortex*, *18*(3), 598–609. https://doi.org/10.1093/cercor/bhm091

Odegaard, B., & Shams, L. (2016). The Brain's Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychological Science*, *27*(4), 583–591. https://doi.org/10.1177/0956797616628860

Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications*, *7*(1), 11543. https://doi.org/10.1038/ncomms11543

Radeau, M., & Bertelson, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Perception & Psychophysics*, *23*(4), 341–343. https://doi.org/10.3758/BF03199719

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian Inference Explains Perception of Unity and Ventriloquism Aftereffect: Identification of Common Sources of Audiovisual Stimuli. *Neural Computation*, *19*(12), 3335–3355. https://doi.org/10.1162/neco.2007.19.12.3335

Stan Development Team. (2022). *Stan Modeling Language Users Guide and Reference Manual, 2.31*. https://mc-stan.org

Trommershauser, J., Körding, K. P., & Landy, M. S. (2011). *Sensory cue integration*. Oxford University Press.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756. https://doi.org/10.3758/BF03193776

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*(1), 12–23. https://doi.org/10.1016/j.actpsy.2006.12.002

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2). https://doi.org/10.1007/s00221-004-1899-9

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Yau, J. M., Olenczak, J. B., Dammann, J. F., & Bensmaia, S. J. (2009). Temporal Frequency Channels Are Linked across Audition and Touch. *Current Biology*, *19*(7), 561–566. https://doi.org/10.1016/j.cub.2009.02.013

S1: Psychometric Curves - All Participants

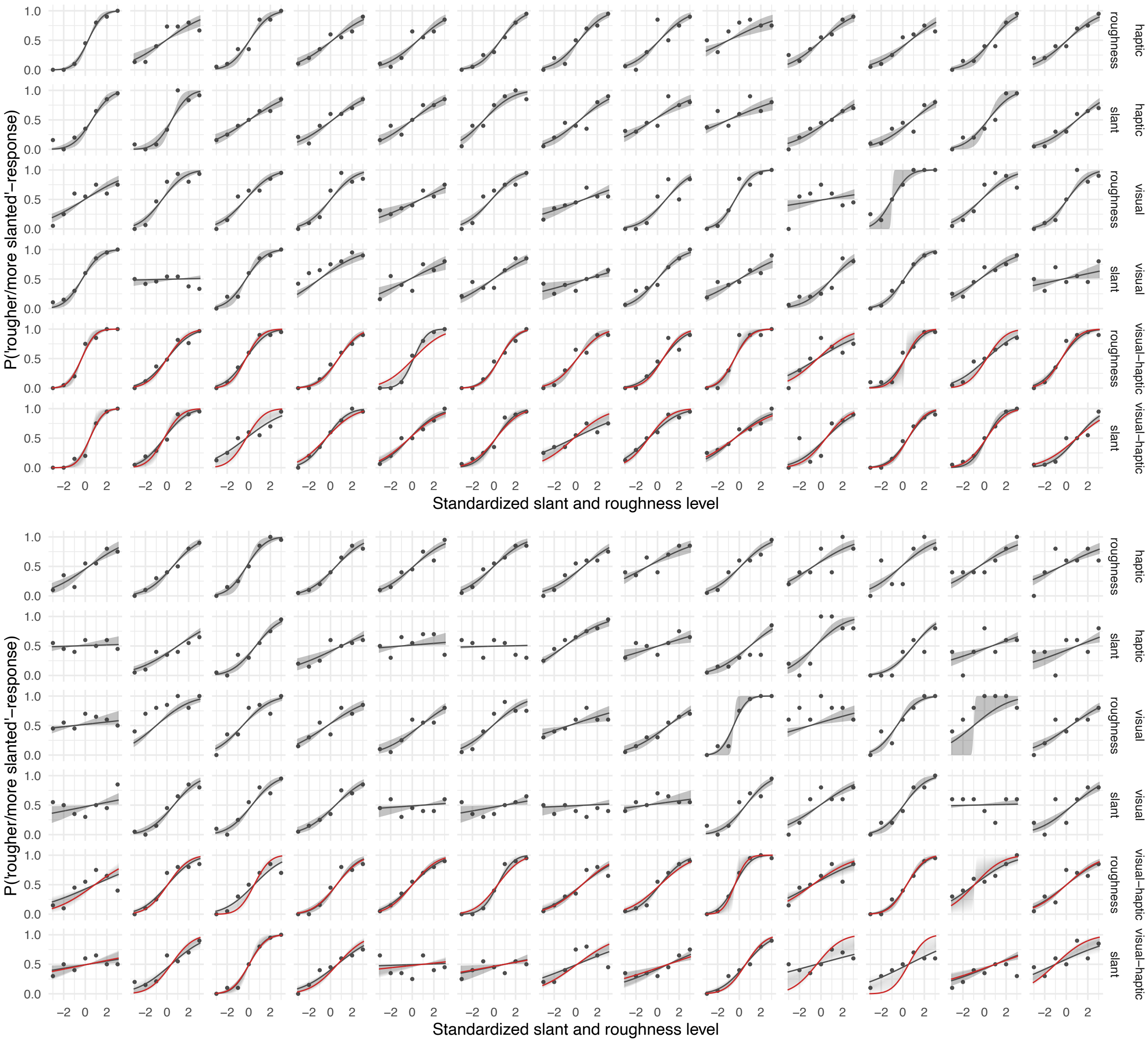

Figure 1: Psychometric curves for all participants (one per column, 13 per panel) in the haptic (top two rows), visual (middle two rows), and visual-haptic (bottom two rows) condition of the roughness and slant task. Markers indicate the observed proportion of 'more rough / more slanted than the standard' responses for each feature level shown on a common scale for roughness and slant. Grey curves show psychometric curves fitted to these data; red curves show psychometric curves predicted by the ideal-observer model based on single-cue performance. Shaded ribbons indicate 95% confidence intervals for the gray curves.

