PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

**Cite this article:** Badde S, Landy MS, Adams WJ. 2023 Multisensory causal inference is feature-specific, not object-based. *Phil. Trans. R. Soc. B* **378**: 20220345.
https://doi.org/10.1098/rstb.2022.0345

Received: 15 January 2023
Accepted: 18 June 2023

One contribution of 16 to a theme issue 'Decision and control processes in multisensory perception'.

**Subject Areas:**
cognition

**Keywords:**
visual–haptic, roughness, slant, cue integration, causal inference

**Author for correspondence:**
Stephanie Badde
e-mail: stephanie.badde@tufts.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6729986.

THE ROYAL SOCIETY
PUBLISHING

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

Stephanie Badde[1], Michael S. Landy[2] and Wendy J. Adams[3]

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

SB, 0000-0002-4005-5503; MSL, 0000-0002-2079-4552; WJA, 0000-0002-5832-1056

Multisensory integration depends on causal inference about the sensory signals. We tested whether implicit causal-inference judgements 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. Modelling of participants' responses revealed that the degree to which their perceptual judgements were based on integrated visual–haptic information varied unsystematically across features. 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 judgements 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.

This article is part of the theme issue 'Decision and control processes in multisensory perception'.

## 1. 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 [1–3]. 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 [4–6].

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 [5,7]. As another example, auditory frequency information interferes with tactile frequency perception but only across similar frequencies [8]. 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 [9–11]. For example, perceptual judgements about visual and haptic stimuli with a large spatial [12] or temporal [13] offset show no integration effects. Instead, given a large spatio-temporal conflict, observers base their perceptual judgements 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

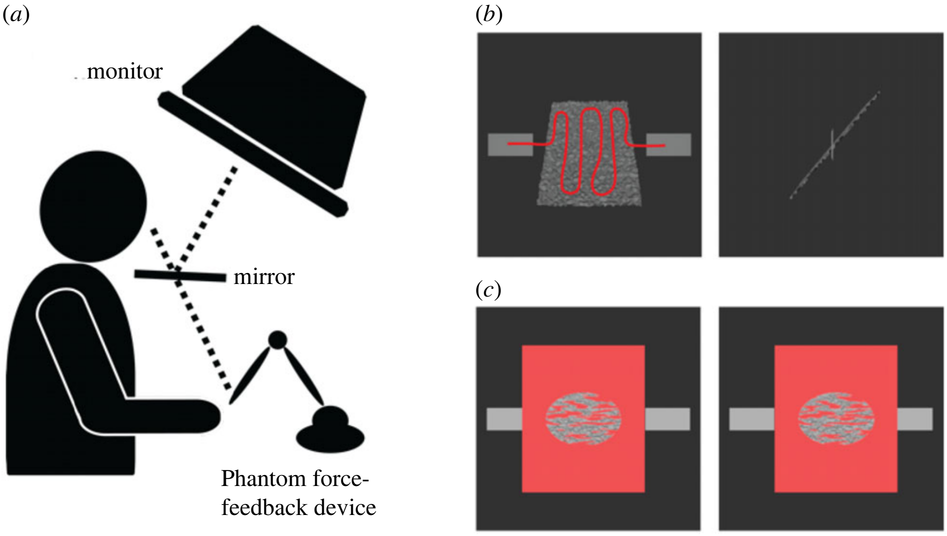

**Figure 1.** Set-up and stimuli. (*a*) Participants viewed stereoscopically presented visual stimuli via a mirror so that they were perceived as co-located with virtual haptic stimuli rendered using a Phantom force-feedback device. (*b*) The stimuli were rough surfaces, slanted top–back from fronto-parallel. 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 centre 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). (Online version in colour.)

