

Bayesian causal inference in visuotactile integration in children and adults

Erik Verhaar¹  | Wijbrand Pieter Medendorp¹ | Sabine Hunnius¹ | Janny C. Stapel²

¹ Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

² Uppsala University, Uppsala, Sweden

Correspondence

Erik Verhaar, Donders Centre for Cognition, Thomas van Aquinostraat 4, 6525 GD Nijmegen, The Netherlands.

Email: erik.verhaar@donders.ru.nl

Funding information

Swedish Research Council, Grant/Award Number: 2016-01725

Abstract

If cues from different sensory modalities share the same cause, their information can be integrated to improve perceptual precision. While it is well established that adults exploit sensory redundancy by integrating cues in a Bayes optimal fashion, whether children under 8 years of age combine sensory information in a similar fashion is still under debate. If children differ from adults in the way they infer causality between cues, this may explain mixed findings on the development of cue integration in earlier studies. Here we investigated the role of causal inference in the development of cue integration, by means of a visuotactile localization task. Young children (6–8 years), older children (9.5–12.5 years) and adults had to localize a tactile stimulus, which was presented to the forearm simultaneously with a visual stimulus at either the same or a different location. In all age groups, responses were systematically biased toward the position of the visual stimulus, but relatively more so when the distance between the visual and tactile stimulus was small rather than large. This pattern of results was better captured by a Bayesian causal inference model than by alternative models of forced fusion or full segregation of the two stimuli. Our results suggest that already from a young age the brain implicitly infers the probability that a tactile and a visual cue share the same cause and uses this probability as a weighting factor in visuotactile localization.

KEYWORDS

Bayesian modeling, causal inference, child development, cue combination, multisensory integration, visuotactile integration

1 | INTRODUCTION

When we perceive the world around us, properties of the environment can often be sensed through more than one modality. For example, when holding an object, information about its size is received both haptically and visually. However, sensory input is noisy, which leads to uncertainty in the perceptual estimates. To reduce this perceptual

uncertainty, adults optimally integrate overlapping sensory cues. More specifically, the different cues are weighted according to their precision (i.e., the reciprocal of the variance), so that the precision of the integrated estimate is higher than those of estimates based on any of the cues in isolation (Ernst & Banks, 2002; van Beers et al., 1996, 1999).

Obviously, integration of two or more sensory cues should only occur if they come from the same cause. Because cues are inherently

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Developmental Science* published by John Wiley & Sons Ltd



noisy, this involves solving a causal inference problem (Kayser & Shams, 2015; Körding et al., 2007; Rohe & Noppeney, 2015). In statistical terms, this means that the brain must weigh the noisy cue information on the likelihood that the cues share a common cause. In this weighting, the brain incorporates its prior expectations on causality (through a *prior on common cause*), based on an accumulated history of comparable sensory circumstances (see Figure 1a,b).

Adults are known to rely on these computations, referred as Bayesian causal inference, in a wide range of perceptual tasks: from multisensory cue localization (Körding et al., 2007; Wozny et al., 2010) and perception of self-motion (Acerbi et al., 2018) to the integration of sensory prediction and sensory reafference in spatial perception (Atsma et al., 2016; Perdreaux et al., 2019). These studies show that multisensory perception in adults is Bayes optimal, not only by integrating cues to maximize multisensory precision, but also by conditioning this integration process on causality.

The developmental time course of the ability to integrate multisensory cues is still under debate. In some studies, only children older than 8 years were found to integrate sensory cues in a Bayes optimal fashion (Dekker et al., 2015; Gori et al., 2008; Nardini et al., 2008, 2010, 2013; Petrini et al., 2014; Scheller et al., 2021), whereas in other studies, younger children also appeared to be able to integrate, potentially as a result of differences in task and feedback (Negen et al., 2019; Rohlf et al., 2020), or due to training (Nava et al., 2020).

From the perspective of Bayesian causal inference, it is unclear whether the typical lack of sensory fusion in young children is related to the fusion process itself or to the process of detecting causality among multiple cues. In case of the latter, the lack of optimal integration in young children may be related to a lower prior on common cause in this age group compared to older children and adults: by applying a lower prior on common cause, cues are more likely to be segregated rather than integrated. Although a recent study argued that young children are able to infer causality in multisensory perception (Rohlf et al., 2020; see also Dekker & Lisi, 2020), their results leave room for the possibility of forced multisensory fusion with sub-optimal weights. To dissociate these explanations experimentally, multisensory perception studies should include multiple disparities between the cues.

Using a psychophysical approach, we tested young children (6- to 8-year-old), older children (9.5- to 12.5-year-old) and adults in a visuo-tactile localization task with different spatial disparities between the visual and tactile cues. They were mechanically stimulated on their forearm while a visual stimulus was simultaneously presented either at the same or at a different location (Figure 1a,b). If causal inference guides the integration of the visual and tactile input, the perceived location of the tactile stimulus should be biased towards the visual stimulus, and the relative magnitude of this bias should decrease with a larger distance between the two stimuli (Figure 1d). To explain the data of each participant, we fitted three causal inference models and three commonly suggested alternatives: full fusion, segregation and cue switching. Moreover, we used the best fitting causal inference model to obtain a prior on common cause for each participant. A smaller prior on common cause in the young children compared to older children and adults would confirm the hypothesis that young chil-

RESEARCH HIGHLIGHTS

- Presentation of a visual cue biases tactile localization in young children (6–8 years), as well as older children (9.5–12.5 years) and adults.
- The relative effect of a visual cue on tactile localization becomes smaller with increased spatial separation between the visual and tactile cues.
- Bayesian causal inference is a better model for visuotactile localization than either full fusion, segregation or switching.
- Children are more likely to combine visual and tactile cues than adults, expressed by a higher prior on common cause.

dren's causality assessment leads to a higher tendency to segregate rather than integrate cues, which would explain why other studies frequently found young children not to integrate.

2 | METHODS

2.1 | Participants

Twenty-one 6- to 8-year-old children ($M = 7.1$ years, $SD = 7$ months; six girls), 25 9.5- to 12.5-year-old children ($M = 11.0$ years, $SD = 8$ months; 14 girls), and 20 adults ($M = 24.7$ years, $SD = 3.8$ years; 14 females) were included in the study. One additional 6-year-old was tested but his data were not included in the analyses because he seemed to randomly press buttons and could not recall the task instructions upon questioning. The data from all other participants were included in the analyses. Of the included children in the youngest age group, six did not complete the full experiment. On average, those six children stopped after finishing two out of four blocks of trials. All other participants finished the whole experiment.

The children were recruited from and tested at a primary school in the Netherlands. Written informed consent was given by their caregivers. As a reward for participation, a science workshop was organized for the whole class. Additionally, children from the youngest age group were rewarded with a sticker for each of the four blocks they completed. Adults enrolled through a participant recruitment website; they signed an informed consent form prior to the experiment. They were rewarded with either a gift voucher (10 euro) or course credits for participation. All procedures were approved by the local Ethics Committee of the Faculty of Social Sciences under approval ECSW-2019-089.