S2: Uncertainty - All Participants

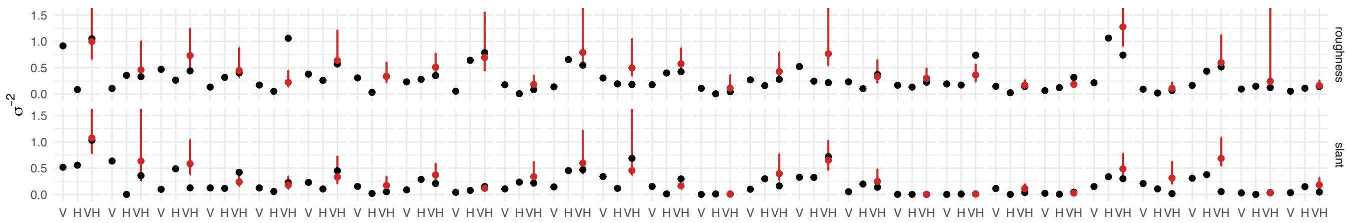

Figure 2: Visual, haptic, and visual-haptic uncertainty estimates for all participants in the roughness (top row) and slant (bottom row) task. Red markers indicate the predicted visual-haptic uncertainty assuming maximal integration effects, i.e., perceptual judgments relying only on optimal cue integration. Red bars indicate 95% confidence intervals for the predictions.

S3: Integration Indices

The variance reduction associated with cross-modal integration is maximal given equally noisy sensory signals. In turn, if the noisiness of the sensory signals is very unbalanced the difference the standard deviation of the more reliable modality might be indistinguishable from that predicted by optimal cue integration. We color-coded the imbalance of the noisiness of the sensory signals to visually check if the lack of a correlation was driven by integration indices based on very imbalanced sensory signals. This is not the case.

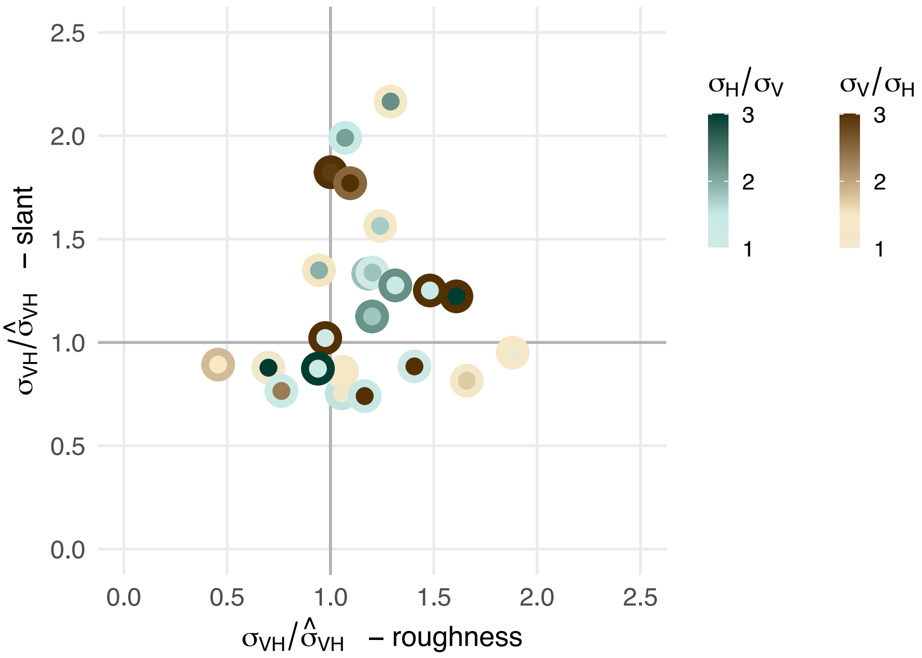

Figure 3: Integration index as function of relative uncertainty of the sensory signals (color-code). Turquoise hues represent haptic-to-visual ratios larger than 1, brown hues represent visual-to-haptic ratios larger than 1, slant is shown as inner color, roughness as color of the outer ring, colors were restricted to a sensible range.

We additionally derived an anchored index to address this blind spot of our more traditional index. This anchored index relates the observed distance between the standard deviation of the best cue and the standard deviation in the visual-haptic condition ($\sigma_{\text{best cue}} - \sigma_{vh}$) to the maximal possible reduction in noise ($\sigma_{\text{best cue}} - \hat{\sigma}_{vh}$) and thus is anchored to 'no integration' (best cue) and to 'full integration' (optimal cue combination). A value of one indicates that the observer relies on the best cue, smaller values indicate integration, whereas larger values indicate that the observer went with the worse cue. The estimated correlation between the anchored integration indices for roughness and slant is $r = 0.04$. We again visually checked whether participants for whom only the traditional index suggests maximal integration effects exerted large influence on the data but found no evidence in that direction.

