

RESEARCH ARTICLE

# The Development of Audio-Visual Integration for Temporal Judgements

Wendy J. Adams\*

Psychology, University of Southampton, Southampton, Hampshire, United Kingdom

\* [w.adams@soton.ac.uk](mailto:w.adams@soton.ac.uk)

## Abstract

Adults combine information from different sensory modalities to estimate object properties such as size or location. This process is optimal in that (i) sensory information is weighted according to relative reliability: more reliable estimates have more influence on the combined estimate and (ii) the combined estimate is more reliable than the component uni-modal estimates. Previous studies suggest that optimal sensory integration does not emerge until around 10 years of age. Younger children rely on a single modality or combine information using inappropriate sensory weights. Children aged 4–11 and adults completed a simple audio-visual task in which they reported either the number of beeps or the number of flashes in uni-modal and bi-modal conditions. In bi-modal trials, beeps and flashes differed in number by 0, 1 or 2. Mutual interactions between the sensory signals were evident at all ages: the reported number of flashes was influenced by the number of simultaneously presented beeps and vice versa. Furthermore, for all ages, the relative strength of these interactions was predicted by the relative reliabilities of the two modalities, in other words, all observers weighted the signals appropriately. The degree of cross-modal interaction decreased with age: the youngest observers could not ignore the task-irrelevant modality—they fully combined vision and audition such that they perceived equal numbers of flashes and beeps for bi-modal stimuli. Older observers showed much smaller effects of the task-irrelevant modality. Do these interactions reflect optimal integration? Full or partial cross-modal integration predicts improved reliability in bi-modal conditions. In contrast, switching between modalities reduces reliability. Model comparison suggests that older observers employed partial integration, whereas younger observers (up to around 8 years) did not integrate, but followed a sub-optimal switching strategy, responding according to either visual or auditory information on each trial.



## OPEN ACCESS

**Citation:** Adams WJ (2016) The Development of Audio-Visual Integration for Temporal Judgements. *PLoS Comput Biol* 12(4): e1004865. doi:10.1371/journal.pcbi.1004865

**Editor:** Ulrik R. Beierholm, University of Birmingham, UNITED KINGDOM

**Received:** July 5, 2015

**Accepted:** March 14, 2016

**Published:** April 14, 2016

**Copyright:** © 2016 Wendy J. Adams. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are from human subjects. All data are available on request from the University of Southampton: DOI:[10.5258/SOTON/386590](https://doi.org/10.5258/SOTON/386590).

**Funding:** The author received no specific funding for this work.

**Competing Interests:** The author has declared that no competing interests exist.

## Author Summary

To complete everyday activities, such as judging where or when something occurred, we combine information from multiple senses such as vision and audition. In adults, this merging of information is optimal: more reliable sensory estimates have more influence (higher weight) in the combined, multisensory estimate. Multisensory integration can result

in illusions: if a single visual flash (e.g. a bright disk appearing briefly on a screen) occurs at the same time as two beeps, we sometimes perceive two flashes. This is because auditory information is generally more reliable than vision for judging *when* things happen; it dominates our audio-visual percept for temporal tasks. Previous work suggests that children don't combine information from different senses in this adult-like way until around 10 years. To investigate this further, we asked children and adults to report the number of visual flashes or auditory beeps when these were presented simultaneously. Surprisingly, all children used appropriate sensory weights: audition—the more reliable signal—tended to dominate perception, with less weight given to vision. However, children didn't show the adult-like reduction in uncertainty until around 8–10 years. Before that age, they switched between using only auditory or only visual information on each trial.

## Introduction

Imagine you are at an academic conference. A heated debate turns nasty and one scientist is repeatedly hit before falling to the floor. You are later asked how many punches were thrown. You confidently answer '3'; you were able to combine information from audition and vision, having both seen and heard the incident. We often receive information about the same object or event from multiple sensory modalities that we can integrate to improve the precision of our perceptual estimates. As adults, we integrate multisensory information for a variety of spatial and temporal tasks, such as judging the size, location, number or duration of objects or events [1–5]. A key benefit of this integration is that uncertainty, or variance (random noise) in the combined, multisensory estimate is reduced, relative to either of the component uni-sensory estimates, see e.g. [6].

Under standard models of integration, sensory estimates are combined via weighted averaging, according to the estimates' relative reliabilities, see e.g. [1, 2, 6]. For example, consider the case in which an aspect of the environment is estimated from vision and audition. The visual and auditory estimates,  $\hat{S}_V$  and  $\hat{S}_A$  are not perfectly precise, but contain noise with variance  $\sigma_V^2$  and  $\sigma_A^2$ . It is commonly assumed that these noise distributions are Gaussian and independent. Under these assumptions, and given that the prior probability distribution over the estimated variable is uniform, then the optimal audio-visual estimate (i.e. that with the lowest possible variance), for a continuous variable is given by:

$$\hat{S}_{VA} = w_V \hat{S}_V + w_A \hat{S}_A$$

with the visual and auditory weights,  $w_V$  and  $w_A$  defined as:

$$w_V = \frac{1/\sigma_V^2}{1/\sigma_V^2 + 1/\sigma_A^2} \text{ and } w_A = \frac{1/\sigma_A^2}{1/\sigma_V^2 + 1/\sigma_A^2}.$$

As can be seen from the equations above, sensory weights give the relative influence of each uni-modal sensory estimate in determining responses to bi-modal stimuli. These weights can be estimated from behavioural data corresponding to bi-modal and uni-modal stimulus conditions. For example, in a size estimation task such as [1], subjects might be required to estimate an object's size from vision alone, from haptics (touch) alone, or from both vision and haptics. If the visual size is 9cm and the haptic size is 12cm, then given unbiased uni-modal estimates, a mean bi-modal response of 10cm would correspond to visual and haptic weights of 2/3 and 1/3, respectively, i.e. vision has double the influence of haptics. These observed weights would be

optimal if the uni-modal visual estimates were twice as reliable as the uni-modal haptic estimates, i.e.  $\sigma_V^2/\sigma_H^2 = 0.5$ .

Observing optimal sensory weights is consistent with optimal integration, i.e. the integration behaviour that minimises variance in the multimodal estimates. However, optimal sensory weights might be observed in the absence of integration: as an alternative to integration, an observer may select one of the uni-modal estimates on each trial, rather than computing a weighted average [7, 8]. In the example above, the observer may select the visual estimate on 2/3 of trials, and the haptic estimate on 1/3 of trials. This ‘switching’ behaviour would produce the same mean response in bi-modal conditions as optimal integration, but with higher variance. Standard models predict that variance will be reduced in bi-modal, relative to uni-modal conditions under optimal integration, see, e.g. [1, 2, 6, 9]. For example in the visual-haptic size example, under optimal integration the predicted variance of the visual-haptic estimates,  $\sigma_{VH}^2$ , is given by  $\sigma_{VH}^2 = \frac{\sigma_V^2 \sigma_H^2}{\sigma_V^2 + \sigma_H^2}$ . In contrast, switching behaviour will result in variance that is at

least as large as the more reliable cue. For this reason, studies of multimodal integration generally determine (i) whether the sensory weights are optimal, given uni-sensory variability, and (ii) whether variability in the bi-modal estimates is reduced, relative to uni-modal estimates.

Recently, a number of studies have asked whether children show optimal integration of sensory cues, as indexed by (i) appropriate cue weighting and (ii) a reduction in variance, relative to single cue conditions. Gori and Burr [10] reported that optimal integration of multisensory information doesn’t appear until surprisingly late—at the age of around 10 years. In two visual-haptic tasks, younger children who were asked to judge object size or orientation relied on only one modality, and not necessarily the most reliable one. Other work has confirmed that children as old as 8 years fail to optimally integrate visual cues with movement-based information (proprioceptive and vestibular) for navigation [7], and another study suggests that optimal integration of auditory and haptic information does not occur until after age 11 [11]. Interestingly, this developmentally late integration is not limited to situations in which information must be combined from different sensory modalities: Nardini and colleagues reported similarly late integration for cues within a modality—optimal integration of two visual depth cues did not emerge until around age 12 [12].