2.2 | Apparatus and stimuli

Participants were seated on a chair behind a table. Tactile and visual stimuli were applied at one of five equidistant nodes that were

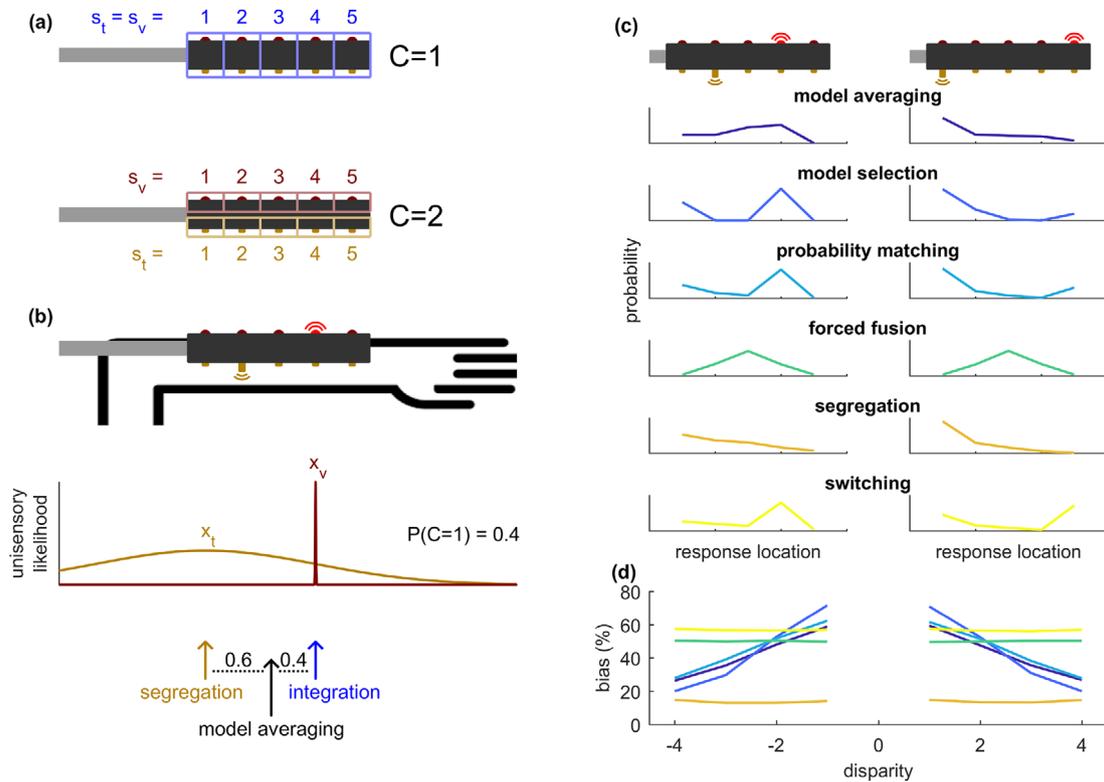


FIGURE 1 Causal inference (a-b) and model predictions (c-d). (a) Two causal models: common cause ($C = 1$) and separate causes ($C = 2$). Causal inference makes use of the generative model that stimuli are generated according to $C = 1$ with a probability of p_{common} (the prior on common cause), and according to $C = 2$ with a probability of $1 - p_{common}$. (c) On each trial, an observer has to infer what the likelihood is of each model, given the sensory input and the prior on common cause. This likelihood is termed the posterior on common cause, denoted by $P(C = 1)$. For *model averaging*, the final cue estimate is based on a weighing of the fully segregated and fully integrated sensory input, with a relative weight of $P(C = 1)$ for the segregated signal. (c) Predictions of various models of the response distributions on two example bimodal stimuli. The upper three models involve causal inference, whereas the bottom three do not. (d) Model predictions for the relation between spatial disparity (the distance between visual and tactile stimulus) and bias (the average shift of the tactile response towards the visual cue, as a percentage of this disparity). Causal inference models predict a negative relation between bias and absolute disparity, while for the alternative models, the percent bias is constant. Parameter values used in (c-d): $p_{common} = .7$, $\sigma_t = 1.5$, $\mu_p = 3$, $\sigma_p = 100$, $w_t = .5$, $p_t = .5$

embedded in a piece of fabric, attached in longitudinal direction on the dorsal side of the left forearm (see Figure 2). Each node contained a miniature electromagnetic solenoid-type stimulator, or “tactor” (Dancer Design, UK) on the inside (i.e., at the skin), and a red LED on the outside. The distance between the centers of adjacent nodes was 2.5 cm. The stimulation device was positioned midway the forearm, such that it felt stable and comfortable for the participant.

For adults, a head- and armrest was used to fix posture across trials (Figure 2b). A computer screen was mounted directly behind the armrest, and was used for presentation of a fixation cross and the response options. The fixation cross was positioned approximately 5 cm above the center node of the stimulus apparatus. Participants could provide their responses with a computer mouse. For the children, the left arm was placed on a piece of cardboard sheathed in fabric, which covered the top half of a tablet (Asus Eee Slate). Children were asked to look at their left arm when the stimuli were presented. Response options were presented on the bottom half of the tablet and responses were given using its touchscreen (Figure 2a).

Tactile stimuli consisted of two taps of 35 ms duration each, separated by a 300 ms interval. Visual stimuli consisted two LED flashes of 35 ms duration each, simultaneous with the tactile taps, at either the same or a different node. Registration of responses and controlling the sleeve was done using Presentation 20.0 (NeuroBehavioral Systems).

2.3 | Procedure

The experimental paradigm was an adapted version of the audiovisual localization task by Körding et al. (2007), but with visual and tactile stimuli. We developed the paradigm first in adults and subsequently tested it in children.

Before the start of the experiment, participants were familiarized with the stimuli by applying five congruent (i.e., spatially aligned) bimodal cues, one at each of the five nodes. Participants were told that, during the experiment, the visual and tactile stimuli could be either at the same or at different locations. For the young children, this was

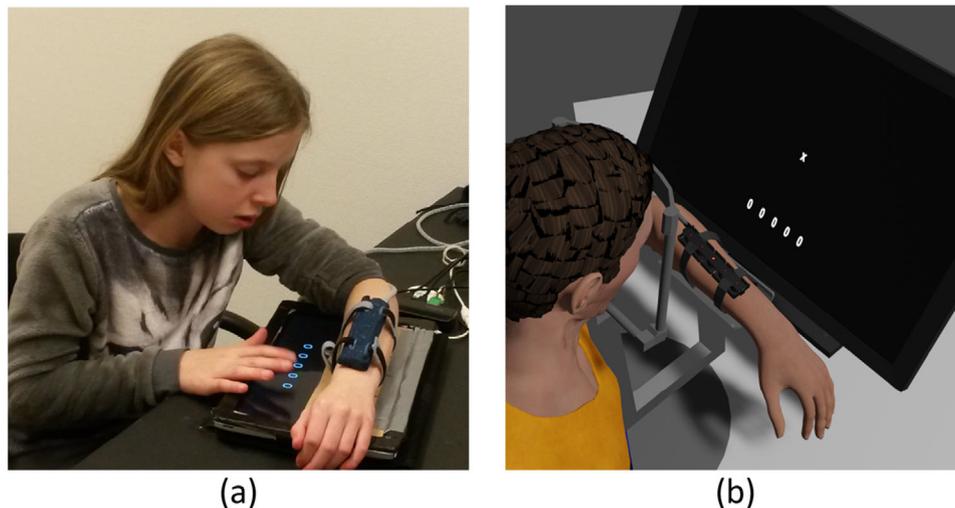


FIGURE 2 Experimental setup for children (a) and adults (b). Participants had to report at which of five nodes a tactile stimulus was felt. On five out of six trials, the tactile stimulus was accompanied by a visual stimulus, either at the same or at a different location. Responses were given by selecting one of five circles on a touchscreen tablet (children) or using a computer screen and mouse (adults)

phrased as that the little light would sometimes, but not always, try to fool them.

