Multisensory uncertainty reduction for hand localization in children and adults

Marko Nardini[1], Katarina Begus[2] and Denis Mareschal[2]

[1] Dept of Visual Neuroscience, UCL Institute of Ophthalmology, London, EC1V 9EL, UK.

[2] Centre for Brain and Cognitive Development, Birkbeck, University of London,

London WC1E 7HX, UK

Corresponding author: m.nardini@ucl.ac.uk

Abstract

Adults can integrate multiple sensory estimates in order to reduce their uncertainty in perceptual and motor tasks. In recent studies children did not show this ability until after 8 years. Here we investigated development of the ability to integrate vision with proprioception to localize the hand. We tested 109 4- to 12-year-olds and adults on a simple pointing task. Participants used an unseen hand beneath a table to point to targets presented on top of the table to vision alone, proprioception alone, or both together. Overall, 7- to 9-year-olds and adults' points were significantly less variable given vision and proprioception together compared with either alone. However, this variance reduction was present at all ages in the subset of participants whose proprioceptive estimates were less than two times more variable than their visual. These results, together with analyses of cue weighting, indicate that all groups integrated vision and proprioception, but only 7-9-year-olds and adults consistently selected cue weights that were appropriate to their own single-cue reliabilities. Cue weights used at 4-6 and 10-12 years still allowed over half of participants at these ages to reduce their pointing variability. One explanation for poorer group-level cue weighting at 10-12 years is that this ages represents a period of relatively rapid physical growth. An existing Bayesian model of hand localisation did not describe either adults' or children's data well, but the results suggest future improvements to the model.

Keywords: multisensory, development, pointing, vision, proprioception

When multiple independent sensory information sources are available, it is possible to integrate them in a manner that reduces the uncertainty of the final estimate (Clark & Yuille, 1990). Recent studies with adults show integration of this kind to be an important and pervasive aspect of human perception and motor control (Ernst & Banks, 2002; Alais & Burr, 2004; reviews Ernst, 2005; Körding & Wolpert, 2006; Trommershäuser, Körding, & Landy, 2011). For example, in the study by Ernst & Banks (2002), participants integrated vision and touch to minimise their uncertainty when judging the heights of blocks. In this and other studies, human adults met the quantitative predictions of Bayesian 'ideal observer' models. Crucially, by integrating multiple senses human adults can attain a greater perceptual sensitivity than they can using any single sense alone. These findings raise a major question for human sensory-motor development: how and when do these sensory-motor optimizations emerge? Are they intrinsic to sensory systems, or do they need to be acquired or to emerge through tuning? Here we report new findings on the time-course of development of multisensory uncertainty reduction in humans using a task in which observers used vision and proprioception to localise their own hand.

The key prediction of Bayesian models is that by taking a weighted average of multiple sensory estimates (*e.g.* estimates for size of an object from both vision and touch), observers can reduce the uncertainty of the final estimate, and so carry out a task more precisely than when relying on single estimates. The key marker of successful cue integration in these studies is a reduction in the variability of sensory estimates, which can also be expressed as a reduction in psychophysical threshold. Developmental studies to date have found no evidence for adult-like uncertainty reduction until relatively late in childhood. Gori, Del Viva, Sandini, & Burr (2008) found that children below 8 years did not integrate visual and tactile information to reduce uncertainty in their estimates of object size or orientation. Nardini, Jones, Bedford, & Braddick (2008) found a similar result for visual and vestibular information for navigation: adults reduced

uncertainty in line with a Bayesian model, while children aged 8 years and below were best fit by a model using only single cues. Nardini, Bedford, & Mareschal (2010) studied integration for two distinct information sources within vision, a single sense: binocular disparity and texture gradient cues to 3D surface slant. Here it was not until 12 years that observers reduced the uncertainty of their judgments in line with a Bayesian model.

None of these studies found abilities to integrate sensory information to reduce uncertainty in children younger than 8 years. A conclusion from these studies is that adults' abilities to reduce uncertainty by integrating sensory cues are acquired over a relatively long time. The difficulties are likely to include learning to translate measures from different sensors into common units that can then be averaged, and taking account of gradually changing body size and sensory precision. Not integrating cues before the senses are well calibrated against each other could even be adaptive (Nardini *et al.*, 2010; Gori, Sandini, Martinoli, & Burr, 2010).

However, developmental studies have so far tested only a few of the many possible situations in which cue integration for uncertainty reduction could occur. Some studies included memory demands (Nardini *et al.*, 2008; Gori *et al.*, 2008) and/or two-alternative forced choice (2AFC) psychophysical procedures (Gori *et al.*, 2008; Nardini *et al.*, 2010). In these the multisensory decision process requires some additional resources (Ernst, 2008). Working memory develops significantly between 4 and 15 years (Gathercole Gathercole, Pickering, Ambridge, & Wearing, 2004; Case Case, Kurland, & Goldberg, 1982), so in tasks requiring information to be held in memory, some failures to integrate information could reflect immaturities in memory rather than in sensory integration. Psychophysical procedures rely on participants' correct identification of a perceptual difference described to them at the beginning of the study, and measures of threshold are vulnerable to participants forgetting or becoming confused about what is being looked for. In the present study we asked whether we might observe sensory integration

for uncertainty reduction at earlier ages than in previous studies by using a simple table-top hand localization task with no memory demands, in which participants simply point directly to targets.

We adapted a type of task used in several previous studies with both children and adults (*e.g.* Smothergill, 1973; von Hofsten & Rosblad, 1988; van Beers, Sittig, & Denier van der Gon, 1996). Participants seated at a table had to use the index finger of the unseen hand below the table to point as accurately as possible to targets presented on top of the table (see Figure 1). Information about the targets' positions was given by vision alone, proprioception alone (the subject's other index finger placed on top of the target without vision), or both. Proprioception-only and vision-only conditions show different patterns of constant error (biases to overshoot the target in azimuth and depth respectively; van Beers *et al.*, 1996). Integration of visual and proprioceptive estimates to reduce uncertainty predicts both a shift in constant error in the combined-cue condition relative to the single-cue conditions, and a reduction in variable error (uncertainty). We are particularly interested in whether children or adults will show reduced variable error (uncertainty) given both cues together *vs.* either alone. If uncertainty associated with the unseen hand under the table is also taken into account, data can be entered into an 'ideal observer' model predicting optimal performance (van Beers *et al.*, 1996). In the present study we modelled data for those participants for whom the assumptions needed to do this were met (see *Model* and *Results*, below).