The current study focuses on the developmental trajectory of audio-visual integration, using a straightforward counting task. The age at which optimal integration emerges for vision and audition is not yet clear. One previous audio-visual study with children aged 5–14 years and adults failed to find optimal integration at any age [13]. We employed a simple audio-visual task in which, on each trial, observers were presented with a number of beeps and / or flashes [14]. In separate blocks, they either reported the number of flashes, or the number of beeps. The task had the benefit of reduced memory and decisional demands, relative to previous studies that have used two-alternative forced choice designs. By comparing observers’ responses to different integration models we ask:

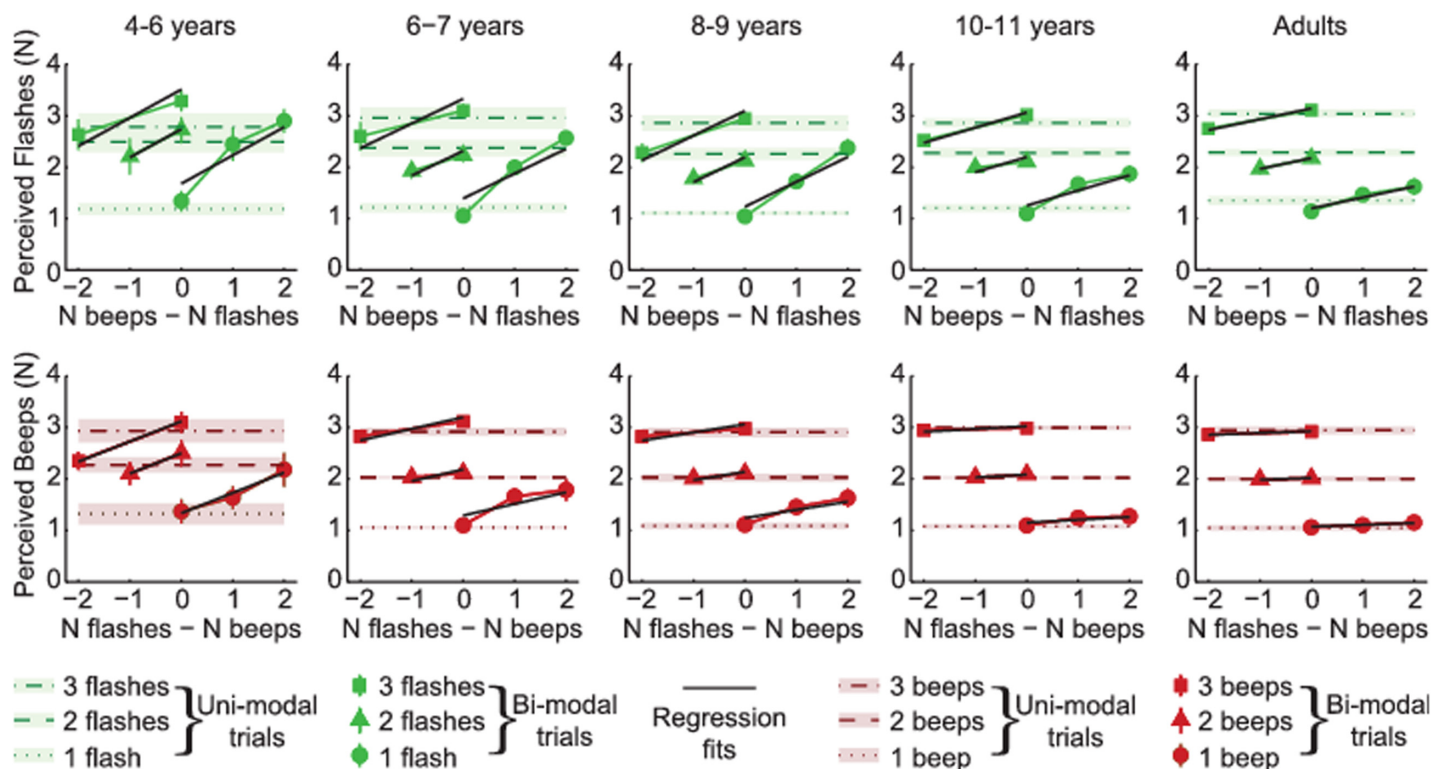
1. Do children show optimal integration of auditory and visual information? If so, from what age?
2. Is integration mandatory? In our task, observers are asked to report only one modality or the other, i.e. either beeps, or flashes, rather than the number of audio-visual events. We ask whether children do ignore the irrelevant (non-reported) modality, and we determine whether the strength of cross-modal interactions changes as function of age.

## Results

Data from 76 observers, split into 5 age groups, are summarised in Fig 1. Observers reported either the number of flashes (upper panels, green) or the number of beeps (lower panels, red). On some (uni-modal) trials, only flashes or only beeps were presented (Fig 1: horizontal dotted and dashed lines). These were intermingled with bi-modal trials in which both flashes and beeps were presented; the number of beeps and flashes could be the same or different.

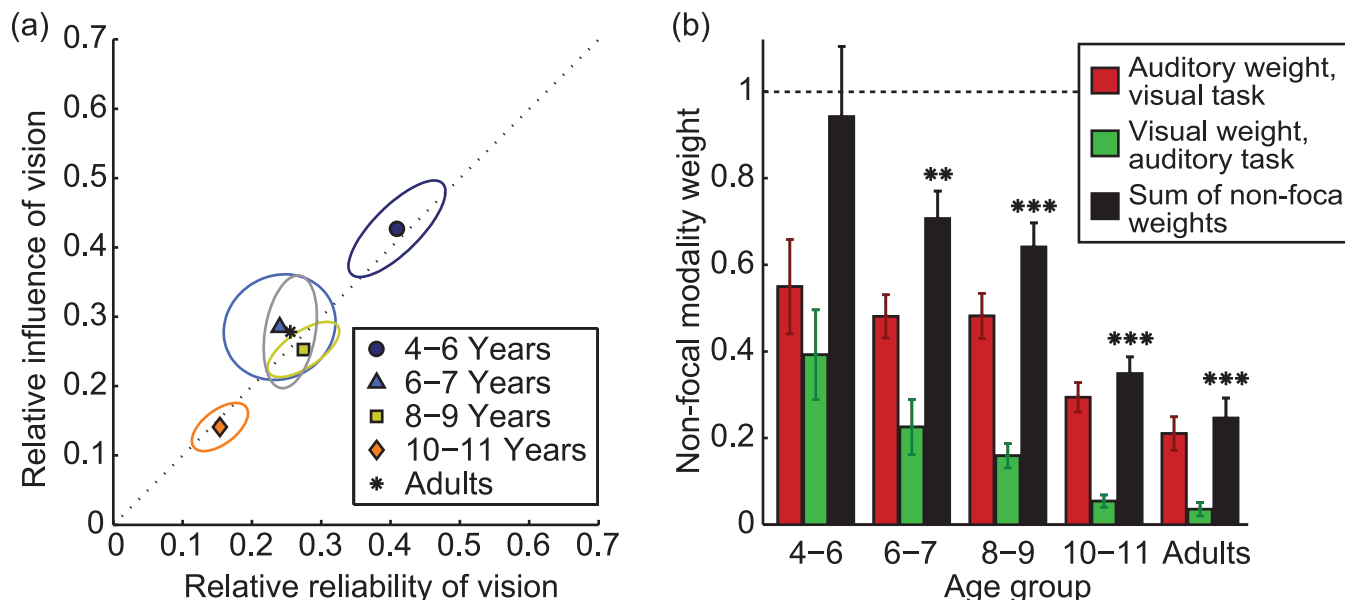
On each trial, observers were explicitly asked to report either the number of flashes or the number of beeps, whilst ignoring the other modality. In the absence of any cross-modal interactions, data for the bi-modal conditions would fall along horizontal lines: increasing or decreasing the number of events in the task-irrelevant modality would have no effect on subjects' responses. However, for all age groups, observers were unable to ignore the irrelevant, non-focal stimulus. The influence, or weight, of the task irrelevant cue can be quantified by the slopes of the regression lines shown in Fig 1. These regression lines were fit to the bi-modal data separately for each observer and modality: one slope parameter quantifies the influence of audition on vision (upper plots) and another quantifies the influence of vision on audition (lower plots).