On each trial, a tactile stimulus was presented at one of the five possible locations. In five out of six trials, a concurrent visual stimulus was presented at a location that was independent of the tactile stimulus location (bimodal trials). In the remaining one-sixth of trials, no visual stimulus was presented (unimodal trials). During each block of 90 trials, every stimulus combination (five tactile \times five visual locations + five tactile-only locations) was presented three times. Within a block, the order of trials was randomized. After each stimulus or stimulus-pair, participants were asked to report the perceived location of the tactile stimulus. Children performed four and adults six blocks of trials. Between blocks there was a rest period between 10 s and 10 min.

Children used a touchscreen to give their response. The response options were depicted as five “o” shapes that were aligned with node locations and turned into a “+” upon selection. Adults selected one of five possible locations depicted on the computer screen, using a computer mouse. There was no time limit for giving a response. Participants could start the next trial by clicking the mouse or touching the touchscreen, which was followed by a randomly varying time interval between 800 and 1300 milliseconds before the next stimulus was presented. No feedback about performance was given.

In addition to the localization task, adult participants performed a two-alternative forced-choice task, in which they had to report whether or not they perceived the visual and a tactile stimulus to be at the same location. A report of this part of the experiment is available as Supplemental Material with the online version of this paper.

2.4 | Data analysis

To test whether a causal inference strategy guided the responses in the localization task, we investigated the relation between the cue dis-

parity and the response bias, expressed as a percentage of this disparity. Per participant, we computed the Spearman’s rank correlation between the percent bias and the absolute disparity over trials. If causal inference guides perception, we expected that the localization responses would be drawn stronger towards the visual cue for small compared to large disparities. Hence, we then expect to find a negative correlation between bias and disparity. For either pure integration (full fusion) or pure segregation, the bias would be independent of disparity (Figure 1d). Correlations were tested against zero at a group level using a one-sample *t*-test, and compared between age groups (young children; older children; adults) using a one-way ANOVA.

2.5 | Modeling

Six different Bayesian models were examined as an account for the response distributions in the localization task. Of those models, three involved causal inference (cf. Wozny et al., 2010): *model averaging*, *model selection*, and *probability matching*. The other models were a pure *segregation* model, a *forced fusion* model and a *switching* model.

The three causal inference models operate under the assumption that participants estimate the posterior probability that the visual and tactile stimulus have a common cause. A generative model is used that states that the two stimuli (tactile and visual) either have a common cause, and hence come from the same location, or alternatively, that they have multiple, independent causes, and hence may come from different locations (Figure 1a). The a priori probability of the two stimuli being generated at the same location is referred to as the *prior on common cause*, or p_{common} . A posterior on common cause is computed by multiplying p_{common} with the likelihood of having the actual sensory input under the common cause assumption. Hence, this posterior on common cause becomes smaller if the disparity between the (unisensory) perceived visual cue and tactile cue becomes larger. The

**TABLE 1** Parameters that were fitted through the MCMC procedure and their prior distributions

Parameter	Used in model						Prior distribution
	Model averaging	Model selection	Probability matching	Forced fusion	Segregation	Switching	
p_{common}	✓	✓	✓				$U(0,1)$
σ_t	✓	✓	✓	✓	✓	✓	$Cau(2)$
μ_p	✓	✓	✓	✓	✓	✓	$N(2,5)$
σ_p	✓	✓	✓	✓	✓	✓	$Cau(5)$
w_t				✓			$U(0,1)$
p_t						✓	$U(0,1)$
lapse rate	✓	✓	✓	✓	✓	✓	$U(0, .2)$

Note: The role of each parameter in the various models is described in the appendix. $U(a,b)$: uniform distribution over the interval $[a,b]$; $Cau(\gamma)$: half-Cauchy distribution with scale parameter γ ; $N(\mu,\sigma)$: normal distribution with mean μ and standard deviation σ . All lengths ($\sigma_t, \mu_p, \sigma_p$) are taken in units of inter-node distance (2.5 cm).

posterior on common cause dictates to what extent the final estimate of the tactile cue location is based on either the segregated tactile percept or the optimally integrated multisensory percept (Wozny et al., 2010). In *model averaging* (Figure 1b), a weighted sum is taken of the tactile percept and the integrated percept, with a relative weight on the integrated percept equal to the posterior on common cause. In *model selection*, the localization corresponds to the most likely causal structure, given the posterior on common cause (a winner-takes-all policy). Finally, in *probability matching*, the participant localizes either based on the integrated multisensory percept or the segregated tactile percept; the probability of choosing the multisensory percept in that trial is set such that it equals the posterior on common cause of that particular trial. Hence, the crucial difference between model selection and probability matching is that the former always discards the least-likely causal structure, while in the latter, there is still a chance for the less likely causal structure to be used. The other three models do not rely on a causal inference computation. According to these models, localization is based solely on the segregated and integrated estimates. In the *forced fusion* model, a weighted average is taken between the sensed tactile and the visual location, with a fixed relative weighting. This model allows sub-optimality by leaving the relative tactile weight as a free parameter during the fitting, rather than forcing it to be equal to the relative precisions of the individual sense. Unlike the causal inference (model averaging) model, the forced fusion model gives the same weight to the tactile cue throughout the whole experiment, independent of cue disparity. In that sense, it assumes that the visual cue is equally informative about the tactile stimulus location for any visual-tactile stimulus pair. According to the *segregation* model, the visual stimulus had no effect and hence the response location should always correspond to the segregated tactile input location. The *switching* model operates under the assumption that, on every trial, a participant chooses either the tactile or the visual stimulus and ignores the other, with a probability for picking either one that is fixed over trials.

In all six models, we assumed that the location of the visual stimulus could be categorized perfectly. Consequently, the optimally integrated multisensory percept always corresponded to the actual visual stimu-

lus location. This was confirmed by a small pilot experiment in which the participant was asked to report which of the five lights had lit up in a unimodal visual stimulus trial. For a detailed mathematical description of the various models, we refer to the appendix.

2.6 | Model fitting

Table 1 provides an overview of the fitted parameters of each model and their prior distributions in the fitting procedure. The goal of the fitting procedure was to obtain a probability distribution over the parameter values for each model, given the data of each individual participant, and use this distribution to compute the model likelihood. We used the Goodman and Weare (2010) affine invariant ensemble Markov Chain Monte Carlo (MCMC) sampler, as implemented in Matlab by Grinsted (2018), to sample parameter values from this posterior distribution. Each step in the MCMC sampler required computation of the log-likelihood of the data given the model and a particular set of parameters. In the current case, the log likelihood was given by:

$$\log L(\text{model}|\{n_{ij}\}) = \sum_{j=\{s_v, s_t\}} \sum_{i=1}^5 n_{ij} \log p_{ij} + \text{constant}, \quad (1)$$

where $\{n_{ij}\}$ are the response counts for each response option i in condition j , and $\{p_{ij}\}$ are the respective modeled response probabilities (Körding et al., 2007).