Several previous studies have tested children's performance on pointing tasks of this kind, using a variety of target configurations and sensory conditions, and analysing a range of different measures (Smothergill, 1973; von Hofsten & Rosblad, 1988; Mon-Williams, Wann, & Pascal, 1999; King, Pangelinan, Kagerer, & Clark, 2010). Crucially, most did not specifically measure the variance of responses, and none found (using any measure) a statistically significant improvement in pointing precision when using vision and proprioception together compared with vision alone.

In general, in both children and adults, proprioception alone was markedly poor compared with vision alone. The potential of proprioception to improve pointing precision, when added to vision, may therefore have been small, making differences between vision alone and vision and proprioception together hard to detect.

Smothergill (1973) found a trend for better target localization given vision and proprioception together *vs.* vision alone from 6 years, but the difference was not statistically significant. von Hofsten & Rosblad (1988), testing 270 4- to 12-year-olds, found performance with vision and proprioception together largely indistinguishable from performance with vision alone. There were trends towards a reduction in constant error from 10 years, and a reduction in random error (a measure of dispersion, although not the same as variance) at 5 and 10 years, however there was no evidence that these patterns were statistically significant. Mon-Williams *et al.* (1999) found a trend for root mean square error (RMSE) of points to be lower for vision and proprioception together than with vision alone at 5 to 7 years but the difference was not statistically significant. King *et al.* (2010), testing 37 7- to 13-year-olds, did not report any age changes in precision of either single or combined cue estimates. However they did find a correlation between proprioceptive error and mean aiming point in an 'incongruent' two-cue condition, in which vision and proprioception specified different locations, a result consistent with the Bayesian principle of giving a higher weighting to a more reliable sense.

Across these developmental studies, robust advantages for adding proprioception to vision while pointing have been difficult to identify. These studies used 3 trials per condition (Smothergill, 1973), 8 trials per condition (von Hofsten & Rosblad, 1988), 5 trials per condition (Mon-Williams *et al.*, 1999), and 16 trials per condition (King *et al.*, 2010). By contrast a study with adults who were given 30 to 40 trials per condition found robust advantages for vision and proprioception together compared with either cue alone (van Beers *et al.*, 1996). To maximise the

sensitivity of our measures of the distributions of responses while keeping the task manageable with young children, we measured performance using 30 trials per condition. We fit each participant's distribution of responses in each condition with a two-dimensional error ellipse using a procedure that is robust to outliers (Rousseeuw, 1984). We measured the abilities of 4- to 12-year-olds and adults to localise targets on a table-top using vision alone, proprioception alone, or both together. Focusing on variable error, which is predicted to reduce given appropriate integration of estimates, we asked at what ages participants gained an advantage from having both vision and proprioception available, and compared performance with an 'ideal observer' model.

We predicted that like adults in previous studies, our adults would show significant reductions in variable error (uncertainty) in localising the hand when given both vision and proprioception compared with either cue alone. We were also interested in whether, using this simple and naturalistic pointing task, we might find evidence for uncertainty reduction *via* integration of sensory estimates at ages younger than in previous studies, in which this ability was not evident until the ages of 8-12 years (Gori *et al.*, 2008; Nardini *et al.*, 2008; Nardini *et al.*, 2010).

Method

Participants

Participants were 92 4- to 12-year-olds and 17 adults. Participants in the analysis were divided into four age bands: 4-6 years (mean age 5.4, sd 0.8 years, n=33; 17 male); 7-9 years (mean age 8.5, sd 0.7 years, n=19; 8 male), 10-12 years (mean age 11.6, sd 1.0 years, n=36; 16 male), and adult (mean age 24.1, sd 4.1 years, n=17; 6 male). Four more participants with outlying scores were excluded from analysis (see Analysis). Both left- and right-handed participants were included; proportions of left-handed participants by group were 9/33 (27%), 3/19 (16%), 1/35 (3%) and 2/17 (12%). Proportions of left-handers were highest in the youngest group, in line with gradual development of handedness in childhood (McManus *et al.*, 1988). The study was approved by the local research ethics board. Children were recruited from local schools and from the lab's database of volunteers; adults were recruited from a university database of volunteers. All participants or their parents gave informed written consent for their participation.

Apparatus and Design

Figure 1a illustrates the layout of three targets that participants attempted to localise. The targets were indicated visually by different 1cm stickers of sea creatures (fish, shell, starfish), and tactually by identical 1.5 mm ball bearings. Participants were seated at a 80 (wide) x 60 (deep) cm table with height 69 cm (Figure 1b). The table's top surface was made of transparent plastic, but for all conditions but one vision through the table's surface was blocked with a layer of cardboard (Figure 1c). To position participants of different ages comparably, chairs of different heights were used together with cushions in order to bring the height of the table-top in line with the mid-point between each participant's shoulder and elbow. Starting positions for each hand were indicated on the table top; for children the positions were as in Figure 1b, for adults the starting