For all groups, the reported number of flashes was significantly modulated by the number of simultaneously presented beeps (one-sample *t*-tests against 0, all  $p < 0.01$ ). Likewise, the reported number of beeps was significantly affected by the number of flashes (all  $p < 0.05$ ). However, it is clear from Fig 1 that the size of this cross-modal interaction depended on which modality was



**Fig 1. Summary of observers' data.** The reported number of flashes (upper row) and reported number of beeps (lower row). Each age group is shown by a separate column. Symbols give data from bi-modal trials. Horizontal dotted and dashed lines give responses on uni-modal trials, with the error bars / shaded regions giving  $\pm 1$  SE across observers for bi- and uni-modal data, respectively. The influence of audition on vision (top row) or of vision on audition (lower row) is characterised by the slope of the best fitting regression lines (black lines). Regressions were performed for individual observers and subsequently averaged (for illustration only).

doi:10.1371/journal.pcbi.1004865.g001



**Fig 2. Sensory weights.** (a) The relationship between the relative reliability of vision and the relative influence of vision, for each age group. It is clear that the relative reliability of vision predicts its relative influence for all groups. Covariance ellipses give 1SE around the mean. Note that the relative reliability of audition,  $rr_A$ , and the relative weight for audition,  $rw_A$  can be calculated in an analogous way such that  $rr_A = 1 - rr_V$  and  $rw_A = 1 - rw_V$ . Thus, the relative reliability of audition predicts the relative weight for audition in exactly the same way as for vision. (b) Sensory weights for non-focal modalities. Red bars give the weight given to (task-irrelevant) auditory information when reporting the number of flashes, while green bars give the visual weight when estimating the number of beeps. Black bars show the amount of integration, as quantified by the sum of the weights given to non-focal cues. Asterisks show the groups for which this sum is significantly less than 1 (one-sample *t*-tests).

doi:10.1371/journal.pcbi.1004865.g002

being reported: task-irrelevant beeps had a much larger effect on the number of reported flashes than vice versa. The mean weight for audition, when reporting flashes,  $w_{VA}$  (as defined by the regression coefficients) was 0.38. The mean weight given to vision when reporting the number of beeps ( $w_{AV}$ ) was significantly smaller: 0.14 (main effect of modality:  $F_{1, 71} = 57.6, p < 0.001$ ).

We can ask whether the relative influence of the two modalities is predicted by their relative reliability, in line with standard models of optimal integration. In general, audition is more reliable than vision for temporal tasks such as the one employed here [15], and is therefore given more weight when integrated with vision, when observers are required to make temporal judgements. We can estimate the reliability of visual and auditory signals from the variance of observers' responses on uni-modal trials. Across observers, audition was indeed more reliable than vision (mean variance for vision,  $\sigma_V^2 = 0.36$ , for audition,  $\sigma_A^2 = 0.18$ ).

Fig 2a summarises the relationship between the relative reliability of vision ( $rr_V$ ), as estimated from the uni-modal responses, and the relative weight of vision ( $rw_V$ ), as estimated from the bi-modal responses, where:

$$rr_V = \frac{1/\sigma_V^2}{(1/\sigma_V^2 + 1/\sigma_A^2)} \text{ and } rw_V = \frac{w_{AV}}{w_{AV} + w_{VA}},$$

where  $w_{AV}$  is the weight given to vision when reporting the auditory stimulus, and  $w_{VA}$  is the weight given to audition when reporting the visual stimulus. The relative influence of the two modalities is well predicted by their relative reliability for all groups (Fig 2a) and this relationship is significant across all individual observers ( $r = 0.54, p < 0.001$ ). It also reaches significance within the youngest ( $r = 0.77$ ), middle ( $r = 0.67$ ) and 4<sup>th</sup> ( $r = 0.57$ ) age groups (all  $p < 0.05$ ).

Interestingly, the strength of cross-modal interactions, as indexed by the sum of the non-focal cue weights, decreased substantially with age (main effect of age group:  $F_{4, 71} = 16.0$ ,

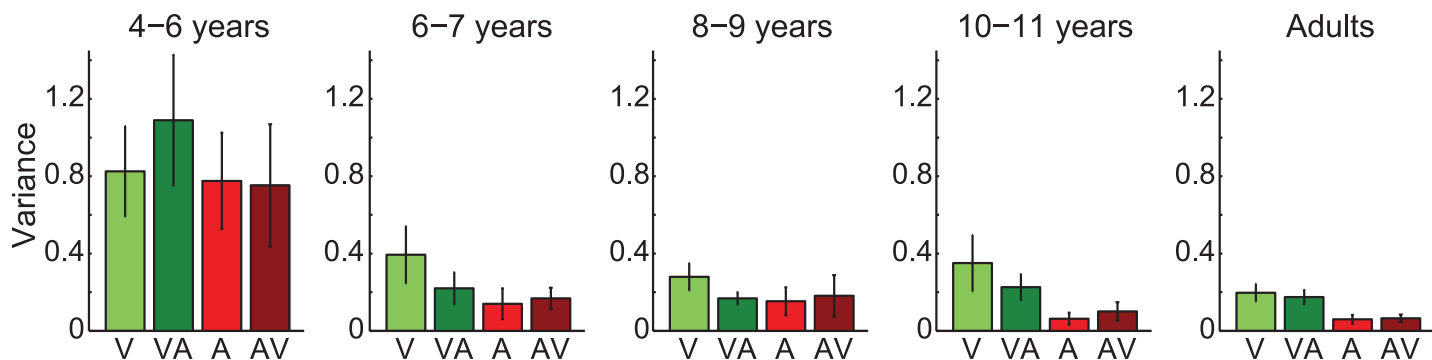


$p < 0.001$ ; see the black bars in Fig 2b). If participants used the same weights for visual and auditory information, irrespective of the task (report beeps vs. report flashes), then the sum of the non-focal weights would be 1 (dashed line, Fig 2b). This is the prediction under standard models of full integration, e.g. [1, 2]. Conversely, if participants gave more weight to the focal, task relevant modality whilst down-weighting the task irrelevant one, the average weight would be less than 1, and would be 0 if observers were able to ignore the task-irrelevant modality completely.

For the youngest observers only, the weights given to visual (red) and auditory (green) information did not vary according to whether observers were reporting beeps or flashes; the sum of the non-focal weights did not differ from 1 (see asterisks in Fig 2b). In other words, 4–5 year olds did not show any selectivity in reporting the focal, rather than non-focal modality. All other groups, however, showed partial cross-modal interactions: the reported number of flashes was dominated by visual information, whilst the reported number of beeps was dominated by auditory information, i.e. more weight was given to the focal modality, and the task-irrelevant modality was increasingly ignored as a function of age. Note that, in order to avoid floor or ceiling effects, the inter-stimulus interval (ISI) decreased with age (see Methods). A decrease in ISI (with all other factors constant) would be expected to increase uncertainty about the number of events, and thus increase the interaction between the sensory signals. For example, in the limiting case, the double flash illusion [16] will be eliminated with a large enough ISI. Note that the opposite pattern is seen here as a function of age—the influence of the non-focal cue *decreases* with age, despite the reduction in ISI. In other words, if the ISI had been more similar across age groups we would expect this age-related decrease in cross-modal interactions to be even larger.

Ernst [17] and Ernst & Di Luca [6] have described a variant of the standard Bayesian optimal integration model that allows partial integration, similarly to the behaviour described above. The model incorporates a ‘coupling prior’ that determines the strength of integration. This coupling prior represents the observer’s prior knowledge about the joint distribution of the two signals, i.e. the extent to which flashes and beeps tend to be correlated in the world and thus the probability that the visual and auditory signals contain redundant information. Under partial integration, variance in sensory estimates is no longer minimised, and in this sense, the integration strategy is no longer optimal. However, partial integration considers both precision (inverse variance) and accuracy (mean error). Under partial integration with potentially biased sensory signals, expected bias (inaccuracy) in the final estimates is reduced relative to full integration, whereas expected variance will be greater than under full integration. Partial integration can thus be described as optimal in the sense that it represents a balanced compromise between precision and accuracy [6]. Note that the standard, full integration model represents a special case of the partial integration model, in which signals are assumed to be accurate and the coupling prior is infinitely narrow (see Model 1: Partial Integration).