For each participant and every model, 24 chains of 417 samples (in total 10,008 samples) were taken from the posterior joint parameter distribution. The initial parameter values of each chain were drawn from their respective prior distributions. No burn-in was used, but to improve chain mixing, every sample was the result of ten sampling-steps (i.e., a total of 100,080 samples were generated in each chain and every tenth sample was stored). Chain mixing and convergence were confirmed through visual inspection of the posterior distributions.

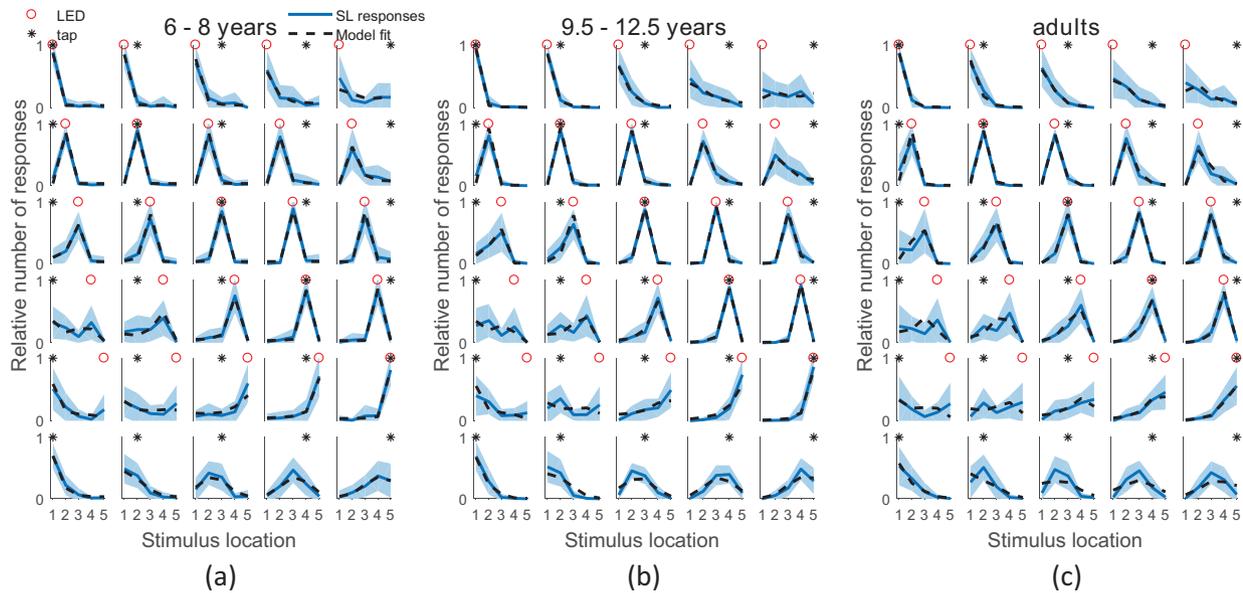


FIGURE 3 Average response distributions for the different conditions, per age group. Columns correspond to different tactile stimulus locations, rows to different visual stimulus locations. Per condition, the fraction of responses for each location are averaged over participants, shaded regions indicate standard errors. Dotted lines display model predictions, based on a group-level MCMC fit of the model averaging causal inference model. Different panels correspond to different age groups: (a) 6- to 8-year-olds, (b) 9.5- to 12.5-year-olds and (c) adults. Node 1 was the most proximal (close to the elbow), node 5 the most distal

The likelihood of a given model for each participant was determined by computing the likelihood of the data for each of the MCMC samples. The mean of those samples was chosen as a proxy for the overall *model likelihood*. For each participant, the best model was defined as the model with the highest likelihood.

Maximum a posteriori probability (MAP) estimates of the model parameters for each participant and each model were obtained by taking the sample from the combined MCMC chains that corresponded to the highest likelihood.

We performed a comparison over the estimated parameter values for the best-fitting causal inference model (i.e., the model that came out as best in the largest number of subjects). For the comparison, a (non-parametric) Kruskal-Wallis test was used, because for all parameters, at least one of the groups showed a non-normal distribution, as revealed by an Anderson-Darling test. In case of a significant age-effect, a Wilcoxon rank-sum test was conducted to pairwise compare the different age groups. We used an alpha level of 0.05 for all statistical tests. All analyses were performed using Matlab R2017a (The Mathworks, Inc.).

3 | RESULTS

3.1 | Behavioral observation

Figure 3 presents the average responses of each age group. Different panels show the different stimuli-combinations, with the bottom panels showing the responses to the unimodal, tactile stimulus. In all

age groups, this unimodal response distribution is shifted leftwards (towards node 1) with respect to the actual location, indicating a bias towards the elbow, in line with the tactile localization study of Badde et al. (2020). When comparing column-by-column the lowest panels (tactile only) to the corresponding panels (bimodal) above, it is evident that responses in bimodal trials were shifted towards the visual stimulus location, indicating that the visual stimulus affected tactile localization.

Figure 4 shows the percent bias as a function of the disparity between the visual and tactile stimulus. Perfect localization is indicated by a bias of zero, while a bias of 100% indicates that the tactile stimulus is localized at the visual stimulus location. As can be seen in the figure, the bias decreased with larger disparities. This finding is in line with a causal inference strategy and contradicts full fusion, pure segregation or switching, all of which predict that the percentual bias would be independent of disparity (see Figure 1d).

Statistical analyses confirmed that percent bias decreased with larger disparities: the mean correlation between the bias and disparity differed significantly from zero for all age groups (mean correlations: $r_s = -0.44$; -0.49 ; -0.22 for young children, older children and adults, respectively; $t(20) = -9.35$; $t(24) = -10.88$; $t(19) = -5.18$; all p 's < 0.001). There was a significant age effect on this correlation ($F(2,63) = 10.33$, $p < 0.001$), with a weaker correlation for adults compared to younger children ($t(39) = 3.53$, $p < 0.01$) and older children ($t(43) = 4.41$, $p < 0.01$), indicating that the decrease in bias with larger disparity was more prominent in children than in adults.