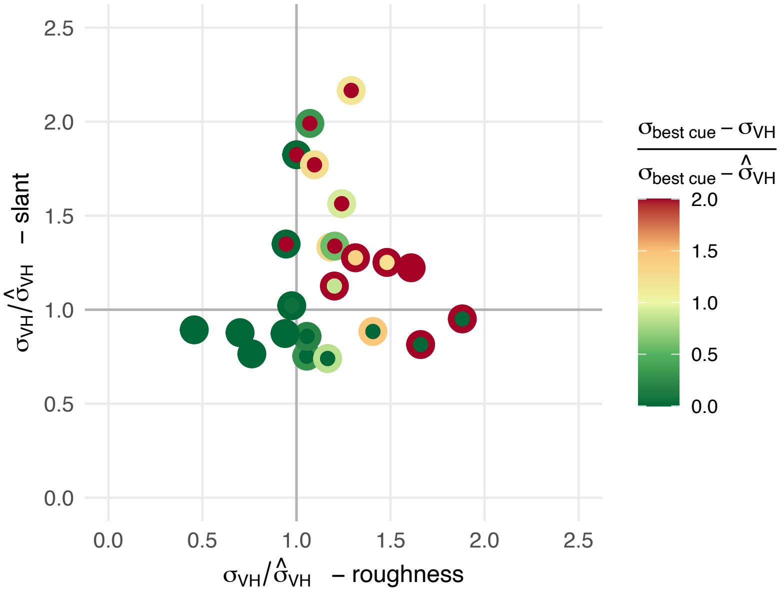

Figure 4: Integration index as function of an anchored integration index (color-code). The anchored index for slant is shown as the inner color, the index for roughness as the color of the outer ring. Colors were restricted to a sensible range. Yellowish inner circles close to y=1 and yellowish outer rings close to x = 1 might be misleading, as only the traditional integration index suggests maximal integration effects for these. However, such data points are rare.

S4: Psychometric Curves and Integration Indices

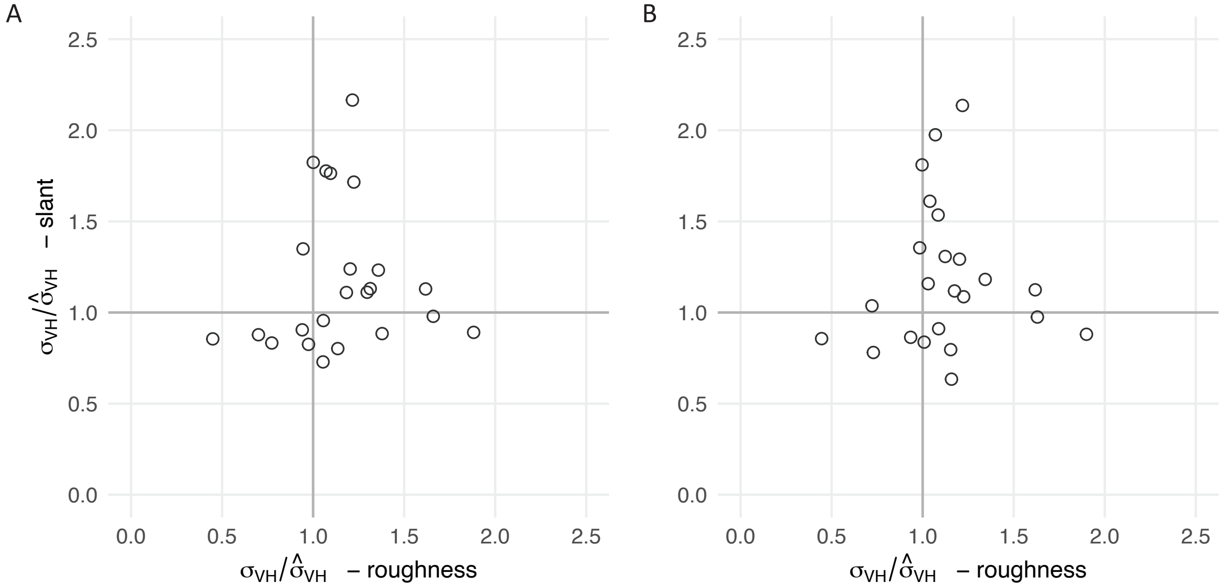

Figure 5: There was still no correlation between participants' integration indices for slant and roughness if the psychometric curve fitting did not allow for (A) lapses or (B) biased representations of the learned standard stimulus.

S5: Uncertainty in the Representation of the Standard Stimulus

In our tasks, participants compared the surface they encounter to a memorized standard stimulus. By taking the standard deviation of the psychometric function as an estimate of the participant's visual, haptic, or visual-haptic noise, we neglect the contribution of the representation of the standard stimulus to the slope of the psychometric function and consequently might be underestimating the reliability of our stimuli. However, as this affects the numerator and denominator of the integration index, the error we potentially introduce to the index remains minuscule as we show below.

We assume that the noise associated with the representation of the standard stimuli is Gaussian. Then we can express the variance of the standard as a proportion of the sensory noise for each modality, for example, $\alpha_h \sigma_h$ for the haptic standard stimulus.

The integration index directly derived from the slopes of the psychometric functions can be expressed as $\frac{\sqrt{\sigma_{vh}^2+\alpha_{vh}\sigma_{vh}^2}}{\sqrt{\frac{(\sigma_v^2+\alpha_v\sigma_v^2)(\sigma_h^2+\alpha_h\sigma_h^2)}{\sigma_v^2+\alpha_v\sigma_v^2+\sigma_h^2+\alpha_h\sigma_h^2}}}$.