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 [14,15]. Such modulations of cross-modal integration across participants are naturally accounted for by Bayesian causal-inference models [5,16]. 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 that 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 [17], with the observer's previous experiences [18,19] as well as their attentional state [20]. In turn, the likelihood of a common cause is driven by sensory information about the stimuli. In some observers, 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 presented on the arm [20]. Such perceptual cross-modal misalignment reduces the inferred 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 [21]. In addition to reduced common cause priors and likelihoods, 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 in a statistical optimal fashion, see electronic supplementary material, 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, according to the object-based account, an observer who assigns a low *a priori* probability to the common cause scenario should do so independently of the task. And an observer's perceptual biases would affect the likelihood of a common cause even if the biased feature is currently irrelevant. By 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 (figure 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

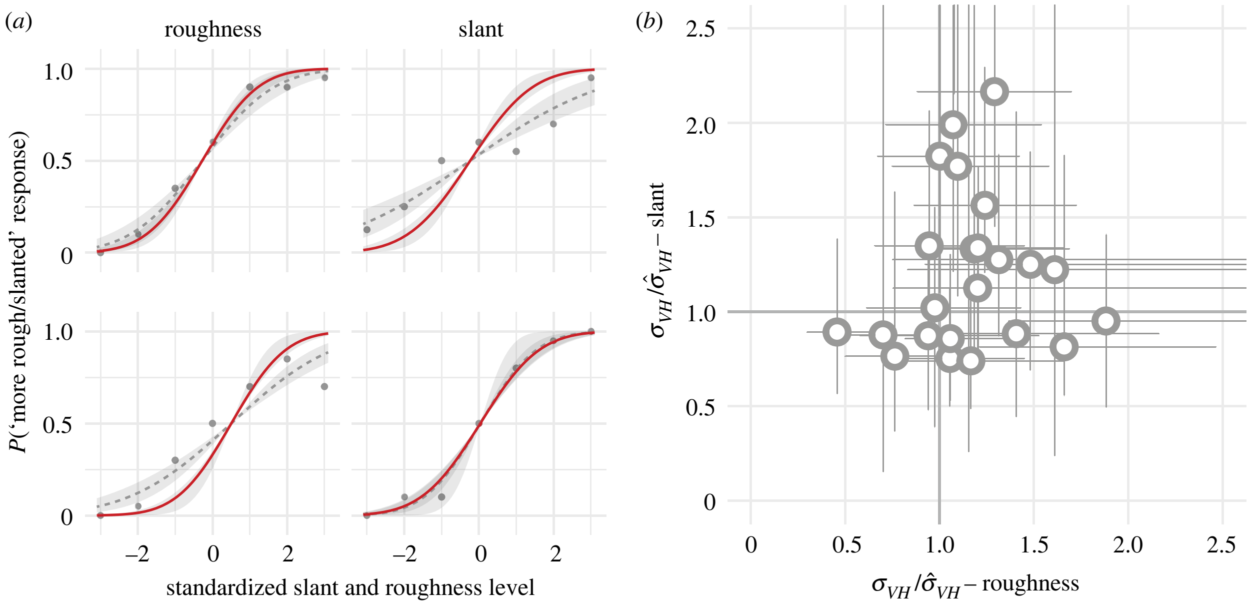

**Figure 2.** Results. (*a*) Psychometric curves for two participants (one per row) in the visual–haptic condition of the roughness (left column) and slant (right column) tasks. Markers indicate the observed proportion of 'more rough / more slanted than the standard' responses for each feature level shown on a common scale for roughness and slant. Grey dashed 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 electronic supplementary material, 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 electronic supplementary material, S3 for an alternative index). An index of one indicates maximal integration effects, while larger values suggest less-than-maximal integration effects, indicating perceptual judgements that are partially based on unimodal information. Error bars indicate 95% confidence intervals obtained by bootstrapping the raw data. (Online version in colour.)

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. By 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.

## 2. 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 for visual and haptic roughness, whereas other participants showed the opposite pattern (figure 2*a*; see electronic supplementary material, S1 for all participants' psychometric curves and electronic supplementary material, S2 for the extracted uncertainty).