position for the right hand was moved further forward to a comfortable distance. For all participants the body midline was aligned with the right-hand edge of the central rectangle of card (see Figure 1b-c). The origin of Figure 1a was 20 cm left of and 15 cm above the point on the table edge in line with the body midline. Relative to this origin, the targets' (x, y) coordinates in cm were (1.6, 5.4), (6.3, 1.8) and (7.4, 8.4).
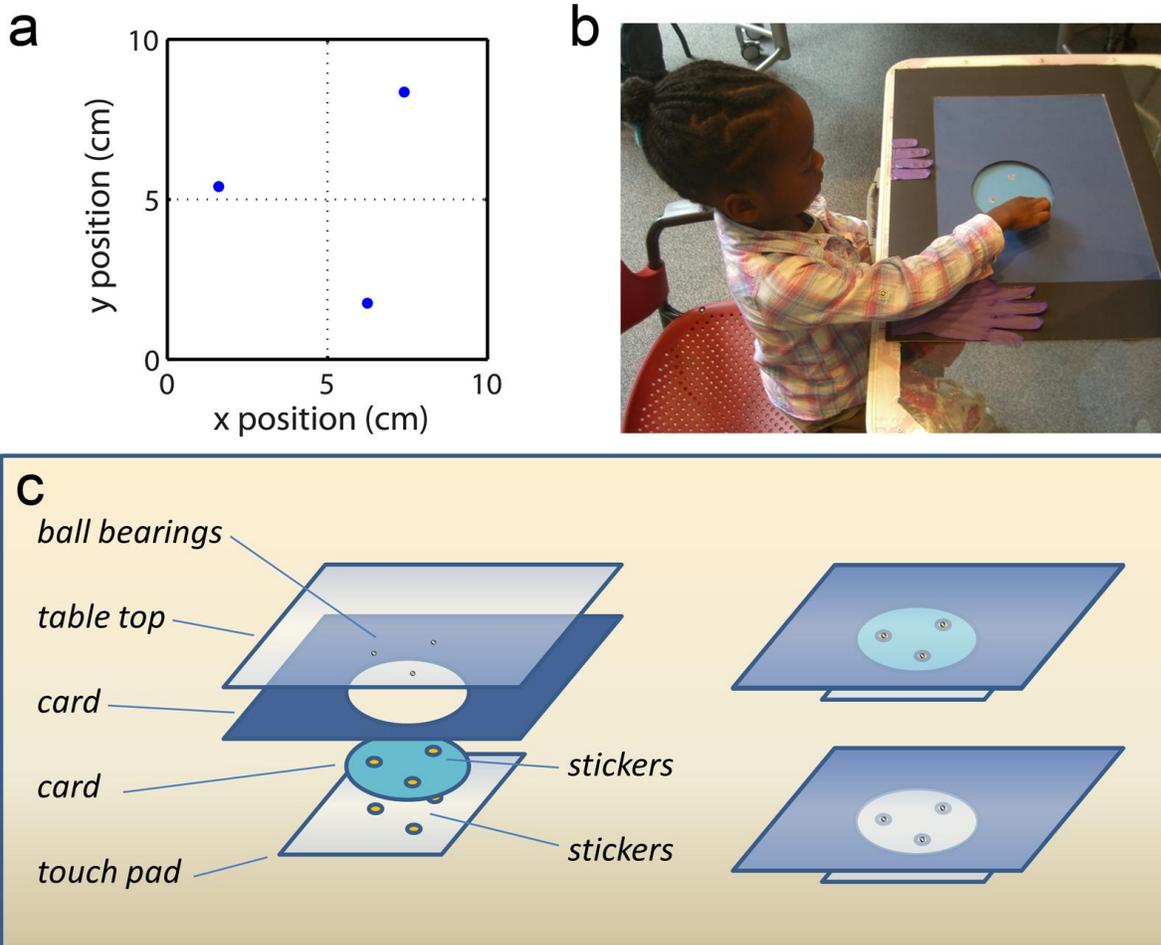
Figure 1. a. Spatial layout of the three targets. b. Illustration of the setup with a participant pointing to a target with the unseen under-table hand using both vision of the target and proprioception of the above-table hand's position (condition VP). c. Illustration of the apparatus, which could be used in two configurations – with the card circle, in which case the stickers on the card but not the responding hand were seen, or without the card circle, in which case the stickers on the transparent touchpad and also the responding hand could be seen. The ball bearings on the table's surface could be seen in all conditions including vision, and felt in all conditions including proprioception.

A transparent 23 cm (wide) x 30.5 cm (long) touch-pad sensor (KeyTec, Garland, TX) was mounted below the table top to record responses. In order for the touch-pad to localise the finger-tip accurately participants wore one finger of a latex glove on the responding finger with a 1.5 mm ball bearing fixed to the end. The touchpad acts as a mouse interface and so returns 'clicked' locations as screen (pixel) coordinates to a PC via USB. Screen coordinates in pixels were converted to table coordinates in cm using a calibration routine written in Matlab in which four corners enclosing the calibrated area are identified.

The task was to use the index finger of the left hand, under the table-top, to point to the touchpad at locations indicated above the table-top. In all conditions but one, the responding hand under the table could not be seen. Vision of the hand responding below the table-top was allowed in one control condition ("full vision"). This was possible because both the table-top and the touch-pad were transparent. However in all conditions but the one in which vision of the hand was allowed, a 14 cm (diameter) cardboard circle inserted between the touchpad and the table top blocked vision of the hand below. Vision through the rest of the table surface was blocked with a cardboard sheet in all conditions (Figure 1c).

We measured localization accuracy below the table under conditions varying the information on top of the table: either vision alone was provided (condition V), or proprioception alone (condition P), or vision and proprioception together (condition VP). Each block of 9 trials included one trial of every combination of condition (3) and target location (3), in a random order. There were 10 blocks in total, giving a total 30 trials for each of the V, P, and VP conditions. A final 15 trials (5 at each of the 3 targets, in random order) were run in a control condition in which the hand responding below the table could be seen ("full vision"). Results from this condition were analysed to check that participants of all ages understood the task and correctly distinguished between the three different targets.

Procedure

At the beginning of each trial, the participant put both hands on the indicated starting positions (see Fig 1b). The right hand was placed on a marked position on the table top in front of them, just to the right of the participant's midline, in a position that allowed the elbow of the right arm to rest comfortably on the table top. The left hand was placed on a marked starting position at the edge of the table top, 20 cm left of the participant's midline.

In all conditions, participants had to match the position of a target indicated on top of the table using the index finger of the unseen left hand, touching the touch pad underneath the table. The conditions differed in the information provided about the target position.

Condition V. The task was to match the position of the unseen left index finger on the underside of the table with the position of a visual target. The experimenter indicated which of three targets (fish, shell, starfish) should be matched. Participants moved their left hand from the starting position to match the target, while the right hand remained in its starting position.

Condition P. The task was to match the position of the unseen left index finger on the underside of the table with the also unseen index of the right finger (a purely proprioceptive target). Participants were first instructed to close their eyes with their hands in the starting positions. For young participants the experimenter also lowered a blindfold. The experimenter then guided the participant's right index finger to the position of the target, indicated by a ball bearing on the surface of the table, directly above the visual target. When the right hand was in its position, participants moved their left hand to match the position of the index finger underneath the table to the one on top. Participants had to keep their eyes closed throughout the whole trial.

Condition VP. The task was to match the position of the unseen left index finger on the underside of the table with the position of a target that was available to both vision and proprioception. As in condition V, the experimenter indicated which of three targets (fish, shell, starfish) should be matched. Participants moved the right index finger to the target. They were then able both to see the target through the transparent table top as in condition V, and to sense its position via the right hand's index finger position as in condition P.

Full vision condition. At the end of the test trials, the cardboard circle covering the touch pad was removed (Fig 1c), so that participants were able to see the targets as well as their hand underneath the table. Participants then had to match the target positions in another 15 trials.

In all conditions, participants were asked to match the target positions as accurately as possible, using the tip of their index finger. Positioning the finger on the touch pad triggered two sounds. The first sound signalled that they had touched the surface of the touch pad. Participants were able to move their finger along the surface of the touch pad, while staying in contact with it, for as long as they wished, in order to position the finger precisely where they thought the target was located. Most participants first made a relatively fast movement towards the position of the target and then adjusted the position of the finger with some slower and smaller movements. Once they were certain about the position, they removed their finger-tip from the touchpad, which triggered another sound. The position at which the participant removed the finger tip from the touch pad was recorded as the chosen position. Any trials on which the pad was touched accidentally, or the participant reported that they had accidentally removed the finger too early, were repeated.