The partial integration model has previously provided a good account of adults’ cross-modal integration in a similar, discretized task [5]. Similarly to standard models of integration, the partial integration model predicts a reliability benefit (i.e. a reduction in variability) when information is combined across modalities. However, the magnitude of this benefit is proportional to the strength of integration, and the reliability of the bi-modal estimates will not necessarily exceed the reliability of *both* of the component uni-modal estimates. However, bi-modal reliability should always improve relative to estimates from the focal modality alone. In other words, it predicts that our observers will be more reliable in reporting the number of flashes when both visual and auditory information is available, than from vision alone. We should expect a similar reliability improvement for bi-modal, relative to uni-modal auditory estimates. Furthermore, if all our observers were integrating optimally (i.e. following the Bayesian partial integration model) then the youngest group would show the largest bi-modal improvement in



**Fig 3. Response variance as a function of age.** Lighter bars give response variance, averaged across uni-modal conditions for vision (V; green) and audition (A; red). Variance for bi-modal conditions is shown by darker bars for visual responses (VA; green) and auditory responses (AV; red). Error bars give  $\pm 1$ SE across observers.

doi:10.1371/journal.pcbi.1004865.g003

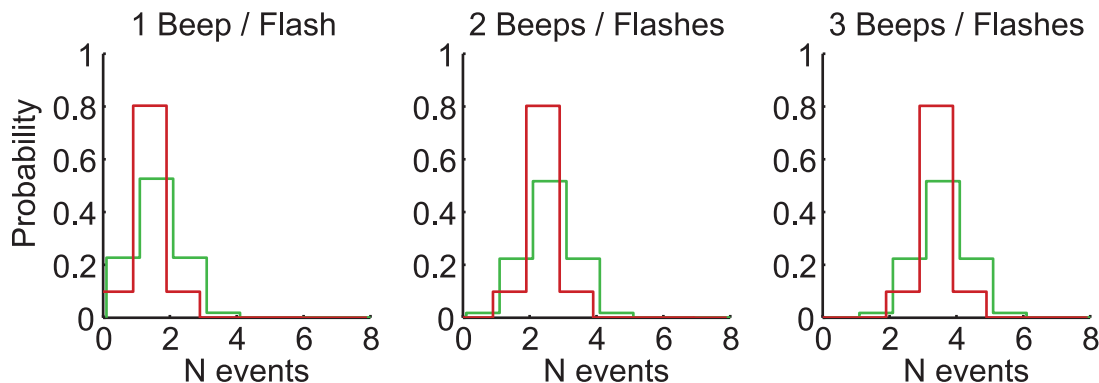
reliability, given that they show the strongest cross-modal interactions (black bars, Fig 2b). Fig 3 compares response variance for uni- and bi-modal estimates for the 5 age groups.

The youngest group shows no evidence of improved reliability in bi-modal, relative to uni-modal conditions for either vision or audition; their responses do not appear to reflect optimal sensory integration. All other groups show reduced variance in bi-modal, relative to uni-modal conditions for visual responses: compare dark and light green bars in Fig 3, this approached significance for 6–7 year olds and 8–9 year olds ( $p = 0.08$ ,  $p = 0.07$  for groups 6–7 and 8–9 years, all other age groups  $p > 0.1$ , from paired  $t$ -tests). However, there is little difference between uni- and bi-modal variance for auditory responses (red bars). Many older observers had little response variance in the uni-modal and bi-modal auditory conditions, and, given the discretized nature of the task, we must be somewhat cautious in using our observers' response variance as an accurate estimator of their underlying sensory noise. Furthermore, older observers gave very little weight to vision in bi-modal auditory conditions (as indicated by the small slopes in the lower plots of Fig 1) and thus the predicted improvement under optimal partial integration is very small. For these reasons, we evaluated whether the Bayesian partial integration model provides a good account of observers' behaviour by calculating the likelihood of each observer's data given this model. We compared it with two other candidate models in which observers do not integrate auditory and visual information, but instead (sub-optimally) switch between them, responding on each trial according to only visual or only auditory information.

## Modelling

Three models were compared: (i) Partial Integration, (ii) Focal Switching, and (iii) Modality Switching. Note that these were evaluated separately for each observer; averaged fits are shown in Figs 4–7 for illustration only. For all three models, because the number of events can take integer values only, noise distributions, and the resultant uni-sensory likelihoods were approximated by discretised Gaussians, i.e. the probability of a sensory estimate equal to  $x$ , is given by  $p(x) = a \exp(-\frac{(x - \mu)^2}{2\sigma^2})$ ,  $\{x \in \mathbb{Z} \mid \geq 0\}$  where  $a$  is a normalising constant. Noise distributions were centred on the true stimulus value,  $\mu$ , but differed in variance,  $\sigma^2$ , for vision and audition (see Fig 4).

In addition, alternative models were evaluated including the Causal Inference model [18, 19], models with logarithmic coding of number (corresponding to skewed likelihoods in linear space), and those that allowed likelihoods to be biased and / or to vary in reliability as a



**Fig 4. Uni-modal likelihoods.** The best-fitting uni-modal likelihoods for vision (green) and audition (red), averaged (for illustration only) across all observers; they have been slightly horizontally offset for visibility. The spread of the likelihood (i.e. the inverse reliability) is fixed as a function of the number of events, but differs between vision and audition. On average, vision was less reliable than audition ( $\sigma_V = 0.772$ ,  $SE = 0.052$ ;  $\sigma_A = 0.488$ ,  $SE = 0.058$ ).

doi:10.1371/journal.pcbi.1004865.g004

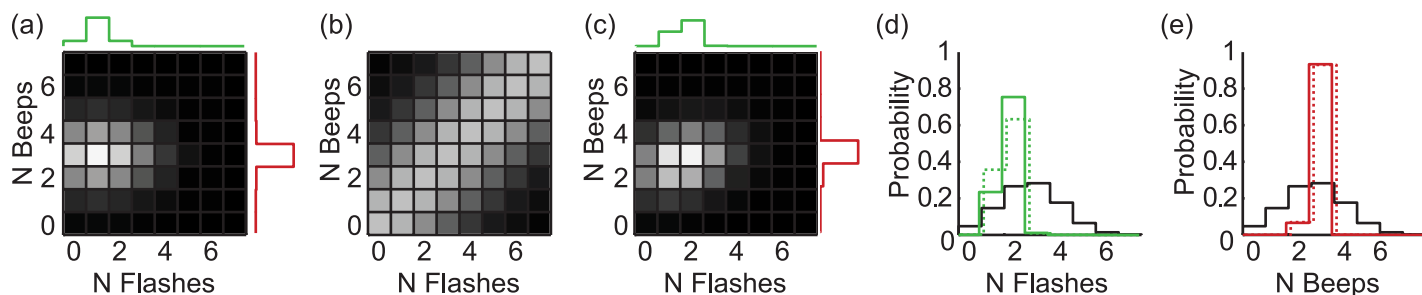
function of the number of events (beeps or flashes). These other models provided an inferior account of the data, as described in the supporting information file: [S1 Text](#).

The three models (Partial Integration, Focal Switching, Modality Switching) differ in the way that sensory information from vision and audition interact:

**Model 1: Partial integration.** The partial integration (PI) model [6, 17] is a variant of the widely used, standard Bayesian model in which sensory estimates are always fully integrated, e.g. [1, 2, 20]. A coupling prior determines the strength of integration: a flat prior results in no integration—visual flashes and auditory beeps are processed independently. Conversely, a 2D prior given by the unity line  $V = A$  is equivalent to the standard full integration model, which would result in observers always perceiving an equal number of flashes and beeps. [Fig 5](#) depicts an intermediate case—the best-fitting coupling prior, averaged across all observers whose responses followed the PI model.

The coupling prior is given by  $p(x_V, x_A) = a \exp(-\frac{(x_V - x_A)^2}{2\sigma_C^2})$ ,  $\{(x_V, x_A) \in \mathbb{Z}(x_V, x_A) \geq 0\}$

where  $x_V$  and  $x_A$  are visual and auditory sensory estimates,  $\sigma_C^2$  is the variance of the coupling prior and  $a$  is a normalising constant. All three models include a prior over the number of events—models without this prior were inferior (see supporting information: [S1 Text](#)). The



**Fig 5. The partial integration model.** (a) An example bi-modal likelihood, centred on 1 flash and 3 beeps. The uni-modal marginals are shown alongside. (b) The coupling prior, and (c) the bi-modal likelihood after combination with the coupling prior; the peak of the distribution has shifted towards  $V = A$ . (d) The visual marginal (dashed green) is multiplied by the prior over the number of events (black) to give the posterior probability distribution of the number of visual events (solid green). (e) The posterior distribution for audition (solid red), given the prior over the number of events (black). Note that to allow easy comparison across the three models, the prior over the number of events is shown as a sequential step after the coupling prior is applied and the subsequent marginals are estimated. The two priors could equivalently be combined and applied in a single step. All plots show the averaged model fit across the set of observers ( $N = 36$ ) who were best characterised by the PI model, as determined by comparing the likelihood of the data, given each of the three models.

doi:10.1371/journal.pcbi.1004865.g005



**Table 1. Fitted parameters for the three models.**

Parameter		Partial Integration (N = 36)		Focal Switching (N = 25)		Modality Switching (N = 15)	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
1	Visual variability, $\sigma_V$	0.5433	0.1749	0.9943	0.5329	0.949	0.5322
2	Auditory variability, $\sigma_A$	0.3441	0.2473	0.729	0.7767	0.4328	0.1832
3	Coupling prior spread, $\sigma_C$	5.31E+05 <sup>a</sup>	1.97E+06	—	—	—	—
	Focal weight, $p_F$	—	—	0.7579	0.2151	—	—
	Visual weight, $p_V$	—	—	—	—	0.2146	0.1534
4	Prior over n events, mean, $\mu_P$	2.6168	1.3401	2.7884	1.023	2.6467	0.8162
5	Prior over n events, std, $\sigma_P$	1.3705	1.1015	1.3723	0.6105	1.2695	0.8499
-Log likelihood		70.15	45.17	143.67	71.42	97.41	35.80

The values of the 5 parameters for each model that maximise the likelihood of observers' data. For each of the three models, the mean and standard deviation of the parameter values are shown, across all observers whose data were best fit by that model.