To estimate the amount of over- or underestimation of the true index that we introduce by dropping the noise of the standard stimulus from the integration index, we simplify the following expression:

$$\begin{aligned}
&\frac{\sqrt{\sigma_{vh}^2+\alpha_{vh}\sigma_{vh}^2}}{\sqrt{\frac{(\sigma_v^2+\alpha_v\sigma_v^2)(\sigma_h^2+\alpha_h\sigma_h^2)}{\sigma_v^2+\alpha_v\sigma_v^2+\sigma_h^2+\alpha_h\sigma_h^2}}} \Bigg/ \frac{\sqrt{\sigma_{vh}^2}}{\sqrt{\frac{\sigma_v^2\sigma_h^2}{\sigma_v^2+\sigma_h^2}}} \\
=&\frac{(\sigma_{vh}^2+\alpha_{vh}\sigma_{vh}^2)(\sigma_v^2+\alpha_v\sigma_v^2+\sigma_h^2+\alpha_h\sigma_h^2)}{(\sigma_v^2+\alpha_v\sigma_v^2)(\sigma_h^2+\alpha_h\sigma_h^2)} \Bigg/ \frac{\sigma_{vh}^2(\sigma_v^2+\sigma_h^2)}{\sigma_v^2\sigma_h^2} \\
=&\frac{(1+\alpha_v+\alpha_{vh}+\alpha_v\alpha_{vh})\sigma_v^2+(1+\alpha_h+\alpha_{vh}+\alpha_h\alpha_{vh})\sigma_h^2}{(1+\alpha_h+\alpha_v+\alpha_v\alpha_h)(\sigma_v^2+\sigma_h^2)}.
\end{aligned}$$

Hence, the index is not affected if the noise introduced by the standard is a constant proportion of the sensory noise, which is not unlikely given that the standard is learned through exploration. We underestimate the index if $(-\alpha_h + \alpha_{vh} - \alpha_v\alpha_h + \alpha_v\alpha_{vh})\sigma_v^2 < (\alpha_v - \alpha_{vh} + \alpha_v\alpha_h - \alpha_h\alpha_{vh})\sigma_h^2$ and overestimate otherwise. The error remains relatively small compared to the integration index our conclusions are based on and thus does not affect our conclusions.

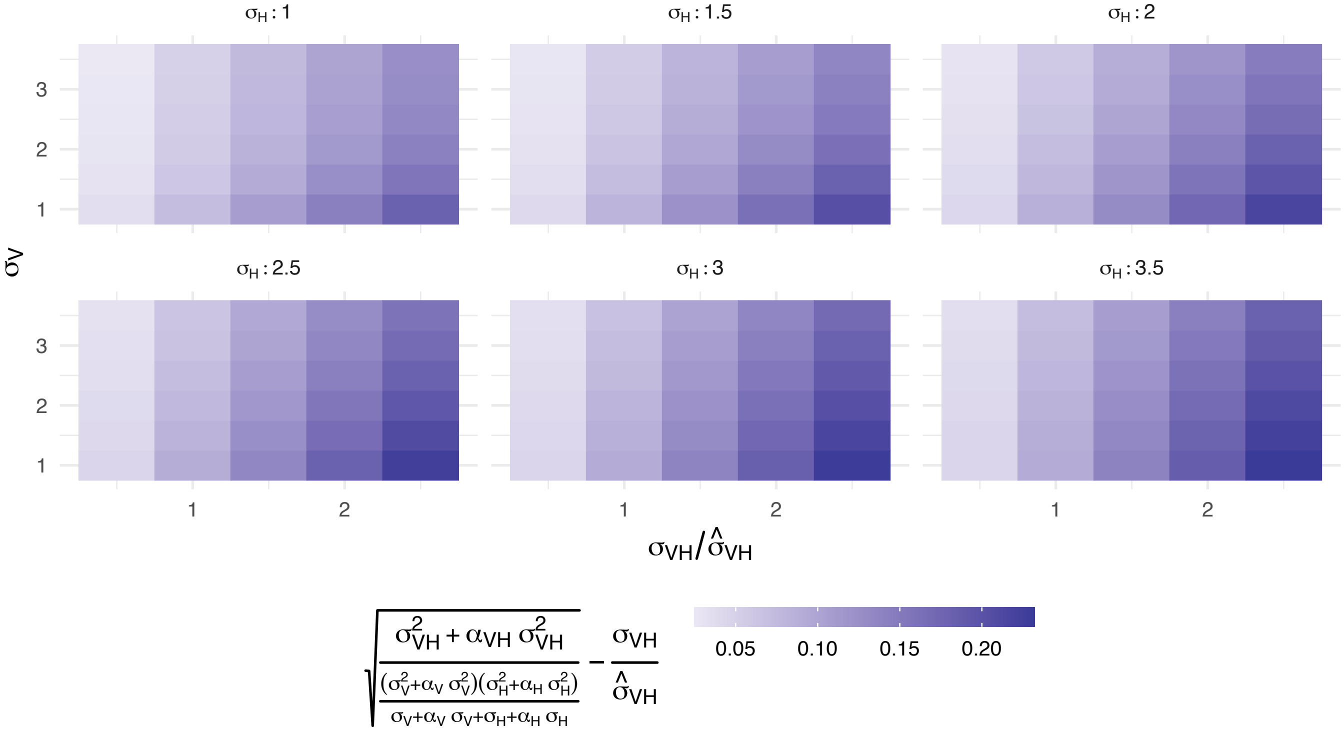

Figure 6: Effect of uncertainty in the representation of the standard stimulus on the integration index. The error in the integration index introduced by not accounting for the noisiness of the standard stimulus is shown as a function of the integration index as used in the manuscript (x-axis), and the standard deviation of the visual (y-axis) and haptic (panels) noise. The data were generated assuming that the visual and haptic standard are represented with an uncertainty that amounts to 40% and 50% of the measurement noise while the visual-haptic standard is at 60% of the visual-haptic measurement noise.