At the beginning of the experiment participants first practiced touching the touchpad in order to get used to the mode of response - in particular, being able to adjust the position until

releasing the pad in order to record the final position. After familiarization with the touchpad all participants also received a total of 6 practice trials (2 from each condition) before starting the main study.

To motivate younger participants to complete the study, we rewarded them with a sticker every 15 trials. We also used sound to make the task engaging: a bubbling sound throughout the study, evoking an undersea ambience consistent with our 'sea creature' targets, and computer game-like sounds when the touch pad was touched or released, or when a sticker was won, every 15 trials.

Analysis

We first collapsed responses across the three target locations. To do this we subtracted each target's location from each pointing response's location. After this translation each pointing response is expressed as an error relative to its target, in *x* (azimuth) and *y* (depth) axes on the table-top. Patterns of pointing error to differently positioned targets can differ across a table-top space (e.g. van Beers *et al.*, 1996), therefore in our configuration targets were placed close together in order to minimise any such differences.

Next, we fit a bivariate normal distribution to each participant's points in each condition. The distribution of responses can be described using five parameters: the *x* mean, the *y* mean, the *x* variance, the *y* variance, and the *x-y* covariance. These measures are strongly influenced by outlying values, which in our setup could easily be generated by responding errors. We therefore estimated these parameters using the Minimum Covariance Determinant (MCD) procedure (Rousseeuw, 1984), which is robust to the undue influence of outliers. We used the FAST-MCD algorithm (Rousseeuw & van Driessen, 1999) as implemented in the Libra toolbox for Matlab (Verboven & Hubert, 2005), with the assumption of 1% aberrant (outlier) values (i.e. a value of

0.99 for the *alpha* parameter). The mean (SD) numbers of trials per condition (out of 30) excluded as outliers in each age group were 1.5 (0.4) at 4-6 years, 1.3 (0.4) at 7-9 years, 1.2 (0.4) at 10-12 years, and 1.1 (0.4) in adults.

Error ellipses describing distributions of points were plotted for visualization. For the main statistical analyses we combined errors across *x* and *y* directions to obtain a single measure of *variable error* – the sum of *x* and *y* variances – and of *constant error* – the length of the vector (*i.e.* the Euclidean distance) between the mean aiming point and the target.

We checked that no individual participant showed a variable error greater than 0.5 cm in the control "full vision" condition. This confirmed that all participants distinguished between the different targets and attended to the task. We also screened variable error scores in the main three conditions (*V, P, VP*) for extreme outliers, excluding participants whose variable error in any condition exceeded the third quartile plus three times the interquartile range of scores in that condition in their age group (Tukey, 1977). Since in a normal distribution fewer than 1 in 800,000 observations would meet this criterion, the most likely interpretation for an outlying value is that either the participant's visuo-motor processing or their strategy for the task is atypical. We excluded four extreme outliers: one in each of the two youngest age groups and two in the 10-12 year group. Data from the remaining participants were entered into the analysis and model.

Pointing tasks without vision of the hand typically show constant errors (biases), which differ in their directions from the target under vision-only and proprioception-only conditions (van Beers *et al.*, 1996; van Beers, Sittig, & Denier van der Gon, 1999). In adults, the trend is for proprioceptive estimates to overshoot the target in azimuth (*x*), and for visual estimates to overshoot it in depth (*y*). There is evidence that constant errors in *VP* conditions fall on a line between *V* and *P* constant errors, with a degree of curvature dependent on the shapes of the *V*

and *P* error ellipses (van Beers *et al.*, 1999). The relative proximity of the mean *VP* location to each of the mean *V* and *P* locations should correspond to the degree of reliance on ("weighting" for) one cue as compared with the other when both cues are provided. To gauge the extent to which responses in the combined-cue condition *VP* were influenced by vision as compared with proprioception, we calculated a measure of weighting for vision in the *VP* condition. Weighting for vision was calculated as the relative proximity of the *VP* distribution's centre to the mean centres of the *V* and *P* distributions (see also Nardini *et al.*, 2008):

$$w_V = \frac{1/d_V}{1/d_V + 1/d_P} = \frac{d_P}{d_V + d_P}$$
<div align="right">Eq. 1</div>

Where d_P and d_V are the distances from the *VP* centre to the *P* centre and from the *VP* centre to the *V* centre respectively. Possible scores on this measure range from 1 (full weight for vision, zero weight for proprioception) to 0 (zero weight for vision, full weight for proprioception). This measure is consistent with the assumption that all three points lie on a straight line, and is also consistent with the assumption that all three points lie on a curved line (van Beers *et al.*, 1999), if the curved line is approximated as two straight lines, one joining the *V-VP* centres, the other joining *VP-P*.

Model

We compared performance in the combined-cue condition *VP* with that predicted by an ideal observer model based on performance in the single-cue conditions *V* and *P*. The major prediction is for a reduction in variance when combining visual and proprioceptive information. We used the model of van Beers *et al.*, 1996, which proposes that:

1. Variability of pointing in the task has parallel contributions from variability in proprioception of the hand below the table (which is constant) and variability in proprioception

and/or vision of the hand and targets above the table (which varies across conditions *V*, *P* and *VP*). These two sources of variance are additive:

$$\sigma_V^2 = \sigma_{pbelow}^2 + \sigma_{vabove}^2 \qquad\qquad \text{Eq. 2}$$

$$\sigma_P^2 = \sigma_{pbelow}^2 + \sigma_{pabove}^2 \qquad\qquad \text{Eq. 3}$$

$$\sigma_{VP}^2 = \sigma_{pbelow}^2 + \sigma_{vpabove}^2 \qquad\qquad \text{Eq. 4}$$

2. The condition *VP* presents in parallel the same two independent sources of sensory information about target location that are available singly in conditions *V* and *P*. By taking a weighted average of these two estimates it is possible for observers to reduce the variance of the estimate in condition *VP* relative to those in either single-cue condition. The lowest possible ("optimal") variance is achieved by weighting each estimate in inverse proportion to its own variance (see Clark & Yuille, 1990; Ernst, 2005). The predicted combined-cue variance for an observer following this optimal strategy is:

$$\sigma_{vpaboveOPT}^2 = \frac{\sigma_{vabove}^2 \sigma_{pabove}^2}{\sigma_{vabove}^2 + \sigma_{pabove}^2} \qquad\qquad \text{Eq. 5}$$

Since our behavioural measure also includes the below-table variance (Eq. 4), we are not able to measure *VP* variance associated with the above-table cues directly. In order to assess whether participants have achieved the optimal variance reduction in their responding, we need to subtract out the below-table variance common to all conditions from the responses.