<sup>a</sup>The distribution of fitted values for this parameter was very skewed across observers; the median value was 0.79.

doi:10.1371/journal.pcbi.1004865.t001

prior distribution over the number of events,  $s$ , is given by  $p(s) = a \exp(-\frac{(s-\mu_P)^2}{2\sigma_P^2})$ ,  $\{s \in \mathbb{Z} | s \geq 0\}$  where  $\mu_P$  and  $\sigma_P^2$  are the mean and variance of the prior and  $a$  is a normalising constant. For the PI model, one consequence of this prior is that responses from uni-modal trials are more biased than responses from congruent bi-modal trials; the prior has more influence on uni-modal trials when the available sensory information is less reliable.

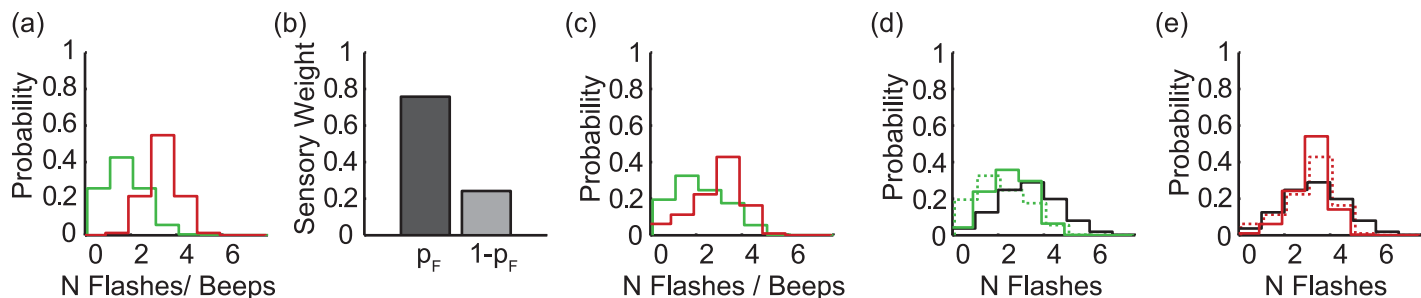
Following the standard Bayes' formulation, the posterior probability of a particular pair of visual and auditory estimates,  $\hat{s}_{V=i}, \hat{s}_{A=j}$ , is given by multiplying likelihoods and priors:

$$p(\hat{s}_V = i, \hat{s}_A = j) \propto \exp(-\frac{(i-\mu_V)^2}{2\sigma_V^2}) \exp(-\frac{(j-\mu_A)^2}{2\sigma_A^2}) \exp(-\frac{(i-j)^2}{2\sigma_C^2}) \exp(-\frac{(i-\mu_P)^2}{2\sigma_P^2}) \exp(-\frac{(j-\mu_P)^2}{2\sigma_P^2})$$

where  $\mu_V$  and  $\mu_A$  give the true number of flashes and beeps, respectively. On any single trial, the observer reports only one estimate: either the number of flashes, or the number of beeps. The posterior probability of a particular response is given by summing over all non-focal response estimates, i.e. finding the marginal probability distributions.

The PI model has 5 free parameters: (i) visual reliability, (ii) auditory reliability, (iii) width of coupling prior,  $\sigma_C$  (iv) mean  $\mu_P$  and (v) spread  $\sigma_P$  of the prior over the number of events (beeps or flashes). Fitted values of these parameters are summarised in [Table 1](#).

**Model 2: Focal switching.** Rather than integrating visual and auditory information, observers might stochastically switch between the two—sometimes responding according to the visual information, and sometimes according to audition [7, 8]. In the focal switching model, the distribution of responses depends on whether the observer is reporting the perceived number of flashes or beeps: observers select their estimate from the focal modality with probability  $p_F$  and the non-focal cue with probability  $(1-p_F)$ . Equivalently, this strategy produces a combined, bi-modal likelihood that is a weighted sum of the two uni-modal likelihoods: if the observer is reporting flashes, the likelihood of a particular estimate,  $i$ , is given by a weighted average of the probabilities of that estimate given the visual and auditory likelihoods:  $p(\hat{s} = i) = p_F p(\hat{s}_V = i) + (1 - p_F) p(\hat{s}_A = i)$ , where  $\hat{s}$  is the estimate from the combined, bi-



**Fig 6. Focal switching model.** Example (a) Uni-modal visual (green) and auditory (red) likelihoods. (b) On visual trials, the observer samples from the visual estimator with probability  $p_F$ , and from the auditory estimator with probability  $1 - p_F$ . On auditory trials, these probabilities, or weights are reversed. The resultant likelihoods are shown in (c). Similarly to the PI model, posterior distributions (d, e, solid lines) are created by combining these likelihoods with a prior (black) over the number of events. All plots show the averaged model fit, averaged across the set of observers ( $N = 25$ ) who were best characterised by the Focal Switching model.

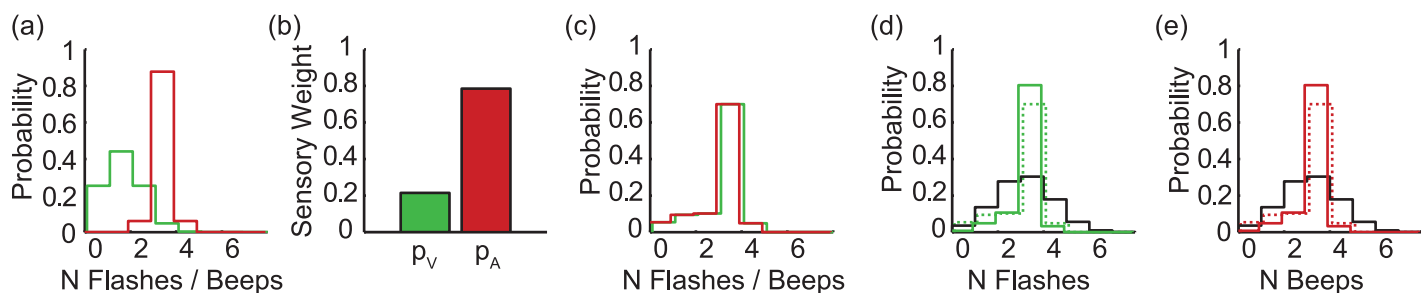
doi:10.1371/journal.pcbi.1004865.g006

modal likelihood, and  $\hat{s}_A$  and  $\hat{s}_V$  are estimates from the visual (focal) and auditory (non-focal) likelihoods.

The model has 5 free parameters: (i) visual reliability, (ii) auditory reliability, (iii) focal probability and (iv) the mean and (v) variance of the prior over the number of events (beeps or flashes). Note that on conflict trials (in which  $n$  flashes  $\neq$   $n$  beeps), such as the example shown in Fig 6, cross-modal interactions produce an *increase* in variance, relative to the uni-modal likelihoods.

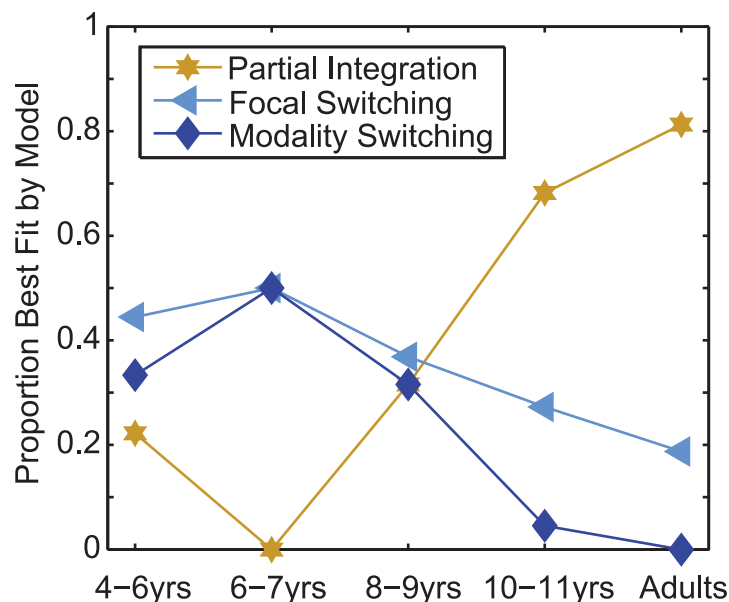
**Model 3: Modality switching.** In the Modality Switching model, observers again stochastically sample from auditory and visual information. However, in this model observers sample visual information with probability  $p_V$ , and auditory information with probability  $p_A$  (where  $p_V + p_A = 1$ ), irrespective of the focal modality. Under Modality Switching,  $p(\hat{s} = i) = p_V p(\hat{s}_V = i) + p_A p(\hat{s}_A = i)$ , where  $\hat{s}$  is an estimate from the combined, bimodal likelihood, and  $\hat{s}_V$  and  $\hat{s}_A$  are estimates from the visual and auditory likelihoods. In other words, for bi-modal conditions with a given number of flashes and beeps (e.g. 1 flash and 3 beeps, as shown in Fig 7), the model predicts the same pattern of responses, irrespective of whether the observer is reporting beeps or flashes. However, similarly to the Focal Switching model, when the visual and auditory estimates differ, bi-modal response variance will be increased, relative to variance in uni-modal conditions.