S6: Bayesian Statistics - Markov Chain Monte Carlo Approximation of the Posterior of the Correlation Coefficient

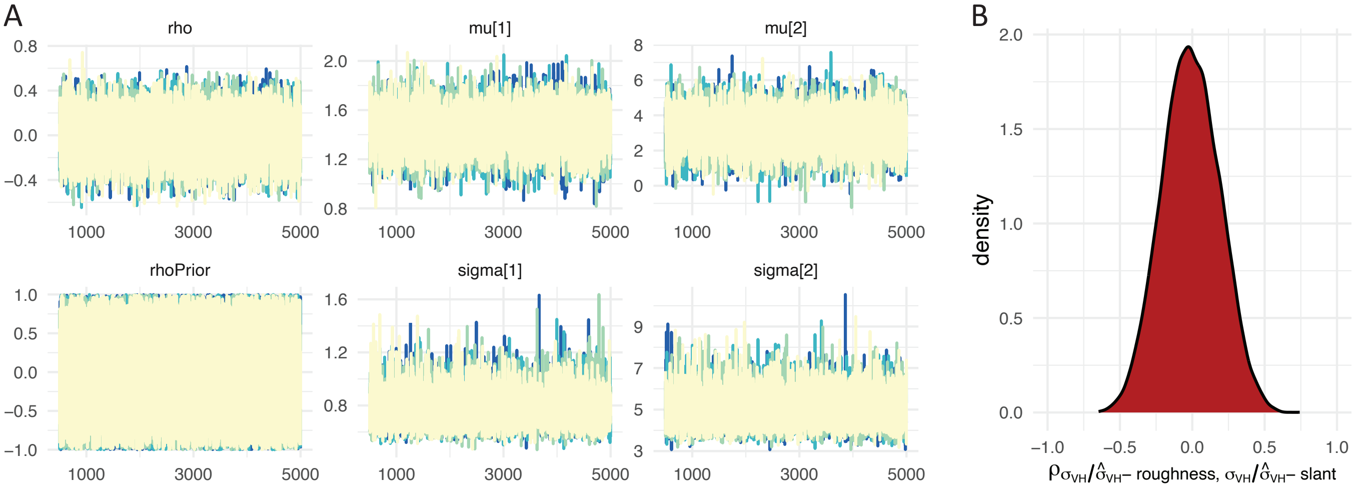

Figure 7: (A) MCMC chains for all parameters of the bivariate Gaussian distribution used to approximate the posterior of the correlation coefficient of the correlation between our participants' integration indices for slant and roughness. Note that rhoPrior was initialized in the same way as rho the variable of the correlation coefficient $\rho_{\sigma_{vh}/\sigma_{\hat{vh}}-r,\sigma_{vh}/\sigma_{\hat{vh}}-s}$ but rhoPrior was not constrained by the data and used to derive the Bayes factors for the point hypothesis $H0 : \rho = \rho_0$. All chains converged without evident problems. (B) Posterior distribution of the correlation coefficient $\rho_{\sigma_{vh}/\sigma_{\hat{vh}}-r,\sigma_{vh}/\sigma_{\hat{vh}}-s}$.

S7: Feature-Specific and Object-Based Causal-Inference Models of Visual-Haptic Integration

An observer might see and touch the same or two different surfaces, i.e., visual and haptic sensory signals can either originate from a single source ($C = 1$) or two sources ($C = 2$) (Körding et al., 2007; Sato, Toyoizumi, & Aihara, 2007). The prior probability of either scenario, $P(C = 1) = p_{C1}$ and $P(C = 2) = 1 - p_{C1}$, might be independent of the task or (in the observer's mind) differ between tasks, i.e., $P(C_{\text{r(oughness)}} = 1) = p_{C1_r}$ might differ from $P(C_{\text{s(lant)}} = 1) = p_{C1_s}$. We assumed that an observer performing object-based causal inference expects that the probability of seeing and touching the same surface is the same across tasks, whereas an observer performing feature-specific causal inference could expect that the probability differs across tasks. We further assumed that the observer understands the prior probability to encounter a surface with a specific roughness $s_{vh,r}$ or slant $s_{vh,s}$ to be the same across all levels of roughness and slant presented in the experiment, $s_{v,r}, s_{v,s}, s_{h,r}, s_{h,s} \sim \mathcal{N}(0, \sigma = 1000)$.

In a single trial, the observer only has access to a noisy measurement of the stimulus feature, e.g., $m_{v,r}$. We assumed the noise to be Gaussian-distributed and biased, for example, an observer might see the surfaces as rougher than they are, $m_{v,r} \sim \mathcal{N}(s'_{v,r}, \sigma_{v,r})$, where $s'_{v,r} = s_{v,r} + \Delta_{v,r}$. The observer has access to the level of noise associated with each modality and object feature (Ma, Beck, Latham, & Pouget, 2006), but not to the biases.