To estimate the below-table contribution to all conditions, we followed the proposal of van Beers *et al.*, 1996, who considered the variability of each hand in the proprioception-only

condition P (in which participants touch the two fingers together with eyes closed) to be approximately equal, *i.e.*

$$\sigma^2_{pbelow} \approx \sigma^2_{pabove} \qquad \text{Eq. 6}$$

From this and Eq. 3 we can estimate the variability of the under-table hand that is common to all conditions as

$$\hat{\sigma}^2_{pbelow} = \frac{\sigma^2_P}{2} \qquad \text{Eq. 7}$$

To obtain our estimates of variability associated with the above-table hand, we subtract this value from the measured total variability in each condition:

$$\hat{\sigma}^2_{vabove} = \sigma^2_V - \frac{\sigma^2_P}{2} \qquad \text{Eq. 8}$$

$$\hat{\sigma}^2_{pabove} = \sigma^2_P - \frac{\sigma^2_P}{2} = \frac{\sigma^2_P}{2} \qquad \text{Eq. 9}$$

$$\hat{\sigma}^2_{vpabove} = \sigma^2_{VP} - \frac{\sigma^2_P}{2} \qquad \text{Eq. 10}$$

Substituting these into Eq. 2 and simplifying, we can now predict optimal performance based on the measured variability:

$$\hat{\sigma}^2_{VPOPT} = \sigma^2_P - \frac{\sigma^4_P}{4\sigma^2_V} \qquad \text{Eq. 11}$$

It is difficult to confirm directly that the assumption of approximately equal above and below table variances is correct. In van Beers *et al.* (1996), a control experiment showed that variability of the above and below table hands was not significantly different, although to test this it was necessary to use conditions that also included vision of the target. Adult observers in van

Beers *et al.* (1996) tended to perform better than the model in condition *VP*, which should not be possible. The model also cannot be fit to participants whose *P* variance is greater than their *V* variance by a factor of 2 times or more. This is because subtracting half the *P* variance as required by Eq. 8 would produce a negative value for the estimated above-table *V* variance. Despite these potential limitations, the model of van Beers *et al.*, 1996 and van Beers *et al.*, 1999 remains the only quantitative model to date of information integration for visual-proprioceptive hand localization. We assessed the model's fit to the children's and adults' data from our study to provide a further test of it, and to inform the development of new models.

Statistical analysis

We report means by age and condition for parameters of error ellipses, variable error and constant error, as well as mean weighting for vision in the VP condition. For the subset of participants whose data can be entered in the model, we report these same values as well as the model's prediction for *VP* variable error given integration of visual and proprioceptive estimates as measured in conditions *V* and *P*.

The model in its present form does not make predictions for *VP* constant error, for which we would need to be able to quantify the different constant errors of the two hands in each condition. While we report both variable and constant error, our primary interest in the analysis is in variable error. Integration (averaging) of estimates predicts reductions in variable error, and the model also allows a quantitative test of this. By contrast, integration (averaging) of estimates is only expected to reduce constant error if existing constant errors given single cues are in opposite directions of the target and so are cancelled out by integration, which need not be the case.

Our major analysis is a set of planned comparisons between variable error in condition *VP* and variable errors in each of conditions *V* and *P* (paired t-tests). This tests the specific prediction, from integration by weighted averaging, that condition *VP* will have lower variable error than *both* conditions *V* and *P*. A significant improvement relative only to one condition would be consistent with the possibility that in the combined-cue condition observers are simply relying on the better of the two single cues. Keeping our significance criterion at the conventional 5% level for each t-test makes this a highly conservative test for integration, as the probability of making a Type I error on both comparisons – i.e. concluding that the *VP* mean is lower than *both* *V* and *P* means when it is not - is 0.0025 (0.05^2).

For an analysis of effects of age and their variation across conditions, we use ANOVAs to analyse changes in both variable error and constant error. Variable errors in the full vision conditions are entered into a separate ANOVA as they were an order of magnitude lower than those in the main conditions. We also use univariate ANOVA to assess changes with age in the weighting measure, and 95% confidence intervals for mean weighting to compare weights with those predicted by full weight for vision, zero weight for vision, and equal weights.

In a secondary analysis including only the subset of participants whose data can be modelled, we report all the same descriptive statistics, but our analysis focuses only on two sets of planned comparisons (paired t-tests) of major interest. First, as in the main analysis, we compare variable error in condition *VP* with variable error in both *V* and *P* to assess whether there is evidence for cue integration by weighted averaging. Second, we compare actual *VP* variable error with model-predicted *VP* variable error, to assess whether performance at any age diverges from that predicted by the model. Here the familywise Type I error (alpha) rate for comparisons including condition *VP*, which comprise (1) the pair of tests that need to be significant for evidence of integration (alpha = 0.0025, i.e. 0.05^2; see above), and (2) the test of *VP* vs the model

prediction (if alpha = 0.05), would be 0.0524. To obtain a familywise alpha rate of 0.05 we reduce alpha (i.e. the threshold for significance) for *VP vs.* model t-tests to 0.0476.