We calculated an integration index for each feature and participant (figure 2*b*). 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 [22,23]. 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 electronic supplementary material, 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 centred at $r = -0.01$ and we obtained a Bayes factor of 3.7 favouring a correlation coefficient equal to zero, as predicted by simulations with a model performing feature-specific causal inference (figure 3*a*). By 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.

## 3. Discussion

We tested whether multisensory causal inference is object-based or selectively refers to task-relevant features. To this

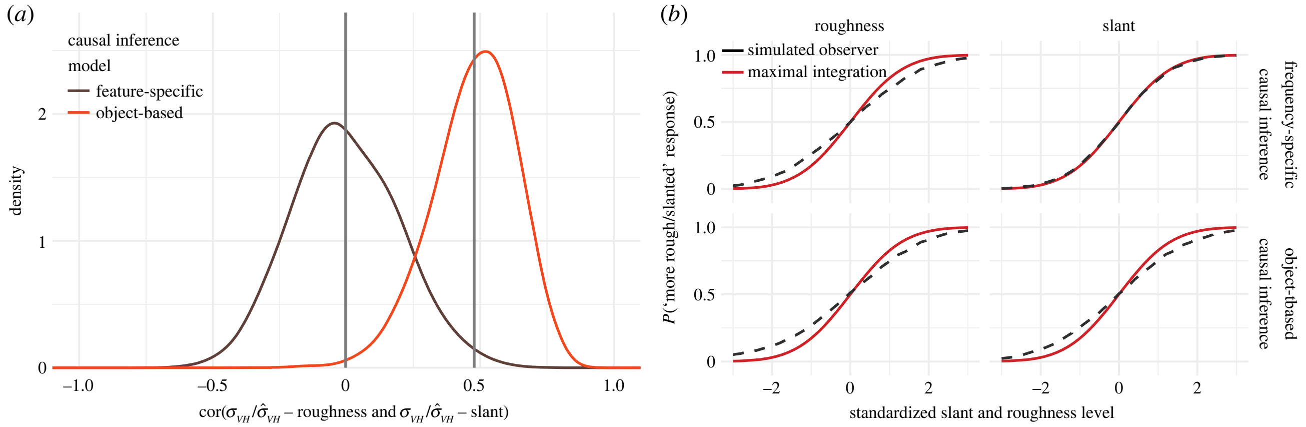

**Figure 3.** Model predictions. (*a*) Distribution of simulated correlation coefficients between the integration indices for roughness and slant. Correlation coefficients are based on 10 000 simulated datasets of the same size as the original data (26 participants, 20 trials per condition). Data were generated using the feature-specific (brown) and object-based (orange) causal-inference models (see Methods, §4f). 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 (solid 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. (Online version in colour.)

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 [5,7,24]. 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 [25]. However, results are mixed regarding the influence of semantic congruency on multisensory integration. Some studies find that congruency between the images and sounds of everyday objects [26] or the gender or emotion of a speaker's face and voice [27,28] facilitates multisensory integration, whereas other studies report no effects of additional high-level cues towards a common cause of visual and auditory signals [29,30]. Furthermore, semantic congruency refers to the category of the presented object rather than cross-modal congruency of sensory feature information [4]. 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 favour 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 (figure 3*a*). 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 favourite modality, the modality they would choose if visual and haptic signals were from different sources. Analogous to previous implementations of Bayesian multisensory causal inference [5,19,20,31], 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 judgements 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 on 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 (figure 3*b*; see §4 and electronic supplementary material, S7 for details). By 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 (figure 3*a*). Based on our data, we can accept the former and reject the latter correlation coefficient and with it the 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.

# 4. Methods

## (a) 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.

## (b) 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 (figure 1*a*). 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 (figure 1*c*) were presented to the left and right eyes using active shutter glasses (Stereographics Crystal Eyes).