In every trial of the experiment, the observer compares the roughness or slant of a test stimulus against a memorized standard, which requires the observer to form an estimate of the test stimulus' feature. If the observer knew the visual and haptic signals shared a common source, the optimal estimate, $\hat{s}_{vh,C=1}$, would be a combination of the measurements and the mean of the prior, each weighted by its respective reliability (Landy, Maloney, Johnston, & Young, 1995; Yuille & Bülthoff, 1996; Ernst & Banks, 2002). This estimate is optimal in the sense that it minimizes squared localization error and maximizes the posterior probability of the estimate:

$$\hat{s}_{vh,r,C=1} = \frac{m_{v,r}\sigma_{v,r}^{-2} + m_{h,r}\sigma_{h,r}^{-2}}{\sigma_{v,r}^{-2} + \sigma_{h,r}^{-2} + 1000^{-2}}. \tag{1}$$

However, when the visual and haptic signals originate from different surfaces ($C = 2$), the estimate of the object feature, e.g., the estimates of roughness $\hat{s}_{v,r,C=2}$ and $\hat{s}_{h,r,C=2}$ should be independent of the haptic measurement and vice versa:

$$\hat{s}_{v,r,C=2} = \frac{m_{v,r}\sigma_{v,r}^{-2}}{\sigma_{v,r}^{-2} + 1000^{-2}} \text{ and } \hat{s}_{h,r,C=2} = \frac{m_{h,r}\sigma_{h,r}^{-2}}{\sigma_{h,r}^{-2} + 1000^{-2}}. \tag{2}$$

We assume that each observer has a favorite modality, vision or haptics, which they rely on if visual and haptic signals are present but do not share a common source. We further assume, in accordance with previous results (Körding et al., 2007; Badde, Navarro, & Landy, n.d.) that the observer derives an estimate of the stimulus feature by averaging the estimates based on the common-cause and separate-causes scenario, each weighted by the posterior probability of the underlying scenario. For the feature-specific model, this probability only depends on the measurements for that specific feature. Thus, for an observer who falls back on vision given separate sources and judges roughness we have

$$\hat{s}_{r,\text{feature-specific}} = P(C = 1|m_{v,r}, m_{h,r})\hat{s}_{vh,r,C=1} + P(C = 2|m_{v,r}, m_{h,r})\hat{s}_{v,r,C=2}. \tag{3}$$

Bayes' rule and the generative model yield

$$P(C = 1|m_{v,r}, m_{h,r}) = \frac{P(m_{v,r}, m_{h,r}|C = 1)p_{C1_r}}{P(m_{v,r}, m_{h,r}|C = 1)p_{C1_r} + P(m_{v,r}, m_{h,r}|C = 2)(1 - p_{C1_r})}, \tag{4}$$

where

$$\begin{aligned} &P(m_{v,r}, m_{h,r}|C = 1) \\ &\quad = \int P(m_{v,r}, m_{h,r}|s'_{vh,r})P(s'_{vh,r})\, ds'_{vh,r} \\ &\quad = \int P(m_{v,r}|s'_{vh,r})P(m_{h,r}|s'_{vh,r})P(s'_{vh,r})\, ds'_{vh,r} \\ &\quad = \frac{\exp\left(-\frac{1}{2}\frac{(m_{v,r}-m_{h,r})^2 1000^2 + m_{v,r}^2\sigma_{h,r}^2 + m_{h,r}^2\sigma_{v,r}^2}{\sigma_{v,r}^2\sigma_{h,r}^2 + \sigma_{v,r}^2 1000^2 + \sigma_{h,r}^2 1000^2}\right)}{2\pi\sqrt{\sigma_{v,r}^2\sigma_{h,r}^2 + \sigma_{v,r}^2 1000^2 + \sigma_{h,r}^2 1000^2}} \end{aligned} \tag{5}$$

and

$$\begin{aligned} &P(m_{v,r}, m_{h,r}|C = 2) \\ &\quad = \iint P(m_{v,r}, m_{h,r}|s'_{v,r}, s'_{h,r})P(s'_{v,r}, s'_{h,r})\, ds'_{v,r}\, ds'_{h,r} \\ &\quad = \int P(m_{v,r}|s'_{v,r})P(s'_{v,r})\, ds'_{v,r} \int P(m_{h,r}|s'_{h,r})P(s'_{h,r})\, ds'_{h,r} \\ &\quad = \frac{\exp\left(-\frac{1}{2}\left(\frac{m_{v,r}^2}{\sigma_{v,r}^2+1000^2} + \frac{m_{h,r}^2}{\sigma_{h,r}^2+1000^2}\right)\right)}{2\pi\sqrt{(\sigma_{v,r}^2 + 1000^2)(\sigma_{h,r}^2 + 1000^2)}}. \end{aligned} \tag{6}$$

In contrast, if causal inference proceeds at the object level, all available information should be taken into account. Thus,

$$\hat{s}_{r,\text{object-based}} = P(C=1|m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s})\hat{s}_{vh,r,C=1} + P(C=2|m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s})\hat{s}_{v,r,C=2} \tag{7}$$

Applying Bayes' rule yields

$$P(C=1|m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}) = \frac{P(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}|C=1)p_{C1}}{P(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}|C=1)p_{C1} + P(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}|C=2)(1-p_{C1})}, \tag{8}$$