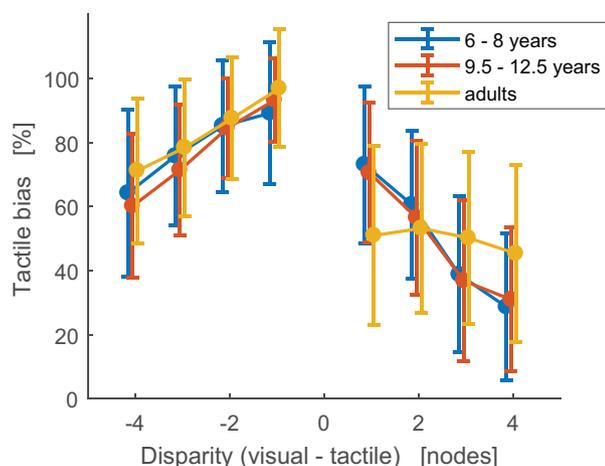


FIGURE 4 Response bias as a function of cue disparity. Disparity is presented in units of inter-node distance, which correspond to 2.5 cm. A bias of zero indicates perfect localization, whereas a bias of 100% indicates that the tactile stimulus is localized at the visual stimulus location. A negative relation between bias and absolute disparity is indicative of a causal inference strategy

3.2 | Model comparison

In all age groups, in nearly all participants, the model averaging model outperformed the other models (Figure 5a). The average (base e) log-likelihood difference of model selection with the best non-causal inference model was 17 for young children, 40 for older children, and 47 for adults (see Figure 5b for the relative log-likelihoods of all models), indi-

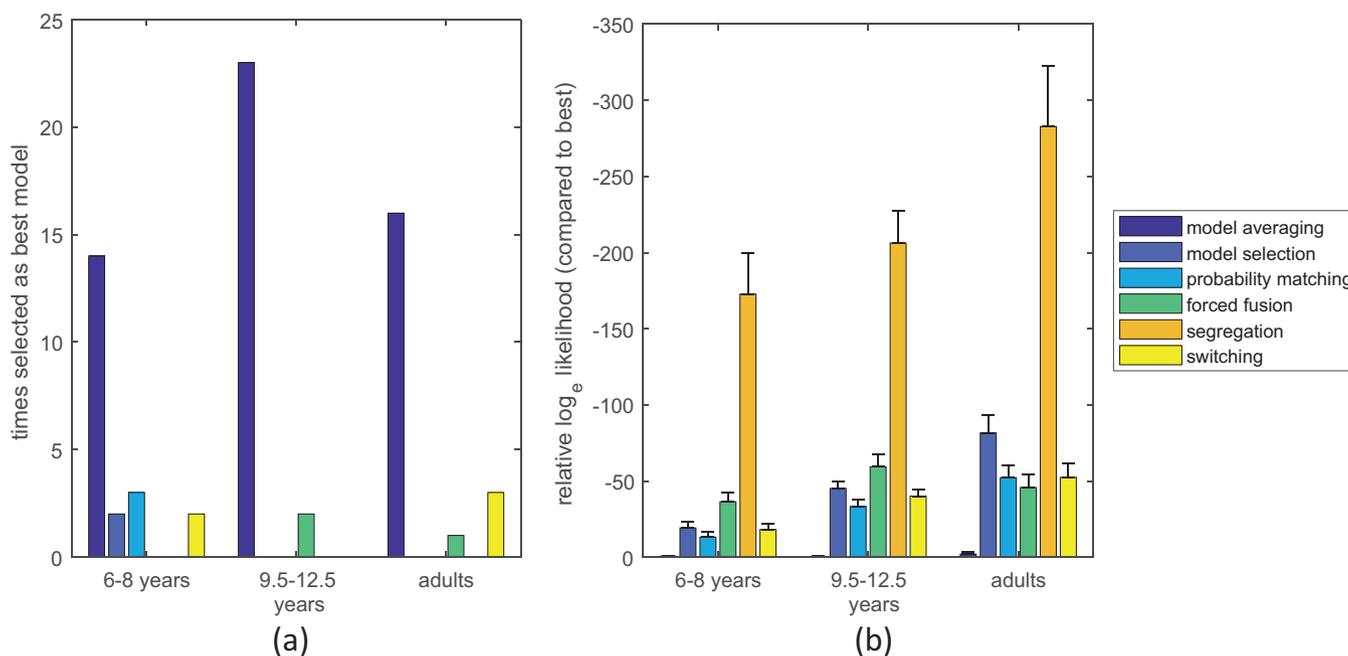


FIGURE 5 Model comparison results. (a) The number of participants for which a particular model gave the best fit. (b) The average difference in log likelihood between each model and the best fitting model for each participant (error bars indicate standard errors)

cating very strong evidence in favor of a causal inference strategy (Kass & Raftery, 1995).

3.3 | Parameter comparison

Because the model averaging model emerged as best fitting model in all groups, we compared its parameter estimates across the three age groups. Specifically, we hypothesized that if young children are more likely to segregate rather than integrate cues, they would have a lower prior on common cause, compared to older children and adults. However, a higher value for this parameter was found for both children groups compared to adults (main effect of age: $H(2) = 10.81, p < 0.005$; planned contrasts: $p = 0.014$ for young children vs. adults, $p < 0.005$ for older children vs. adults, $p = 0.12$ for young children vs. older children). Figure 6 shows the mean and individual MAP estimates of p_{common} , together with the other mean MAP parameter estimates of the model averaging model.

To check the robustness of our statistical outcomes, we re-ran the comparison of p_{common} using only the participants in which “model averaging” was the best model. The results were the same as in the original analyses: a significant main effect of age ($H(2) = 12.73, p < 0.005$), with significantly higher p_{common} for young children compared to adults ($p < 0.005$) and older children compared adults ($p < 0.005$), but no significant difference between young children and older children ($p = 0.28$).

Regarding the other model parameters, we only found significant effects of age on the tactile standard deviation σ_t ($H(2) = 9.46, p < 0.01$), with adults ($\sigma_t = 2.7$) being less precise than young ($\sigma_t = 1.4, p < 0.005$) and older ($\sigma_t = 1.8, p = 0.013$) children, and on lapse rate ($H(2) = 11.15, p < 0.005$), with younger children showing higher lapse rates ($M = 0.07$).

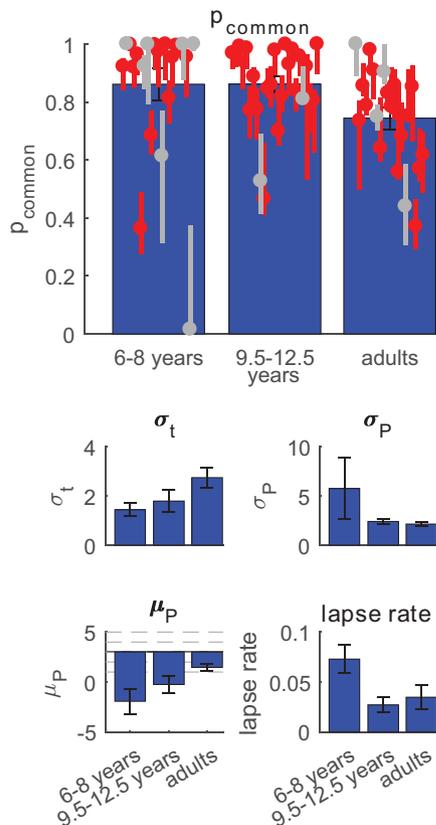


FIGURE 6 Estimates of parameter values for the model averaging model. For p_{common} , individual MAP estimates and 95% credible intervals are shown. Red data points correspond to participants for whom model averaging outperformed the other models, grey points correspond to participants for whom one of the other models was best. Bar plots show the mean MAP estimates for all parameters. For μ_p the dotted lines correspond to the five node locations; $\mu_p = 3$ corresponds to a prior centered at the middle node. p_{common} : prior on common cause; σ_t : standard deviation of tactile noise; σ_p , μ_p : standard deviation and center of Gaussian location prior; lapse rate: proportion of trials in which a random response is given

than older children ($M = 0.03$, $p < 0.005$) and adults ($M = 0.03$, $p < 0.005$).