The virtual stimuli were textured rectangles (20 cm wide, 30 cm high; figure 1*b*), which were slanted top–back from fronto-parallel (defined with respect to the visual plane, figure 1*a*) by 26–38°. To create a rough plane, first a two-dimensional grid of 400×600 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 three-dimensional 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 (figure 1*b*). 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 (figure 1*c*). 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.

## (c) Task and design

Participants compared the roughness or slant of a visual, haptic or visual–haptic test stimulus with 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° 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°; test stimuli in the slant experiment were slanted top–back by 38, 36, 34, 32, 30, 28 or 26° and had a roughness of 4.5 mm.

## (d) 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 (figure 1*b*, left panel). In visual trials, the textured plane was not haptically rendered and 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 centre of the occluder (figure 1*c*) 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 judgements. 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 h to complete. Testing was spread across several sessions.

## (e) 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 $\Phi$, with a lapse rate $\lambda$, were fit to the proportion of more rough / more top–back responses as a function of stimulus level $s$, $p(s) = \lambda/2 + (1-\lambda)\Phi(s;\, \mu,\sigma^2)$ using maximum likelihood. (We did not fix $\mu$ 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 electronic supplementary material, 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). Ninety-five per cent 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, $\widehat{\sigma^2_{vh}} = \sigma^2_v\sigma^2_h/(\sigma^2_v + \sigma^2_h)$ [22,23]. We quantified the degree to which participants relied on visual–haptic integration by calculating the ratio of the estimated and predicted visual–haptic variances, $\sigma^2_{vh}/\widehat{\sigma^2_{vh}}$. 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 electronic supplementary material, 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 (electronic supplementary material, S5).

The two alternative accounts of multisensory causal inference (object-based and feature-specific) make specific 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 covariance 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 [32] leapfrog algorithm (see electronic supplementary material, S6 for details). Bayes factors for point hypotheses $H0{:}\rho = \rho_0$ and $H1{:}\rho \neq \rho_0$ were calculated based on the ratio of the posterior and prior densities at $\rho_0$ [33]. 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 analysed in the same way as the original data and the correlation between the integration indices for roughness and slant was stored (see electronic supplementary material, 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:10.17605/OSF.IO/EVKBF).

## (f) 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 electronic supplementary material, 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 ‘favourite’ modality. Their ‘favourite’ 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. $\hat{x}_i = P(C = 1|m_{v,i}, m_{h,i})\, \hat{x}_{vh,i,C=1} + P(C = 2|m_{v,i}, m_{h,i})\hat{x}_{v\text{ or }h,i,C=2}$. This is equivalent to model averaging in Bayesian causal inference [5], with the only difference being that typically one of the modalities is the ‘favourite’ 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 $\sigma^2_v$ and $\sigma^2_h$. We additionally allowed the sensory signals to be biased [19,20,31], as modality-specific biases in the sensory signals are a root cause of reduced cross-modal integration effects [21].

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, $m_{v,\text{s},\text{i}},m_{h,\text{s},\text{i}},m_{v,\text{r},\text{i}},m_{h,\text{r},\text{i}}$) and a task-independent prior probability that visual and haptic signals originate from a common cause $p_{\text{c}=1}$,

$$P(C_{\text{s(lant) or r(oughness)}}=1|m_{v,s,i},m_{h,s,i},m_{v,r,i},m_{h,r,i})=\frac{p_{C=1}P(m_{v,s,i},m_{h,s,i},m_{v,r,i},m_{h,r,i}|C=1)}{p_{C=1}P(m_{v,s,i},m_{h,s,i},m_{v,r,i},m_{h,r,i}|C=1)+(1-p_{C=1})P(m_{v,s,i},m_{h,s,i},m_{v,r,i},m_{h,r,i}|C=2)}$$