Results

Figure 2a plots 75% error ellipses using for each age group the mean of individual participants' parameters (*x* and *y* means, *x* and *y* variances and *x-y* covariance) describing distributions of points under conditions *V*, *P*, and *VP*. Constant errors are present at all ages, with centres of all distributions displaced some way from the target, and *VP* distributions centred intermediate to the centres of *V* and *P* distributions. The *P* ellipse is generally more elongated in the *x* direction and the *V* ellipse in the *y*, consistent with greater uncertainty in azimuth for proprioception and depth for vision from this perspective (van Beers, Wolpert, & Haggard, 2002). Sizes of error ellipses reduce with age, indicating development of precision in hand localization. The *P* ellipse is the largest at all ages, showing greatest localization uncertainty with proprioception alone. The *V* and *VP* ellipses are smaller, and at ages 7-9 years and in adults the VP ellipse is also visibly smaller than the *V*, consistent with an advantage for having both vision and proprioception available compared with vision alone. Figure 2b plots 75% error ellipses for the control "full vision" condition. Here error ellipses are very small and well centred at all ages. The sizes of the ellipses also decline with age, particularly between 4-6 and 7-9 years.
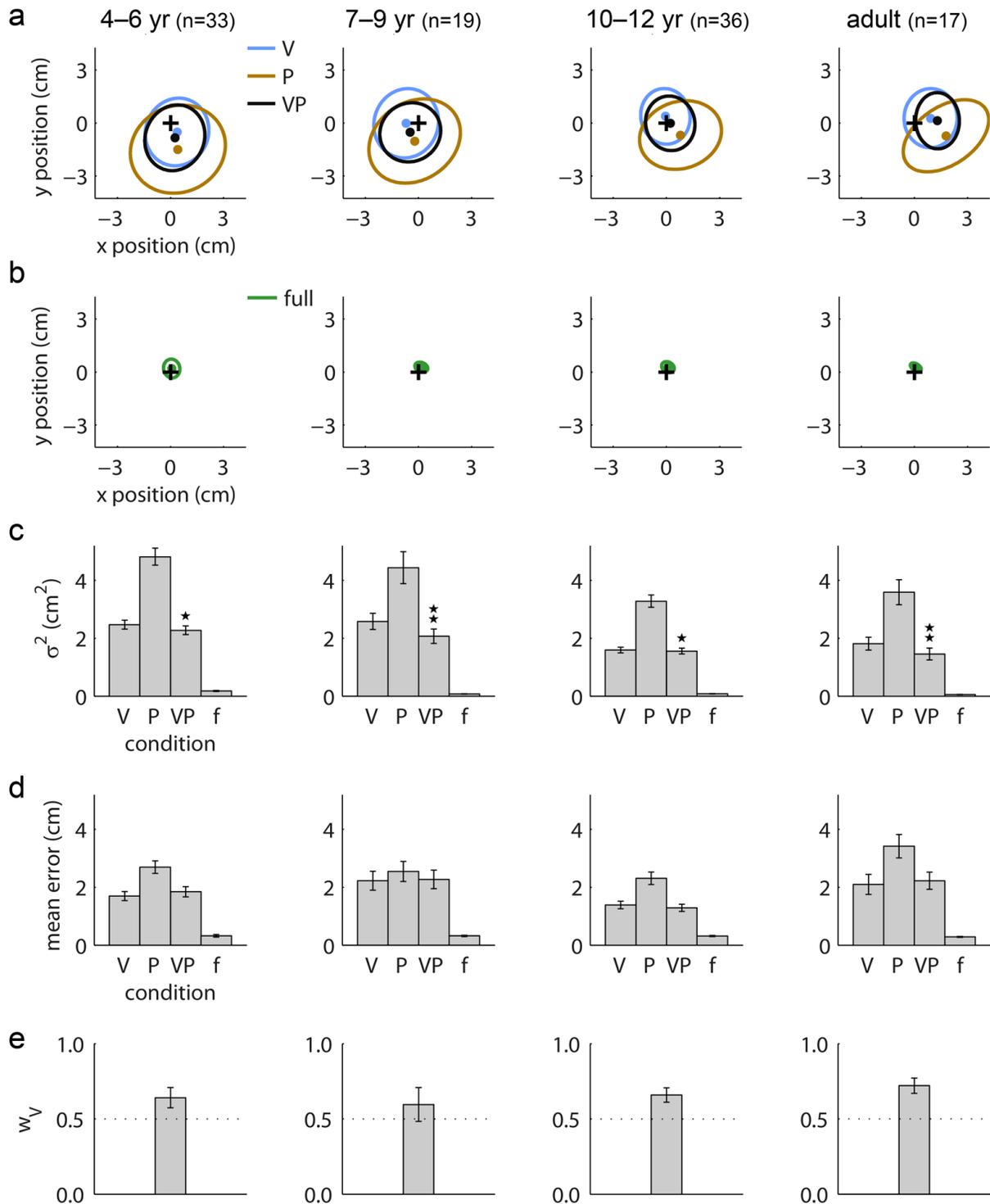
Figure 2. *Results by age group for all participants.* **a-b.** *Mean shapes and positions of error ellipses for conditions V, P, VP, and "full vision".* **c.** *mean total variable error (s.e. bars) by condition; *, VP differs from P at the 5% level on paired t-test; **, VP differs from P and VP differs from V at the 5% level on paired t-test.* **d.** *mean total constant error (s.e. bars) by condition.* **e.** *mean weighting for vision (95% c.i. bars).*

Figures 2c-d plot mean total variable error and mean total constant error respectively, corresponding to the sizes and positions of the ellipses shown in Fig 2a-b. Our key question is whether *VP* variance is reduced compared with either *V* or *P*. We compared *VP* with *V* and *VP* with *P* in each group using planned paired t-tests. Performance significantly better than with either single cue is required to show that participants cannot be simply relying on the best single cue in the two-cue case. Results of the t-tests are in Table 1. In Figure 2c, "*" indicates that one of these comparisons was significant, while "**" indicates that both were significant. *VP* responses were significantly less variable than *P* responses at all ages. *VP* responses were also significantly less variable than *V* responses in the 7-9 year and adult groups, but not in the 4-6 or 10-12 year groups (Figure 2c). This shows evidence for ability to use combined cues in order to improve pointing precision from 7-9 years, but a seeming absence of this ability at 10-12 years, despite an overall reduction in variability, before its re-emergence by adulthood.

	VP vs. P			VP vs. V		
	t	d.f.	p	t	d.f.	p
4 – 6 yr.	9.5	32	<0.001*	1.3	32	0.202
7 – 9 yr.	6.0	18	<0.001*	2.7	18	<0.02*
10 – 12 yr.	9.0	35	<0.001*	0.5	35	0.639
adult	8.2	16	<0.001*	3.8	16	<0.01*

Table 1. *Results of two-tailed paired t-tests comparing variable error in the VP condition with P and V conditions, * difference significant at the 5% level.*

Participants' variable errors (Figure 2c) were entered into a mixed ANOVA with *condition* (*V*, *P*, *VP*) as a within-subjects factor and *group* as a between-subjects factor. The assumption of sphericity was not met, so degrees of freedom were adjusted using the Greenhouse-Geisser correction. The main effect of *condition* was significant, $F(1.4, 139.1) = 195.7$, $p<0.001$, as was the main effect of *group*, $F(3, 139.1) = 7.8$, $p<0.001$. The *condition* x *group* interaction was not significant, $F(4.1, 139.1) = 1.9$, $p = 0.10$. This shows that variability changed with the type of sensory information available, and with age. However over the course of the overall change in variability with age, the pattern of performance did not change significantly across conditions.