4 | DISCUSSION

The aim of this study was to examine whether children use causal inference in multisensory perception, and if so, whether younger children (6- to 8-year-old) are less likely to integrate different cues than older children (9.5- to 12.5-year-old) or adults. An increased tendency to segregate cues in younger children, expressed as a lower a priori belief that different cues originate from a single environmental source, would explain earlier findings that suggested that younger children do not integrate multisensory cues.

The current results suggest that children from a young age apply a causal inference strategy, just like older children and adults. In all three age groups, the relative effect of the visual cue on the felt tactile cue

location became smaller for larger disparities, which is a characteristic property of a causal inference strategy. Moreover, for the majority of participants, a causal inference model described the data better than either a forced fusion, a segregation or a switching model. In the few participants for whom one of the alternative models was a better descriptor, the model averaging model described the data almost equally well, as evidenced by the small differences in the model likelihood. Hence, from a Bayesian perspective, we conclude that causal inference plays a role in multisensory integration already in 6- to 8-year-olds.

We found a negative relation between the visually induced tactile bias and the absolute disparity. However, the results in Figure 4 suggest an uneven relation between bias and disparity: for negative disparities (i.e., when the visual cue was closer to the elbow than the tactile cue), a larger bias was observed than for positive disparities. Moreover, for positive disparities, the adult group showed a consistent percentual bias (i.e., bias as a percentage of the disparity), rather than a bias that decreases with increasing disparity. The explanation of these findings is that we used the actual tactile stimulus location when computing bias and disparity, while the perceived tactile stimulus location was generally shifted towards the elbow (see bottom panels of Figure 3a-c). This unisensory tactile shift leads to leftward biases being overestimated and rightward biases being underestimated, in particular for disparities close to zero. Hence, the asymmetry in the relation between bias and disparity was due to a shift in tactile likelihood (cf. Badde et al., 2020), rather than due to an asymmetry in causal inference.

Importantly, out of the candidate causal inference models, model averaging outperformed model selection and probability matching. This is an interesting finding, because in contrast to the two alternatives, model averaging involves weighted contributions of visual and tactile information on a single trial. Taking a weighted average of different sensory cues is a requirement for optimal cue integration as often described to occur in adults (Ernst & Banks, 2002; van Beers et al., 1996, 1999). It has been debated whether young children take weighted averages when integrating cues, or rather pick a single cue (Gori et al., 2008) or switch between cues on a trial-by-trial basis (Nardini et al., 2010). Our results suggest that the ability to weigh different cues, at least when the weights are based on inference about the causal structure of sensory input, is already present before 8 years of age. Moreover, out of all the proposed models, model averaging is the only one that assumes optimality in the sense that it aims at minimizing the squared error of the tactile location estimate. Hence, in the current experiment, all age groups showed behavior that can be considered optimal, given their individual prior belief that the visual and tactile cue share a common cause.

We note that gender distributions were not matched across the three participant groups. Based on adult studies, it has been suggested that women have a stronger tendency to bind multisensory signals than men (Barnett-Cowan et al., 2010; Claypoole & Brill, 2019; Colignon et al., 2010; but see Magnotti & Beauchamp, 2018). However, our results suggests that the adult group, which consisted of slightly more females, showed the smallest prior on common cause, and thus



the weakest tendency to bind signals. Therefore, it is unlikely that the gender imbalance strongly affected the outcomes of the current study.

When successfully integrating multiple cues, the precision of the combined estimate should be better than the precision for any of the cues presented in isolation. This property is often used as a criterion for cue integration in children (Dekker et al., 2015; Gori et al., 2008; Nardini et al., 2008, 2010, 2013; Nava et al., 2020; Negen et al., 2019; Petrini et al., 2014). However, for two reasons, our study design did not allow the quantification of precision improvement through cue weighting. Firstly, when causal inference is applied, improved precision for combined-cue trials compared to single-cue trials is not generically true for an ideal observer. Secondly, a visual cue was used that could be perceived nearly infinitely precisely, at least with regard to the categorical localization task used here. Therefore, practically no precision improvement was possible by combining cues, compared to using the best unimodal cue. Under this assumption, any deviation from the visual stimulus location should be regarded as the result of segregation of the visual and tactile cues. Hence, the current study apparatus was particularly suitable to address causal inference, at the cost of not being able to address precision improvement through optimal cue integration.

To date, only one study explicitly addressed the possibility that causal inference could explain the (apparent) sub-optimal cue integration that is often found in young children. Rohlf et al. (2020) used an auditory-visual localization experiment to study the ventriloquist-effect and cross-sensory recalibration in 5- to 9-year-old children. They showed, as in the current study, that a causal inference model explained their data better than an optimal integration (forced fusion) model or a switching model. However, because their study only included a single disparity, the corresponding data could not tease apart a causal inference model from a forced fusion model with sub-optimal weights. While both “causal inference with model averaging” and “forced fusion” predict that an observer takes a weighted average of cues, the crucial difference between the predictions of these models is that in causal inference the respective weights depend on the disparity between cues. Hence, by using a range of stimulus disparities, we could distinguish causal inference from sub-optimal forced fusion, and found support for the first over the latter model. The current results thus strengthen the conclusion of Rohlf and colleagues that causal inference is indeed a better descriptor than “simple” integration, when it comes to multisensory cue localization in young children.

A priori, we hypothesized that young children would be less likely to take two cues to represent the same environmental property, compared to older children and adults. However, we did not find a lower prior on common cause for the 6- to 8-year-old children compared to the two older groups. Therefore, it is unlikely that the lack of multisensory integration in young children, as found in earlier studies (Gori et al., 2008; Nardini et al., 2013; Petrini et al., 2014), can be explained by a generally higher tendency to segregate rather than integrate multiple cues. If anything, the current study suggests that the opposite is true: children (from both age groups) had a harder time than adults in disentangling information from different sensory streams, compared to adults. This reduced ability in children to ignore or suppress an irrel-

evant visual cue is in line with earlier work by Petrini et al. (2015), who found that 7- to 10-year-old children were biased towards a visual cue when judging the location of a sound, while adults' responses were unaffected. In this light, a recent study by Badde et al. (2020) provides an interesting insight. Using a visuotactile localization task very similar to the one used in the current study, these authors showed, in adults, that the prior on common cause can be lowered by directing the attention towards either the visual or the tactile cue before the stimulus is presented. In the current study, participants were asked to specifically attend the tactile cue, which could have lowered the effective prior on common cause in adults. Potentially, this could explain why children were found to have a higher prior on common cause than adults: because inhibitory control is still developing up to 13 years of age at least (Davidson et al., 2006), children might have struggled to suppress the uninformative but salient visual cue.

An alternative interpretation of the larger prior on common cause in children compared to adults is that this difference is an expression of an age related narrowing of multisensory binding windows. It has been found before that both the spatial and temporal windows in which two cues are judged as being concurrent shrink with age during childhood (Greenfield et al., 2017; Hillock-Dunn & Wallace, 2012; Lewkowicz & Flom, 2014). A with age decreasing prior on common cause fits within this general developmental trend. Combining those findings of narrowing binding windows with earlier findings of sub-optimal cue integration in young children, it seems that over the course of development, humans become less likely to integrate information from different cues, but become more proficient when they do.