(see electronic supplementary material, S7 for a full specification of the likelihoods). Thus, the posterior probability of a common cause is derived identically across tasks. By 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 $i$, $m_{v,\text{s},\text{i}},m_{h,\text{s},\text{i}}$) and allow for separate common-cause priors, one for slant, $p_{\text{c}=1,\text{s}}$, and one for roughness, $p_{\text{c}=1,\text{r}}$. Thus, the posterior probability for a common cause in the slant task,

$$P(C_{\text{s(lant)}}=1|m_{v,s,i},m_{h,s,i})=\frac{p_{C=1,s}P(m_{v,s,i},m_{h,s,i}|C=1)}{p_{C=1,s}P(m_{v,s,i},m_{h,s,i}|C=1)+(1-p_{C=1,s})P(m_{v,s,i},m_{h,s,i}|C=2)},$$

differs from the one in the roughness task

$$P(C_{\text{r(oughness)}}=1|m_{v,r,i},m_{h,r,i})=\frac{p_{C=1,r}P(m_{v,r,i},m_{h,r,i}|C=1)}{p_{C=1,r}P(m_{v,r,i},m_{h,r,i}|C=1)+(1-p_{C=1,r})P(m_{v,r,i},m_{h,r,i}|C=2)}.$$

Ethics. This work was approved by the institutional review board of the University of Southampton.

Data accessibility. Data and code are publicly available from OSF: https://osf.io/10.0.68.197/EVKBF.

The data are provided in electronic supplementary material [34].

Authors' contributions. S.B.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, visualization, writing—original draft, writing—review and editing; M.S.L.: conceptualization, funding acquisition, methodology, supervision, writing—review and editing; W.J.A.: conceptualization, methodology, resources, supervision, writing—review and editing.

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

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

Funding. This work was supported by the National Institutes of Health, grant no. NIH EY08266 to M.S.L. and the German Research Foundation, grant no. BA 5600 1/1 to S.B.

Acknowledgements. We thank Walter Wang and Davide Deflorio for help with data acquisition and Erich Graf for technical and logistical support.

## References

1. Ernst MO, Bülthoff HH. 2004 Merging the senses into a robust percept. *Trends Cogn. Sci.* **8**, 162–169. (doi:10.1016/j.tics.2004.02.002)
2. Fetsch CR, DeAngelis GC, Angelaki DE. 2013 Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nat. Rev. Neurosci.* **14**, 429–442. (doi:10.1038/nrn3503)
3. Trommershauser J, Körding KP, Landy MS. 2011 *Sensory cue integration*. Oxford, UK: Oxford University Press.
4. Chen Y-C, Spence C. 2017 Assessing the role of the 'unity assumption' on multisensory integration: a review. *Front. Psychol.* **8**, 445. (doi:10.3389/fpsyg.2017.00445)
5. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. 2007 Causal inference in multisensory perception. *PLoS ONE* **2**, e943. (doi:10.1371/journal.pone.0000943)
6. Welch RB, Warren DH. 1980 Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* **88**, 638–667. (doi:10.1037/0033-2909.88.3.638)
7. Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA. 2004 Unifying multisensory signals across time and space. *Exp. Brain Res.* **158**, 252–258. (doi:10.1007/s00221-004-1899-9)
8. Yau JM, Olenczak JB, Dammann JF, Bensmaia SJ. 2009 Temporal frequency channels are linked across audition and touch. *Curr. Biol.* **19**, 561–566. (doi:10.1016/j.cub.2009.02.013)
9. Alais D, Newell F, Mamassian P. 2010 Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving* **23**, 3–38. (doi:10.1163/187847510X488603)
10. Calvert GA, Spence C, Stein BE. 2004 *The handbook of multisensory processes*. New York, NY: MIT Press.
11. Murray MM, Wallace MT. (eds) 2012 *The neural bases of multisensory processes*. New York, NY: CRC Press/Taylor & Francis. See http://www.ncbi.nlm.nih.gov/books/NBK92848/.
12. Gepshtein S, Burge J, Ernst MO, Banks MS. 2005 The combination of vision and touch depends on spatial proximity. *J. Vis.* **5**, 7. (doi:10.1167/5.11.7)
13. Parise CV, Ernst MO. 2016 Correlation detection as a general mechanism for multisensory integration. *Nat. Commun.* **7**, 11543. (doi:10.1038/ncomms11543)
14. Battaglia PW, Jacobs RA, Aslin RN. 2003 Bayesian integration of visual and auditory signals for spatial localization. *J. Opt. Soc. Am. A* **20**, 1391–1397. (doi:10.1364/JOSAA.20.001391)
15. 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. (doi:10.1016/j.cortex.2019.03.026)
16. Sato Y, Toyoizumi T, Aihara K. 2007 Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.* **19**, 3335–3355. (doi:10.1162/neco.2007.19.12.3335)
17. Odegaard B, Shams L. 2016 The brain's tendency to bind audiovisual signals is stable but not general. *Psychol. Sci.* **27**, 583–591. (doi:10.1177/0956797616628860)
18. Gau R, Noppeney U. 2016 How prior expectations shape multisensory perception. *Neuroimage* **124**, 876–886. (doi:10.1016/j.neuroimage.2015.09.045)
19. Hong F, Badde S, Landy MS. 2022 Repeated exposure to either consistently spatiotemporally congruent or consistently incongruent audiovisual stimuli modulates the audiovisual common-cause prior. *Sci. Rep.* **12**, 15532. (doi:10.1038/s41598-022-19041-7)
20. 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. (doi:10.1016/j.cognition.2019.104170)
21. Negen J, Slater H, Bird L-A, Nardini M. 2022 Internal biases are linked to disrupted cue combination in children and adults. *J. Vis.* **22**, 14. (doi:10.1167/jov.22.12.14)
22. Ernst MO, Banks MS. 2002 Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433. (doi:10.1038/415429a)
23. Landy MS, Maloney LT, Johnston EB, Young M. 1995 Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res.* **35**, 389–412. (doi:10.1016/0042-6989(94)00176-M)
24. McGovern DP, Roudaia E, Newell FN, Roach NW. 2016 Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Sci. Rep.* **6**, 24673. (doi:10.1038/srep24673)
25. Doehrmann O, Naumer MJ. 2008 Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain*