For older observers, responses were strongly modulated by the response modality, with more weight given to the focal cue (see Fig 2b). However, this was not the case for the youngest observers, who gave similar weight to vision and audition, irrespective of which was focal



**Fig 7. Modality switching model.** (a) Unimodal visual (green) and auditory (red) likelihoods. (b) On both visual and auditory trials, the observer samples the visual estimator with probability  $p_V$ , and the auditory estimator with probability  $1 - p_V$ . The resultant likelihoods (slightly offset for visibility) are shown in (c). Posterior distributions (d, e, solid lines) are created by combining these likelihoods with a prior (black) over the number of events. All plots show the averaged model fit across the set of observers ( $N = 15$ ) who were best characterised by the Modality Switching model.

doi:10.1371/journal.pcbi.1004865.g007



**Fig 8. The best-fitting model of audio-visual interactions, as a function of age.**

doi:10.1371/journal.pcbi.1004865.g008

(compare the leftmost pair of bars in Fig 2b). The Modality Switching model could, therefore, provide a good fit to younger observers' behaviour.

## Modelling results

For each observer and each model, the values of the 5 free parameters were found (Matlab: `fminsearch`) that maximised the joint likelihood of the observer's data across all uni-modal and bi-modal conditions. To avoid the problem of local minima, 288 iterations of the search were performed, making use of the University of Southampton's IRIDIS High Performance Computing facility, with initial values uniformly sampled from the multidimensional space of plausible parameters. Fig 8 shows how multisensory interactions change as a function of age. Observers in the two youngest groups were best described by the switching models. Children aged 8–9 years were evenly split, and by 10 years the majority of observers followed the partial integration model.

As the three different models have common parameters, (visual and auditory noise, and the mean and variance of the prior over the number of events) we can consider how the best fitting values of these change as a function of age. Recent work [21] suggests that children as young as 7 quickly learn the statistics of a stimulus set and bias their estimates towards the mean. In the current study, knowledge of the stimulus statistics would be represented within the prior over the number of events. As participants learn these statistics we might expect both the mean and standard deviation of the prior to decrease, as participants learn that only a small numbers of beeps and / or flashes are presented. The youngest group had the weakest prior (largest standard deviation) of all age groups; this parameter varied significantly as a function of age ( $F_{4, 71} = 3.03, p < 0.05$ ). Post hoc comparisons showed that the youngest group had a significantly weaker prior than the 6–7 and 8–9 year olds ( $p < 0.05$  from independent *t*-tests, after correction for multiple comparisons), no other comparisons were significant. Whilst the fitted prior for youngest group also had the largest mean, this did not vary significantly across groups. This provides some evidence that the youngest group may have been slower to learn the stimulus statistics.

As might be expected from the raw response variance shown in Fig 3, the fitted visual and auditory noise parameters also varied as a function of age ( $F_{4, 71} = 6.8$ ,  $p < 0.001$ ;  $F_{4, 71} = 6.9$ ,  $p < 0.0001$ , for  $\sigma_V$  and  $\sigma_A$ , respectively). Visual noise decreased monotonically with age, auditory noise decreased across each age group pair that shared a common stimulus ISI. Posthoc *t*-tests showed that, based on the fitted noise parameters, the youngest group was significantly more variable in both vision and audition than all other groups  $p < 0.01$ , after corrections for multiple comparisons). In the current paradigm ISI decreased with age (in order to broadly equate task difficulty). With a fixed ISI we would expect a larger increase in visual and auditory temporal acuity as a function of age.

## Discussion

A simple task was used to investigate the developmental trajectory of audio-visual integration. Importantly, we evaluated three different models of integration that together provide a good account of sensory integration behaviour at all stages of development. Key findings emerged:

1. Observers of all ages combined visual and auditory information using appropriate sensory weights, as determined by the relative reliability of visual and auditory signals (see Fig 2a). This contrasts with previous findings in which younger observers relied entirely on one sensory estimate, and not always the most reliable one [10]. In that study, children viewed the front of the object, while touching the reverse side. It has been suggested that the spatial offset and / or the fact that the active hand was obscured from view, prohibited cross-modal interactions in younger participants [22]; adults show reduced integration when sensory information is spatially offset [23]. Studies in which the sensory signals are aligned have found evidence for cross-modal interactions, i.e. switching behaviour, but not optimal integration, in younger children [7, 22].
2. Integration was automatic, and younger children were far less able to ignore task-irrelevant sensory information than older observers. This is in broad agreement with recent work which suggests that 7–10 year olds are unable to ignore irrelevant visual stimuli when performing an auditory spatial discrimination task [24]. In the current study, the youngest observers fully combined auditory and visual information, such that for bi-modal stimuli they perceived the number of flashes and the number of beeps to be the same. Older observers' behaviour was well modelled by a partial integration model in which the coupling of visual and auditory information was relatively weak.
3. Optimal integration, as indexed by increased reliability, emerged by 10 years—this is broadly in line with previous findings [7, 10, 12]. Before this age, model comparison suggests that observers do not integrate sensory information, but stochastically sample from each modality. Our finding suggests that optimal integration of auditory and visual signals develops at a similar age to integration across and within other modalities. Why did a previous study fail to find optimal audio-visual integration of temporal signals [13]? The study used a temporal bi-section task in which observers estimated which of two empty intervals was longer. Subsequent work has shown that for this type of task, with empty intervals, observers integrate auditory and visual information to optimally estimate the time points at the ends of the interval, rather than integrating duration per se [4]. With filled intervals, it is likely that optimal integration of duration estimates would be found with children aged 10 or so, as it is in adults [4].

Sensory integration has the potential to provide benefits for virtually all of our everyday activities—precision is improved by combining redundant information sources either within

or across modalities. An obvious question remains unanswered—why does this ability fail to appear until around 10 years? One proposed explanation is that the lack of integration is beneficial during early childhood, and facilitates recalibration [10, 12]. During this period of growth and sensory development, constant sensory recalibration is required in order to maintain accurate (unbiased) perceptual estimates. Recalibration requires the estimation of inter-sensory conflict—if this were only possible in the absence of integration, i.e. by keeping sensory estimates separate, then the developing sensory system might forego integration in favour of recalibration. The importance of cross-sensory interaction for sensory calibration and development is supported by studies in populations with sensory impairments—congenital visual deficits appear to have a detrimental effect on the precision of haptic estimates and vice versa [25, 26].

Studies with adult observers, however, suggest that integration and recalibration are not mutually exclusive. For example, when glasses distort the relationship between binocular disparity and depth, the perceptual system recalibrates accordingly, whilst continuing to integrate binocular disparity with other depth cues [27]. Moreover, the sensory system adapts relatively quickly (within hours) when sensory statistics change [17, 28–30]. In fact, recalibration and integration both rely on establishing the correspondence between signals—identifying which signals are redundant and only integrating (or recalibrating) when they arise from the same source. It might be that younger children find this correspondence hard to learn [31]. In the current study, observers were told to ignore one modality—adults were able to do this to a large extent, whereas children were sub-optimal in the sense that cross-modal influences were larger, even though vision and audition were discrepant on the majority of trials. A previous study also found that the effect of auditory beeps on the reported number of flashes was larger in children than adults [32]. However, that study did not use a design that allowed optimal integration to be evaluated.