At all ages, variable error from the "full vision" condition was an order of magnitude lower than that in other conditions (see Fig 2c, including error bars). "Full vision" variable errors were entered into a separate univariate ANOVA. The effect of *group* was significant, $F(3, 101) = 14.7$, $p < 0.001$. Although very small even at the youngest age (4-6 year mean = 0.18 cm), variable error on the "full vision" condition declined with age (adult mean = 0.05 cm).

Constant errors (Figure 2d) for conditions *V*, *P*, and *VP* were entered into a mixed ANOVA. The main effect of *condition* was significant, $F(1.4, 144.9) = 59.6$, $p<0.001$, as were the main effect of *group*, $F(3, 144.9) = 3.5$, $p<0.05$, and the *condition x group* interaction, $F(4.3, 144.9) = 2.9$, $p<0.05$ (Greenhouse-Geisser corrected F values). These effects reflect the results that mean pointing locations changed with age and also differed across conditions. The change with age

(Figure 2a) represents a gradual drift from below the target to right of the target. A consistent pattern across conditions (Figure 2a) is for *V* to be localised further forward than *P*, and *VP* to be intermediate. This pattern, coupled with the drift, means that which condition happens to be closest to the target, *i.e.* to have the lowest constant error, changes with age (Figure 2a). The lowest constant errors were those of the 10-12 year group in condition *VP*, while the highest were those of the adult group in condition *P*.

Constant error from the "full vision" condition was entered into a separate univariate ANOVA. There was no significant change with age, $F_{(3, 101)} = 0.2$, p=0.90 (see Figure 2b) .

Figure 2e plots the mean weighting for vision by age, in terms of the mean distance of centres of *VP* distributions to the centres of *V* and *P* distributions. A score of 0.5 would indicate an equal weighting for vision and proprioception. Scores in all age groups were higher than 0.5, indicating greater weighting for vision than for proprioception in the *VP* condition. The error bars represent 95% confidence intervals. From these it is clear that mean weightings at all ages were significantly different from either 1 (full weight for vision) or 0 (full weight for proprioception). Confidence intervals also exclude 0.5, showing that vision was weighted higher than proprioception, at all ages except 7-9 years. A univariate ANOVA found no significant changes in weighting with age, $F_{(3, 101)} = 1.5$, p=0.22. Overall these results indicate a propensity to be guided more by vision than by proprioception on combined-cue trials, with a similar and stable vision weight of between 0.60 (at 7-9 years) and 0.72 (in adults) for the whole age range from 4 years to adulthood.

Note that the weighting measures (Figure 2e) do not directly correspond to those apparent from the centres of error ellipses in Figure 2a. The configurations of ellipse centres plotted in Figure 2a come from averaging the positions of these centres across all observers,

24

whereas the weighting calculations in Figure 2e come from first calculating each observer's weight based on their own specific configuration of ellipse centres, then averaging these weights across observers.

We considered possible effects that our groups' uneven sizes could have had on results from the main analysis: significantly reduced variable error given *VP* vs either single cue at 7-9 years and in adults (N = 19, 17), but no significant reduction at 4-6 and 10-12 years (N = 33, 36); Table 1 and Fig 2c. Since groups in which effects were *not* found were larger than those in which they were, it seems unlikely that lack of power accounted for failures to detect differences in these groups – although power could in principle be low in these groups if variability across participants was high. To investigate this issue we calculated power for *VP vs. V* comparisons, which were not significant at all ages (*VP vs. P* comparisons were highly significant at all ages; see Table 1 and Fig. 2c).

In the two groups in which the *VP vs. V* comparison was significant at the 5% level, the power to detect the magnitude of *VP-V* difference that was detected (based on each group's *N* and SD of the *VP-V* difference), was 0.71 (7-9 years) and 0.96 (adults). In these groups variable error was reduced by 24.7% and 24.5% respectively given *VP* rather than *V* alone. We asked what the power was in the two groups in which a significant *VP vs. V* difference was *not* detected to detect a similar level of difference, a 25%, reduction in *VP* compared with *V*, given the *N* and the SD of the *VP-V* difference in each group. At 4-6 years the power for detecting a 25% improvement in VP vs V was 0.96; at 10-12 years the power was >0.99. In the latter group a difference as small as 17% could be detected with power 0.9. These calculations indicate that there was equal or greater power to detect differences between conditions in those groups that did not show significant differences. Lack of power is unlikely, therefore, to account for failures to find differences at these ages.