*Res.* **1242**, 136–150. (doi:10.1016/j.brainres.2008.03.071)

26. Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ. 2008 The effect of prior visual information on recognition of speech and sounds. *Cereb. Cortex* **18**, 598–609. (doi:10.1093/cercor/bhm091)
27. Dolan RJ, Morris JS, de Gelder B. 2001 Crossmodal binding of fear in voice and face. *Proc. Natl Acad. Sci. USA* **98**, 10 006–10 010. (doi:10.1073/pnas.171288598)
28. Vatakis A, Spence C. 2007 Crossmodal binding: evaluating the 'unity assumption' using audiovisual speech stimuli. *Percept. Psychophys.* **69**, 744–756. (doi:10.3758/BF03193776)
29. Radeau M, Bertelson P. 1978 Cognitive factors and adaptation to auditory-visual discordance. *Percept. Psychophys.* **23**, 341–343. (doi:10.3758/BF03199719)
30. Vatakis A, Spence C. 2008 Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. *Acta Psychol.* **127**, 12–23. (doi:10.1016/j.actpsy.2006.12.002)
31. Hong F, Badde S, Landy MS. 2021 Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception. *PLoS Comput. Biol.* **17**, e1008877. (doi:10.1371/journal.pcbi.1008877)
32. Stan Development Team. 2022 *Stan modeling language users guide and reference manual, 2.31*. See https://mc-stan.org
33. Wagenmakers E-J, Lodewyckx T, Kuriyal H, Grasman R. 2010 Bayesian hypothesis testing for psychologists: a tutorial on the Savage–Dickey method. *Cognit. Psychol.* **60**, 158–189. (doi:10.1016/j.cogpsych.2009.12.001)
34. Badde S, Landy MS, Adams WJ. 2023 Multisensory causal inference is feature-specific, not object-based. Figshare. (doi:10.6084/m9.figshare.c.6729986)