where, given that surface roughness and slant are independent,

$$\begin{aligned}
&P(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}|C=1)\\
&= \iint P(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}|s'_{vh,r}, s'_{vh,s})P(s'_{vh,r}, s'_{vh,s})\, ds'_{vh,r}\, ds'_{vh,s}\\
&= \int P(m_{v,r}|s'_{vh,r})P(m_{h,r}|s'_{vh,r})P(s'_{vh,r})\, ds'_{vh,r} \int P(m_{v,s}|s'_{vh,s})P(m_{h,s}|s'_{vh,s})P(s'_{vh,s})\, ds'_{vh,s}\\
&= \frac{\exp\left(-\frac{1}{2}\left(\frac{(m_{v,r}-m_{h,r})^2 1000^2 + m_{v,r}^2\sigma_{h,r}^2 + m_{h,r}^2\sigma_{v,r}^2}{\sigma_{v,r}^2\sigma_{h,r}^2 + \sigma_{v,r}^2 1000^2 + \sigma_{h,r}^2 1000^2}\right) + \left(\frac{(m_{v,s}-m_{h,s})^2 1000^2 + m_{v,s}^2\sigma_{h,s}^2 + m_{h,s}^2\sigma_{v,s}^2}{\sigma_{v,s}^2\sigma_{h,s}^2 + \sigma_{v,s}^2 1000^2 + \sigma_{h,s}^2 1000^2}\right)\right)}{4\pi^2\sqrt{\sigma_{v,r}^2\sigma_{h,r}^2 + \sigma_{v,r}^2 1000^2 + \sigma_{h,r}^2 1000^2 + \sigma_{v,s}^2\sigma_{h,s}^2 + \sigma_{v,s}^2 1000^2 + \sigma_{h,s}^2 1000^2}}
\end{aligned} \tag{9}$$

and

$$\begin{aligned}
&P(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}|C=2)\\
&= \iint P(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s}|s'_{v,r}, s'_{h,r}, s'_{v,s}, s'_{h,s})P(s'_{v,r}, s'_{h,r}, s'_{v,s}, s'_{h,s})\, ds'_{v,r}\, ds'_{h,r}\, ds'_{v,s}\, ds'_{h,s}\\
&= \int P(m_{v,r}|s'_{v,r})P(s'_{v,r})\, ds'_{v,r} \int P(m_{h,r}|s'_{h,r})P(s'_{h,r})\, ds'_{h,r} \int P(m_{v,s}|s'_{v,s})P(s'_{v,s})\, ds'_{v,s} \int P(m_{h,s}|s'_{h,s})P(s'_{h,s})\, ds'_{h,s}\\
&= \frac{\exp\left(-\frac{1}{2}\left(\frac{m_{v,r}^2}{\sigma_{v,r}^2+1000^2} + \frac{m_{h,s}^2}{\sigma_{h,s}^2+1000^2} + \frac{m_{v,s}^2}{\sigma_{v,s}^2+1000^2} + \frac{m_{h,s}^2}{\sigma_{h,s}^2+1000^2}\right)\right)}{2\pi\sqrt{(\sigma_{v,r}^2+1000^2)(\sigma_{h,r}^2+1000^2)(\sigma_{v,s}^2+1000^2)(\sigma_{h,s}^2+1000^2)}}.
\end{aligned} \tag{10}$$

These estimates are conditional on the noisy sensory measurements $(m_{v,r}, m_{h,r}, m_{v,s}, m_{h,s})$ for a single trial. To simulate an observer's performance in our experiment, we randomly drew 20 measurements for each level of each feature and each modality and used these equations to determine the simulated observer's estimates of roughness and slant in each of the simulated trials. The observer's response was generated by comparing the estimate to the internal representation of the standard stimulus.

S8: Model Simulation - Representative Samples

The causal-inference models described were used to simulate our experiment under each hypothesis. The resulting simulation data were analyzed in the same way as our original data (see Methods). This was done to check whether our sample size and trial numbers were sufficient to measure the correlation between the integration indices for slant and roughness predicted by object-based multisensory causal inference. Below we show randomly selected samples of integration indices generated by each model.

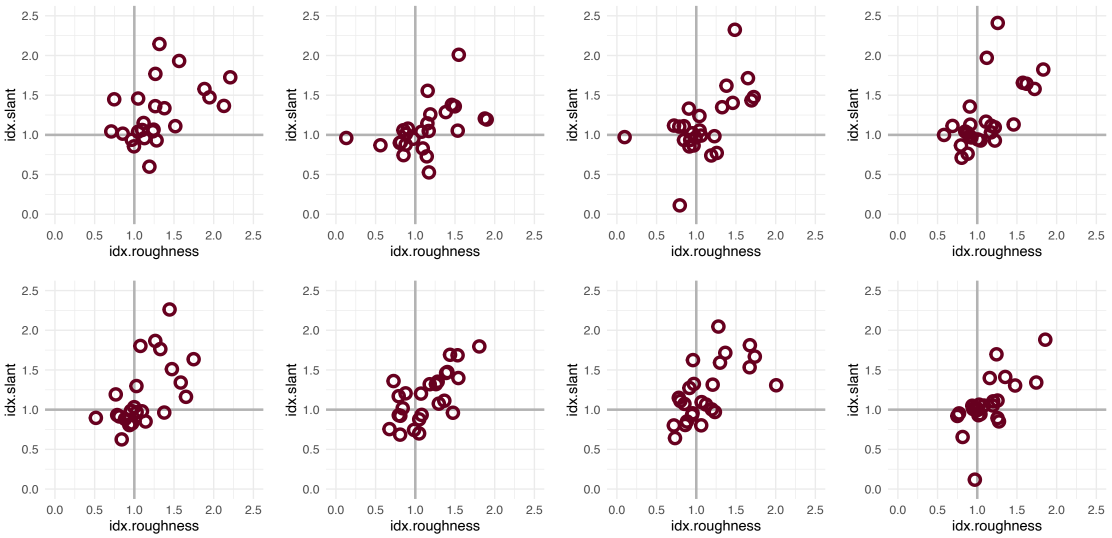

Figure 8: Integration indices for samples generated with the object-based causal inference model.