Until a few years ago, children younger than 8–10 years of age were thought not to integrate different sensory modalities because they might still be calibrating their senses (Gori, 2015; Gori et al., 2008), resulting in dominance of one sense over the other (Gori et al. 2008; Petrini et al., 2014). Also, it has been suggested that “the developing visual system may be optimized for speed and detecting sensory conflicts” (Nardini et al., 2010, p. 17041). Recently, these ideas have been challenged by studies showing near-optimal integration in this age group (Bejjanki et al., 2020; Nava et al., 2020; Negen et al., 2019), and in particular by Rohlf et al. (2020), who showed that integration develops prior to cross-sensory recalibration. In the current study, we found that even though perception of tactile stimuli is strongly affected by the presence of a visual cue, the less-dominant modality (touch) still had an effect on the perceived stimulus location. Hence, the current study provides additional evidence that children's multisensory perception is not fully dominated by a single modality. Nonetheless, the lower effective prior on common cause in adults compared to children suggests that visual dominance becomes less pronounced with age, allowing adults to better segregate information from different senses.

In conclusion, we found that 6- to 8-year-old children, like older children and adults, apply causal inference when presented with two simultaneous cues in different sensory modalities. The current experiment was set up in such a way that causal inference could be distinguished from cue integration, by allowing the cue in one modality to be perceived nearly infinitely precise, and showed that in such a case, cue localization indeed appears to be under the influence of causal



inference. However, in real-world situations, causal inference and weighted cue integration often go hand in hand, for example when trying to swat a mosquito off one's arm: one wants to be accurate in localizing *and* to disentangle which of the sensations on the skin is caused by the mosquito. Future developmental research could aim at the interplay between these two aspects of multisensory perception, causal inference and integration, by using tasks in which both are required. The current results highlight that causal inference is an important factor to take into account when studying the development of cue integration.

ACKNOWLEDGMENTS

This work was supported by a grant of the Swedish Research Council (JCS, grant nr. 2016-01725). The authors thank all the participating adults and children. We also thank primary school De Werkplaats for welcoming us and providing the opportunity to run our experiment at their location. Lastly, we thank the research assistants who helped collecting data.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICAL APPROVAL

The study conforms to the Declaration of Helsinki. All procedures were approved by the local Ethics Committee of the Faculty of Social Sciences of Radboud University Nijmegen under approval ECSW-2019-089.

ORCID

Erik Verhaar  <https://orcid.org/0000-0002-5902-0212>

REFERENCES

- Acerbi, L., Dokka, K., Angelaki, D. E., and Ma, W. J. (2018). Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLoS Computational Biology*, *14*(7):e1006110.
- Atsma, J., Maji, F., Koppen, M., Irwin, D. E., and Medendorp, W. P. (2016). Causal inference for spatial constancy across saccades. *PLoS Computational Biology*, *12*(3):1–20.
- Badde, S., Navarro, K. T., and 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.
- Bejjanki, V. R., Randrup, E. R., and Aslin, R. N. (2020). Young children combine sensory cues with learned information in a statistically efficient manner: But task complexity matters. *Developmental Science*, *23*(3):1–13.
- Barnett-Cowan, M., Dyde, R. T., Thompson, C., & Harris, L. R. (2010). Multisensory determinants of orientation perception: task-specific sex differences. *European Journal of Neuroscience*, *31*(10), 1899–1907.
- Claypoole, V. L., & Brill, J. C. (2019). Sex differences in perceptions of asynchronous audiotactile cues. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* *63*(1):1204–1208.
- Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., & Lassonde, M. (2010). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia*, *48*(1), 220–225.
- Davidson, M. C., Amso, D., Anderson, L. C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11):2037–2078.
- Dekker, T., Ban, H., Van Der Velde, B., Sereno, M. I., Welchman, A. E., and Nardini, M. (2015). Late development of cue integration is linked to sensory fusion in cortex. *Current Biology*, *25*(21):2856–2861.
- Dekker, T. and Lisi, M. (2020). Sensory development: Integration before calibration. *Current Biology*, *30*(9):R409–R412.
- Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870):429–433.
- Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1):65–80.
- Gori, M. (2015). Multisensory integration and calibration in children and adults with and without sensory and motor disabilities. *Multisensory Research*, *28*(1-2):71–99.
- Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*(9):694–698.
- Greenfield, K., Ropar, D., Themelis, K., Ratcliffe, N., and Newport, R. (2017). Developmental changes in sensitivity to spatial and temporal properties of sensory integration underlying body representation. *Multisensory Research*, *30*(6):467–484.
- Grinsted, A. (2018). Ensemble mcmc sampler (<https://github.com/grinsted/gwmcmc>), GitHub. Retrieved June 21, 2018.
- Hillock-Dunn, A. and Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, *15*(5):688–696.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430):773–795.
- Kayser, C. and Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biology*, *13*(2):1–7.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9).
- Lewkowicz, D. J. and Flom, R. (2014). The audiovisual temporal binding window narrows in early childhood. *Child Development*, *85*(2):685–694.
- Magnotti, J. F., & Beauchamp, M. S. (2018). Published estimates of group differences in multisensory integration are inflated. *PLoS one*, *13*(9), e0202908.
- Nardini, M., Bedford, R., and Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(39):17041–17046.
- Nardini, M., Begus, K., and Mareschal, D. (2013). Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3):773–787.
- Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*(9):689–693.
- Nava, E., Föcker, J., and Gori, M. (2020). Children can optimally integrate multisensory information after a short action-like mini game training. *Developmental Science*, *23*(1):1–9.
- Negen, J., Chere, B., Bird, L.-A., Taylor, E., Roome, H. E., Keenaghan, S., Thaler, L., and Nardini, M. (2019). Sensory cue combination in children under 10 years of age. *Cognition*, *193*:104014.
- Perdreau, F., Cooke, J. R., Koppen, M., and Medendorp, W. P. (2019). Causal inference for spatial constancy across whole body motion. *Journal of Neurophysiology*, *121*(1):269–284.
- Petrini, K., Jones, P. R., Smith, L., & Nardini, M. (2015). Hearing where the eyes see: Children use an irrelevant visual cue when localizing sounds. *Child Development*, *86*(5), 1449–1457.
- Petrini, K., Remark, A., Smith, L., and Nardini, M. (2014). When vision is not an option: Children's integration of auditory and haptic information is suboptimal. *Developmental Science*, *17*(3):376–387.



- Rohe, T., & Noppeney, U. (2015). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, 15(5), 22-22.
- Rohlf, S., Li, L., Bruns, P., and Röder, B. (2020). Multisensory integration develops prior to crossmodal recalibration. *Current Biology*, 30(9):1726-1732.e7.
- Scheller, M., Proulx, M. J., De Haan, M., Dahmann-Noor, A., & Petrini, K. (2021). Late-but not early-onset blindness impairs the development of audio-haptic multisensory integration. *Developmental Science*, 24(1), e13001.
- van Beers, R. J., Sittig, A. C., and Denier van der Gon, J. J. (1996). How humans combine simultaneous proprioceptive and visual position information. *Experimental Brain Research*, 111(2):253-261.
- van Beers, R. J., Sittig, A. C., and Denier van der Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, 81(3):1355-1364.
- Wozny, D. R., Beierholm, U. R., and Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, 6(8).