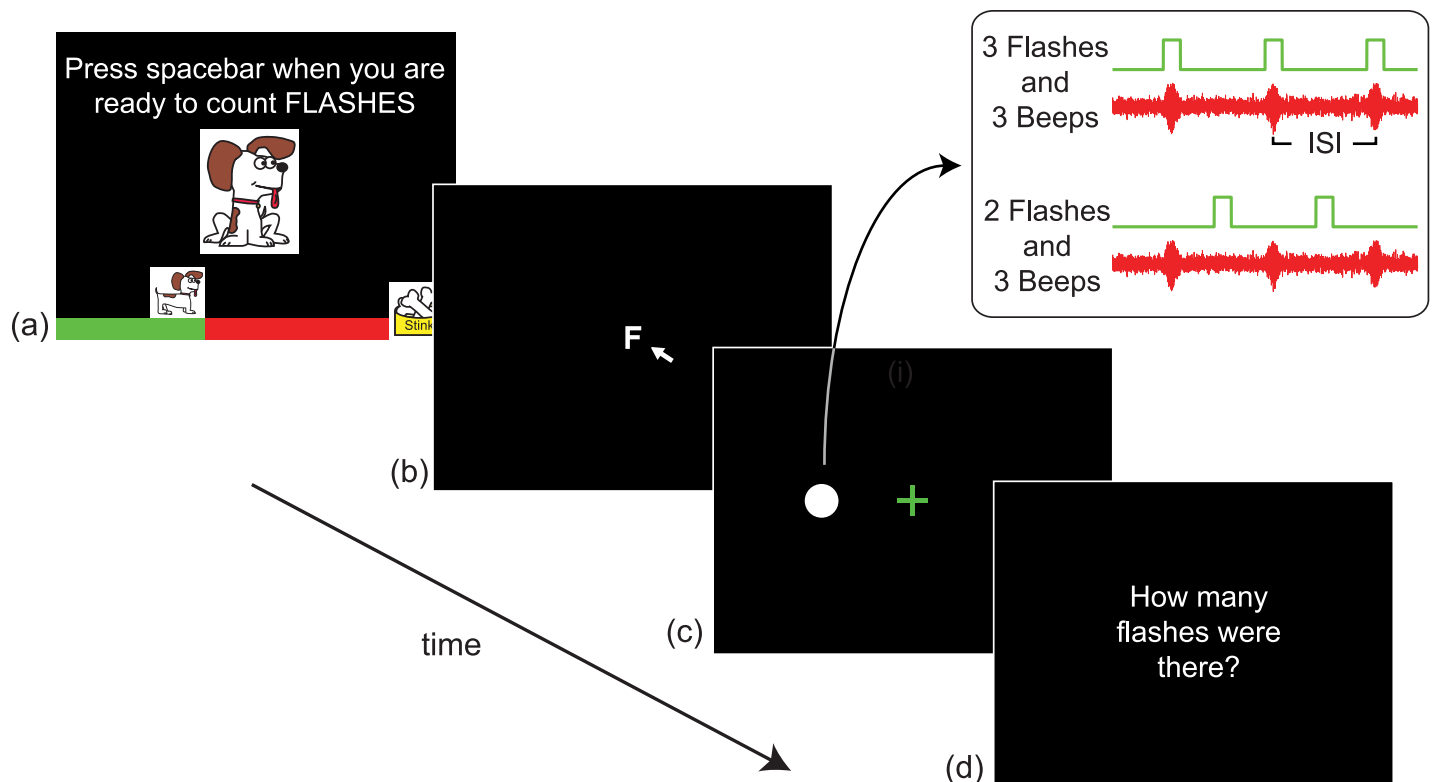
One recent study using a visual-proprioceptive reaching task did find some evidence of optimal integration, as evidenced by a reliability benefit, in children as young as 4–6 years [33]. However, this was only for the subset of observers who showed similar reliability for visual and proprioceptive estimates. Sub-optimal behaviour in other observers was attributed to inappropriate weighting. However, because the study did not include cue-conflict conditions, precise estimation of cue weightings was not possible. Our data suggest that, at least for the current task, the lack of integration shown by our observers was not due to a failure to weight the available signals appropriately.

In summary, the current work suggests that optimal integration does not emerge until around 10 years. Model comparison suggests that before that age, observers switch between the information provided by the two modalities, but do so in accordance with their relative reliabilities. This behaviour does result in responses centred on optimal values, but variance is larger than under optimal integration. In contrast with previous work, our younger observers did not rely on a single modality—in fact they were less able to ignore task-irrelevant information. Instead, they instead showed stronger, mandatory cross-sensory interactions than older observers.

## Methods

### Stimuli

Visual stimuli were white discs subtending 2.2 degrees of visual angle (dva) at the viewing distance of 45 cm with a luminance of 196 cd/m<sup>2</sup>. These were presented briefly (1 flash = 16.7msec), centred at 5.7dva to the left or right (randomly across trials) of a central fixation cross on an otherwise black screen. Auditory stimuli were presented via small speakers placed either side of the screen. These consisted of short beeps: 440Hz tones in a Gaussian temporal envelope of  $\sigma = 21$ msec. To reduce the reliability of the auditory stimuli, these beeps were embedded in



**Fig 9. Trial schematic.** (a) Instructions were shown at the start of each block of trials, and the voice of Stinker the dog gave the same instructions. A progress bar showed Stinker getting closer to his treats, as more trials were completed. (b) Either an 'F' or 'B' in the centre of the screen reminded the participant of the task. (c) After the letter was clicked, flashes, beeps or both were presented. The inset shows an example congruent trial (upper) and conflict trial (lower). (d) The participant was prompted to respond. An image of Stinker the dog appeared every few trials, with Stinker's voice offering words of encouragement or comments, e.g. 'You're great', or 'I'm hungry'.

doi:10.1371/journal.pcbi.1004865.g009

continuous white noise [4]. As in previous studies, sequences of flashes and beeps were temporally aligned [14], see Fig 9a.

The spacing between events (the ISI) was varied as a function of age group, as determined by pilot work. This was done to roughly equate task difficulty across groups such that floor or ceiling effects were avoided: pilot work showed that a fixed ISI across groups resulted in floor effects for the youngest group (such that the number of perceived events did not systematically increase as a function of the true number of events) and / or ceiling effects in the adult group (no response errors). For children in school years 1–3 (infant school; age 4–7 years) beeps and / or flashes were spaced by an ISI of 200msec. For junior school children (school years 4–7; age 7–11 years) the ISI was 167msec and for adults it was 117msec.

## Procedure

All participants were given detailed instructions, and completed 8 practice trials in which they reported flashes (4 trials) or beeps (4 trials). When counting flashes, subjects were told to ignore any beeps and vice versa. To help with motivation and concentration, participants were told that they needed to help Stinker the dog count beeps or flashes in order to get his treats. At the start of each block of experimental trials, Stinker appeared on the screen and instructed the participant to 'count the flashes' or 'count the beeps'.

Each trial began with an 'F' or a 'B' presented at the screen's centre to remind participants of the current task. To ensure fixation, participants were required to use the mouse to click this



letter. The letter then changed to a green fixation cross and the sequence of flashes and / or beeps was presented. Each sequence consisted of 0–3 beeps and 0–3 flashes, such that the trial could be uni-modal (only beeps or only flashes), bi-modal congruent (equal number– 1, 2, or 3 –of flashes and beeps) or bi-modal conflict (the number of beeps and flashes differed by 1 or 2). Uni-modal and bi-modal trials were randomly intermingled, but trials were blocked by focal modality (i.e. report flashes, or report beeps).

Participants gave their response on each trial by selecting the appropriate number on the keyboard (1–9); they were not told the maximum or minimum number of possible beeps or flashes. To keep the task duration within the concentration span of the child participants (approximately 20 minutes, based on pilot work), infant school children completed 140 trials (2 modalities: judging beeps or flashes) x 10 conditions (3 uni-modal, 7 bi-modal) x 7 repetitions. Junior school children completed 8 repetitions (160 trials) and adults completed 12 repetitions (240 trials).

## Participants

We report data from 76 observers (60 children, 16 adults). A further 5 children from the 4–6 age group were excluded who failed to complete the task and / or could not reliably count up to 3. To check for counting ability / task comprehension, we used leave-one-out cross validation to compare regression models for each observer's data, to ensure that the reported number of events across uni-modal and bi-modal congruent trials increased significantly as a function of the true number of events.

Children were *a priori* divided into four age groups, by splitting the infant and junior school children at the midpoint of each age range, such that all children within a group were given the same stimulus set (i.e. the same ISI). The resultant 5 groups were (i) '4–6 Years': Range: 4 years 9 months to 6 years 3 months,  $n = 9$ , 6 males (ii) '6–7 Years': Range 6yrs 5m to 7yrs 8m,  $n = 11$ , 7 males (iii) '8–9 Years': 7yrs 9m to 9 yrs 8m,  $n = 19$ , 9 males (iv) '10–11 Years': 9yrs 11m to 11yrs 5m,  $n = 22$ , 8 males and (v) 'Adults': Range 18–41 years,  $n = 16$ , 8 males). The study was approved by the ethics committee at the University of Southampton and all participants gave informed consent. Parents / guardians gave consent on behalf of their children and children also provided consent on the day of the experiment.