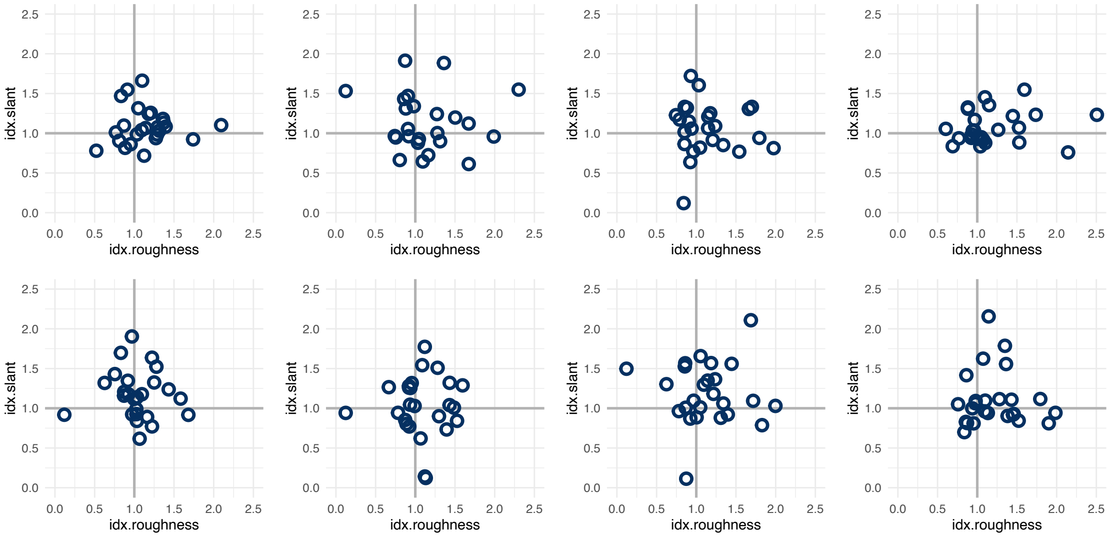

Figure 9: Integration indices for samples generated with the feature-specific causal inference model.

S9: Model Simulation - Incorrect Likelihoods

Less-than-maximal integration effects, that is, less variance reduction in cross-modal conditions than predicted by optimal cue integration, are sometimes interpreted as evidence that observers use an incorrect estimate of the sensory noise. The prevalence of such sub-optimal behavior remains unclear as older studies typically assume that the posterior probability of a common cause equals 1, i.e., they assume optimal cue integration without any causal inference. Nevertheless, we re-ran our simulations while accounting for this possibility. Specifically, we randomly selected for each observer, modality, and feature whether the standard deviation of the likelihood ($\sigma_v, \sigma_h, \sigma_{vh}$ for both features) was 1) identical, 2) 20% larger, or 3) 20% smaller than the standard deviation of the measurement function used to generate Monte Carlo samples. Though the simulated correlation coefficients were slightly lower than assuming accurate estimates of sensory uncertainty, the object-based model still reliably predicted that the integration indices for both features correlate with each other and thus is still at odds with the observed data.

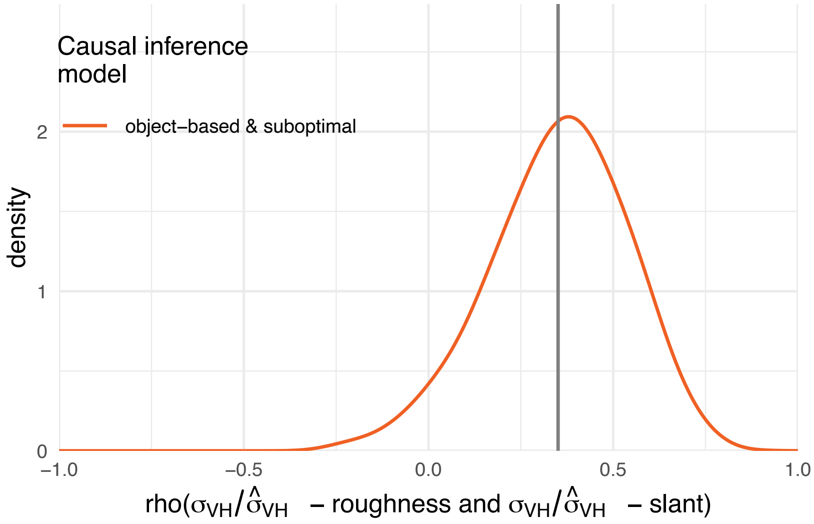

Figure 10: Distribution of simulated correlation coefficients between the integration indices for roughness and slant assuming object-based multisensory causal inference and misestimation of an observer's sensory uncertainty (33.3% chance that a sensory uncertainty was overestimated by 20%, 33.3% chance that a sensory uncertainty was underestimated by 20%). Correlation coefficients are based on 10,000 simulated datasets of the same size as the original data (26 participants, 20 trials per condition). Vertical lines indicate distribution means.

References

Badde, S., Navarro, K. T., & Landy, M. S. (n.d.). Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. , *197*, 104170.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. Retrieved from `http://dx.doi.org/10.1038/415429a`

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007, September). Causal inference in multisensory perception. *PloS one*, *2*, e943.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res*, *35*(3), 389–412.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006, Nov). Bayesian inference with probabilistic population codes. *Nat Neurosci*, *9*(11), 1432–1438. Retrieved from `http://dx.doi.org/10.1038/nn1790`

Sato, Y., Toyoizumi, T., & Aihara, K. (2007, December). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural computation*, *19*, 3335–3355.

Yuille, A., & Bülthoff, H. (1996). Bayesian theory and psychophysics. In D. Knill & W. Richards (Eds.), *Perception as bayesian inference* (p. 123-161). Cambridge University Press.