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Verhaar, E., Medendorp, W. P., Hunnius, S., and Stapel, J. C. (2022). Bayesian causal inference in visuotactile integration in children and adults. *Developmental Science*, 25, e13184. <https://doi.org/10.1111/desc.13184>

APPENDIX A

Appendix Computational models

We considered six candidate models to describe the responses in the localization task. Three of those were adaptations of the causal inference models proposed by Körding et al. (2007) and Wozny et al. (2010). In our versions of these models, we assumed that in the current task visual input was precise enough to be modeled as noiseless, that is, we assumed that participants could always unambiguously indicate at which of the five positions the LED was activated on each trial.

Computing the posterior on common cause

In all used models, the first step is to obtain unisensory estimates of the tactile and visual stimulus locations. In this step, a possible bias in tactile localization on the forearm is included, which was modeled by introducing a prior on tactile stimulus location. The tactile input signal x_t was assumed to be drawn from a normal distribution with variance σ_2 ; t centered at the actual tactile stimulus location s_t . The tactile prior was modeled as a Gaussian distribution with variance σ_2 ; p and mean μ_p . Integration of tactile input and prior results in a biased tactile signal:

$$\bar{x}_t = \frac{\sigma_p^2 x_t + \sigma_t^2 \mu_p}{\sigma_p^2 + \sigma_t^2}. \quad (2)$$

The variance associated to \bar{x}_t (that is, the width of the posterior tactile stimulus location) is given by

$$\bar{\sigma}_t^2 = \frac{\sigma_p^2 \sigma_t^2}{\sigma_p^2 + \sigma_t^2}. \quad (3)$$

Including a unisensory tactile bias in our model was motivated by the observation that participants tended to report the tactile stimulus to be shifted towards the elbow in unimodal trials (see the bottom rows of plots in Figure 3). The location of the visual input signal x_v was set equal to the visual stimulus location s_v .

The causal inference models had the underlying assumption that when participants report the perceived tactile location, they have to make an inference on whether or not the visual stimulus location was informative about the location of the tactile stimulus. In the models, such an inference is based on the (biased) sensory signals (\bar{x}_t and x_v), the tactile uncertainty $\bar{\sigma}_t$ and a prior probability on common cause p_{common} .

According to the generative model used in each model, either the visual and tactile signal were generated at the same location (denoted by $C = 1$), or the two signals were generated independently (denoted by $C = 2$), which allowed them to originate from different locations. In both cases, stimuli are assumed to be drawn from a uniform distribution over the five possible locations. Following Bayes rule, the probability of $C=1$ given the sensory input is given by:

$$p(C = 1 | \bar{x}_t, x_v) = \frac{p(\bar{x}_t, x_v | C = 1) p_{common}}{p(\bar{x}_t, x_v | C = 1) p_{common} + p(\bar{x}_t, x_v | C = 2) (1 - p_{common})}. \quad (4)$$

Here, the likelihoods are given by:

$$\begin{aligned} p(\bar{x}_t, x_v | C = 1) &= \sum_{s \in S} p(\bar{x}_t | s) p(x_v | s) p(s) \\ &= \frac{1}{5} \mathcal{N}(s_v; \bar{x}_t, \bar{\sigma}_t) \end{aligned} \quad (5)$$

and

$$\begin{aligned} p(\bar{x}_t, x_v | C = 2) &= p(\bar{x}_t | C = 2) p(x_v | C = 2) \\ &= \left(\sum_{s \in S} p(\bar{x}_t | s) p(s) \right) \left(\sum_{s \in S} p(s_v | s) p(s) \right) \\ &= \frac{1}{25} \mathcal{N}(s; \bar{x}_t, \bar{\sigma}_t) \end{aligned} \quad (6)$$

where S denotes the set of five possible stimulus locations and $N(s; \mu, \sigma)$ a normal distribution with mean μ and standard deviation σ , evaluated at s . Note that the prior $p(s)$ is equal to $1/5$ for each location s . The left side of Equation (4) is called the posterior on common cause. This posterior has a value between 0 and 1 and reflects the probability that the visual cue was informative about the tactile stimulus location.

Estimation of tactile stimulus location

Following Wozny et al. (2010), we considered three different ways in which causal inference may have led to an estimate of tactile stimulus location. In all cases, two separate estimates of the tactile stimulus location are computed: the integrated estimate and the segregated



estimate. For the integrated estimate, the tactile and visual cues are thought to be at the same location. Because we assumed that the visual stimulus could be perceived with infinite precision, this integrated location is equal to the visual stimulus location:

$$\hat{s}_{t,C=1} = s_v. \quad (7)$$

The segregated estimate, in which the visual cue is ignored, is given by the (biased) unisensory tactile signal (Equation (2)):

$$\hat{s}_{t,C=2} = \bar{x}_t. \quad (8)$$

According to *model averaging*, the final percept is obtained by taking a weighted average of these two estimates, with the posterior on common cause as weighing factor:

$$\hat{s}_t = p(C = 1\bar{x}_t, x_v) \hat{s}_{t,C=1} + (1 - p(C = 1\bar{x}_t, x_v)) \hat{s}_{t,C=2} \quad (\text{model averaging}). \quad (9)$$

According to *model selection*, an observer selects the integrated or segregated signal based on which of the two causal structures was more likely according to the posterior on common cause:

$$\hat{s}_t = \begin{cases} \hat{s}_{t,C=1}, & \text{if } p(C = 1\bar{x}_t, x_v) > 0.5 \\ \hat{s}_{t,C=2}, & \text{if } p(C = 1\bar{x}_t, x_v) \leq 0.5 \end{cases} \quad (\text{model selection}). \quad (10)$$

Lastly, in *probability matching*, an observer randomly picks either the integrated or segregated signal, with the probability of picking the integrated signal equaling the posterior on common cause:

$$\hat{s}_t = \begin{cases} \hat{s}_{t,C=1}, & \text{with probability } p(C = 1\bar{x}_t, x_v) \\ \hat{s}_{t,C=2}, & \text{with probability } 1 - p(C = 1\bar{x}_t, x_v) \end{cases} \quad (\text{probability matching}). \quad (11)$$

Non-causal-inference models

Aside from the three causal inference models described above, three models which did not involve causal inference were fitted. In the *forced fusion* model, the estimate of tactile location is computed by taking a weighted average of the visual and tactile cue:

$$\hat{s}_t = w_t \bar{x}_t + (1 - w_t) s_v \quad (\text{forced fusion}), \quad (12)$$

where w_t is the weight assigned to the (unisensory) tactile percept. In the *segregation* model the visual cue is completely ignored:

$$\hat{s}_t = \bar{x}_t \quad (\text{segregation}). \quad (13)$$

Lastly, the *switching* model postulates that an observer alternatively follows either the visual or the tactile cue, with a fixed probability:

$$\hat{s}_t = \begin{cases} \bar{x}_t, & \text{with probability } p_t \\ s_v, & \text{with probability } 1 - p_t \end{cases} \quad (\text{switching}). \quad (14)$$

where p_t is the probability of picking the tactile cue.

Providing a localization response

In the experiment, participants reported the perceived tactile stimulus location by selecting one of five possible locations. In the model fitting, this was realized by selecting the location closest to \hat{s}_t , following Körding et al. (2007). Lastly, to allow for attentional lapses, a *lapse rate* was included in each model, representing the fraction of trials in which an observer randomly selects one of the five possible response locations, independent of the stimulus presented.