## Supporting Information

**S1 Text. Description and evaluation of alternative models.** The supporting information provides a description and evaluation of alternative models of observers' data. First, we apply the Causal Inference model [18], with three different decision rules [19]. Second, we test whether participants' responses are more consistent with log-based coding of number. Finally, we show that more complex models, such as those in which the likelihoods can be biased, or noise changes as a function of the number of events, do not provide a significantly better fit to the data.

(PDF)

## Acknowledgments

The author thanks Iona Kerrigan for conducting pilot work that informed the current study; Kristiana Newton for collecting data and Erich Graf for assisting in data collection and providing the voice for Stinker the dog. She also acknowledges the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

## Author Contributions

Conceived and designed the experiments: WJA. Performed the experiments: WJA. Analyzed the data: WJA. Wrote the paper: WJA.

## References

1. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 2002; 415(6870):429–33. PMID: [11807554](#)
2. Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*. 2004; 14(3):257–62. PMID: [14761661](#)
3. Helbig HB, Ernst MO. Optimal integration of shape information from vision and touch. *Experimental Brain Research*. 2007; 179(4):595–606. PMID: [17225091](#)
4. Hartcher-O'Brien J, Di Luca M, Ernst MO. The Duration of Uncertain Times: Audiovisual Information about Intervals Is Integrated in a Statistically Optimal Fashion. *Plos One*. 2014; 9(3).
5. Bresciani J-P, Dammeier F, Ernst MO. Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*. 2006; 6(5):554–64. PMID: [16881788](#)
6. Ernst MO, di Luca M. Multisensory perception: from integration to remapping. In: Trommershauser J, Kording K, MS L, editors. *Sensory cue integration* 2012. p. 224–50.
7. Nardini M, Jones P, Bedford R, Braddick O. Development of cue integration in human navigation. *Current Biology*. 2008; 18(9):689–93. doi: [10.1016/j.cub.2008.04.021](#) PMID: [18450447](#)
8. Ghahramani Z, Wolpert DM, Jordan MI. Computational models of sensorimotor integration. In: Morasso PG, V S, editors. *Self-organization, computational maps and motor control*. Amsterdam: Elsevier Press; 1997. p. 117–47.
9. Hillis JM, Watt SJ, Landy MS, Banks MS. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*. 2004; 4(12):967–92. PMID: [15669906](#)
10. Gori M, Del Viva M, Sandini G, Burr DC. Young children do not integrate visual and haptic form information. *Current Biology*. 2008; 18(9):694–8. doi: [10.1016/j.cub.2008.04.036](#) PMID: [18450446](#)
11. Petrini K, Remark A, Smith L, Nardini M. When vision is not an option: children's integration of auditory and haptic information is suboptimal. *Developmental Science*. 2014; 17(3):376–87. doi: [10.1111/desc.12127](#) PubMed WOS:000334693300005. PMID: [24612244](#)
12. Nardini M, Bedford R, Mareschal D. Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(39):17041–6. doi: [10.1073/pnas.1001699107](#) PMID: [20837526](#)
13. Gori M, Sandini G, Burr D. Development of visuo-auditory integration in space and time. *Frontiers in integrative neuroscience*. 2012; 6:77-. doi: [10.3389/fnint.2012.00077](#) PMID: [23060759](#)
14. Shams L, Ma WJ, Beierholm U. Sound-induced flash illusion as an optimal percept. *Neuroreport*. 2005; 16(17):1923–7. PMID: [16272880](#)
15. Burr D, Banks MS, Morrone MC. Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*. 2009; 198(1):49–57. doi: [10.1007/s00221-009-1933-z](#) PMID: [19597804](#)
16. Shams L, Kamitani Y, Shimojo S. Illusions—What you see is what you hear. *Nature*. 2000; 408(6814):788-.
17. Ernst MO. Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*. 2007; 7(5).
18. Koerding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal Inference in Multisensory Perception. *Plos One*. 2007; 2(9).
19. Wozny DR, Beierholm UR, Shams L. Probability Matching as a Computational Strategy Used in Perception. *Plos Computational Biology*. 2010; 6(8).
20. Adams WJ, Mamassian P. Bayesian combination of ambiguous shape cues. *Journal of Vision*. 2004; 4(10):921–9. PMID: [15595895](#)
21. Sciutti A, Burr D, Saracco A, Sandini G, Gori M. Development of context dependency in human space perception. *Experimental Brain Research*. 2014; 232(12):3965–76. doi: [10.1007/s00221-014-4021-y](#) PMID: [25183158](#)
22. Drewing K, Jovanovic B. Visuo-haptic Length Judgments in Children and Adults. *Haptics: Generating and Perceiving Tangible Sensations, Pt II*, Proceedings. 2010; 6192:438–44.
23. Gepshtein S, Burge J, Ernst MO, Banks MS. The combination of vision and touch depends on spatial proximity. *Journal of Vision*. 2005; 5(11):1013–23. PMID: [16441199](#)

24. Petrini K, Jones PR, Smith L, Nardini M. Hearing Where the Eyes See: Children Use an Irrelevant Visual Cue When Localizing Sounds. *Child Development*. 2015; 86(5):1449–57. doi: [10.1111/cdev.12397](https://doi.org/10.1111/cdev.12397) PMID: [26228618](https://pubmed.ncbi.nlm.nih.gov/26228618/)
25. Gori M, Sandini G, Martinoli C, Burr D. Poor Haptic Orientation Discrimination in Nonsighted Children May Reflect Disruption of Cross-Sensory Calibration. *Current Biology*. 2010; 20(3):223–5. doi: [10.1016/j.cub.2009.11.069](https://doi.org/10.1016/j.cub.2009.11.069) PMID: [20116249](https://pubmed.ncbi.nlm.nih.gov/20116249/)
26. Gori M, Tinelli F, Sandini G, Cioni G, Burr D. Impaired visual size-discrimination in children with movement disorders. *Neuropsychologia*. 2012; 50(8):1838–43. doi: [10.1016/j.neuropsychologia.2012.04.009](https://doi.org/10.1016/j.neuropsychologia.2012.04.009) PMID: [22569216](https://pubmed.ncbi.nlm.nih.gov/22569216/)
27. Adams WJ, Banks MS, van Ee R. Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience*. 2001; 4(11):1063–4. PMID: [11584290](https://pubmed.ncbi.nlm.nih.gov/11584290/)
28. Adams W, Graf E, Ernst M. Experience can change the 'light-from-above' prior. *Nature Neuroscience*. 2004; 7(10):1057–8. PMID: [15361877](https://pubmed.ncbi.nlm.nih.gov/15361877/)
29. Kerrigan IS, Adams WJ. Learning different light prior distributions for different contexts. *Cognition*. 2013; 127(1):99–104. doi: [10.1016/j.cognition.2012.12.011](https://doi.org/10.1016/j.cognition.2012.12.011) PMID: [23376295](https://pubmed.ncbi.nlm.nih.gov/23376295/)
30. Burge J, Ernst MO, Banks MS. The statistical determinants of adaptation rate in human reaching. *Journal of Vision*. 2008;8(4).
31. Ernst MO. Multisensory integration: A late bloomer. *Current Biology*. 2008; 18(12):R519–R21. doi: [10.1016/j.cub.2008.05.002](https://doi.org/10.1016/j.cub.2008.05.002) PMID: [18579094](https://pubmed.ncbi.nlm.nih.gov/18579094/)
32. Innes-Brown H, Barutcu A, Shivdasani MN, Crewther DP, Grayden DB, Paolini A. Susceptibility to the flash-beep illusion is increased in children compared to adults. *Developmental Science*. 2011; 14(5):1089–99. doi: [10.1111/j.1467-7687.2011.01059.x](https://doi.org/10.1111/j.1467-7687.2011.01059.x) PMID: [21884324](https://pubmed.ncbi.nlm.nih.gov/21884324/)
33. Nardini M, Begus K, Mareschal D. Multisensory Uncertainty Reduction for Hand Localization in Children and Adults. *Journal of Experimental Psychology-Human Perception and Performance*. 2013; 39(3):773–87. doi: [10.1037/a0030719](https://doi.org/10.1037/a0030719) PMID: [23163790](https://pubmed.ncbi.nlm.nih.gov/23163790/)