Model

Participants who met the criterion that condition *P* total variable error was less than two times the condition *V* variable error were entered into a separate analysis, with results plotted in Fig. 3. This analysis included just over half (57 of 105) participants, in similar proportions in each age group (compare *n* in Figure 2a and Figure 3a). The aim of this analysis was to compare *VP* performance with an ideal observer model (van Beers *et al.*, 1996) that can only be fit to data meeting this criterion. Figures 3a-b plot error ellipses for the subset of participants included in this analysis. The shapes and layouts of the distributions are similar to those in the main analysis (Fig 2), except that the sizes of *V* and *P* ellipses are more similar, a consequence of excluding observers in whom these were highly discrepant. Unlike in the main analysis, *VP* ellipses are visibly smaller than both *V* and *P* at all ages.
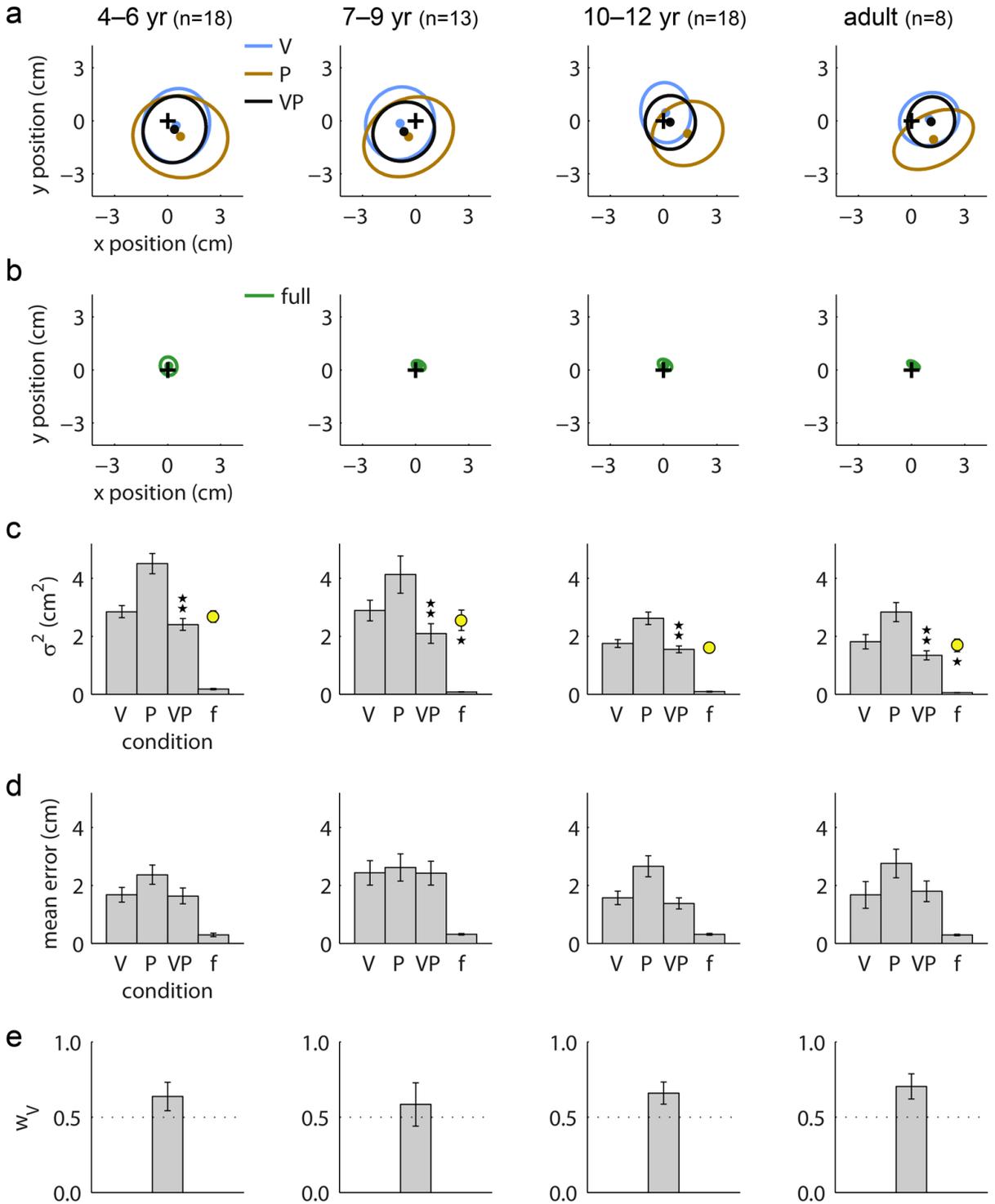
Figure 3. *Results by age group for participants meeting the criterion that P variable error < 2 * V variable error.* ***a-b.*** *Mean shapes and positions of error ellipses for conditions V, P, VP, and "full vision".* ***c.*** *mean total variable error (s.e. bars) by condition; *, VP differs from P at the 5% level on paired t-test; **, VP differs from P and VP differs from V at the 5% level on paired t-test. Circle: mean model prediction for VP variable error (s.e. bars); *, model differs from actual VP variable error at the 5% level on paired t-test.* ***d.*** *mean total constant error (s.e. bars) by condition.* ***e.*** *mean weighting for vision (95% c.i. bars).*

Figure 3c plots mean total variable error by age and condition as in the main analysis, as well as model predictions (circles) for ideal observer *VP* performance. The pattern of results is similar to that in the main analysis, except for a reduced discrepancy in *V vs. P* variance at all ages, and a greater advantage for *VP vs.* the best single cue (*V*) at all ages.

As in the main analysis we used planned paired t-tests to compare *VP* with *V* and *VP* with *P*. We also compared *VP* with the model prediction. Results of the t-tests are in Table 2. In Fig 3c, "**" above the *VP* mean for every age group shows that both the comparisons of *VP* with *V* and *VP* with *P* were significant at all ages. That is, in this subset of participants, points in the *VP* condition were significantly less variable than those in either of the single-cue conditions at every age. This shows evidence for ability to use combined cues in order to improve pointing precision at 4-6 years, 7-9 years, 10-12 years, and in adults, in participants whose single-cue variances were not too discrepant (*P* variance < 2 * *V* variance).

In Fig 3c, "*" below the model prediction (circle) indicates that predicted and actual *VP* performance were significantly different. *VP* performance did not differ significantly from the model prediction at 4-6 or 10-12 years, but did differ from the prediction at 7-9 years and in adults. These differences were also significant at the 4.76% level after correction for multiple comparisons (see Methods / Analysis). In both these groups participants performed *better* than the ideal observer model. This result replicates the findings of van Beers *et al.*, 1996 with adults, and indicates that there are important limitations to the model. One possibility that could explain both the 'impossible' results of some observers showing *P* variance > 2 * *V* variance, and the abilities of other observers to outperform the model, would be that condition *P* consistently over-estimates proprioception-only variance (van Beers *et al.*, 1996). This possibility is discussed further below.

	VP vs. P			VP vs. V			VP vs. model		
	t	d.f.	p	t	d.f.	p	t	d.f.	p
4 – 6 yr.	8.0	17	<0.001*	2.4	17	<0.04*	1.6	17	0.133
7 – 9 yr.	5.2	12	<0.001*	3.6	12	<0.01*	2.3	12	<0.04*
10 – 12 yr.	6.7	17	<0.001*	3.2	17	<0.01*	0.7	17	0.480
adult	6.8	7	<0.001*	3.3	7	<0.02*	2.9	7	<0.03*

Table 2. *Results of paired t-tests comparing variable error in the VP condition with P and V conditions and with the model prediction, for participants meeting the criterion that P variable error < 2 * V variable error. * difference significant at the 5% level.*

Relationships between single cue variability, two-cue variability and weighting

Our results indicate group-level "u-shaped" development of the ability to benefit from combined *vs.* single cues to localise the hand (Figure 2). However, there is evidence for this ability at all ages in those observers whose localization variability using proprioception alone is less than 2 times than when using vision alone (Figure 3). Thus, there seems to be a relationship between individuals' discrepancies in visual and proprioceptive variabilities and their abilities to benefit from combining these cues. To understand this relationship we carried out a further analysis relating measures of single-cue discrepancy to measures of combined-cue benefits and cue weighting.

Figure 4a plots the relationship between ratios of single-cue variances (P/V; a measure of discrepancy in the reliabilities of single cues) and combined- to single-cue variances (VP/V; a measure of the benefit given combined cues *vs.* vision alone) at each age. Three lines indicate the predictions of different cue combination rules: relying on the worst single cue (red line), the best single cue (green line), or using the Bayesian model (black curve; predictions cannot be made past the vertical P / V = 2 line).
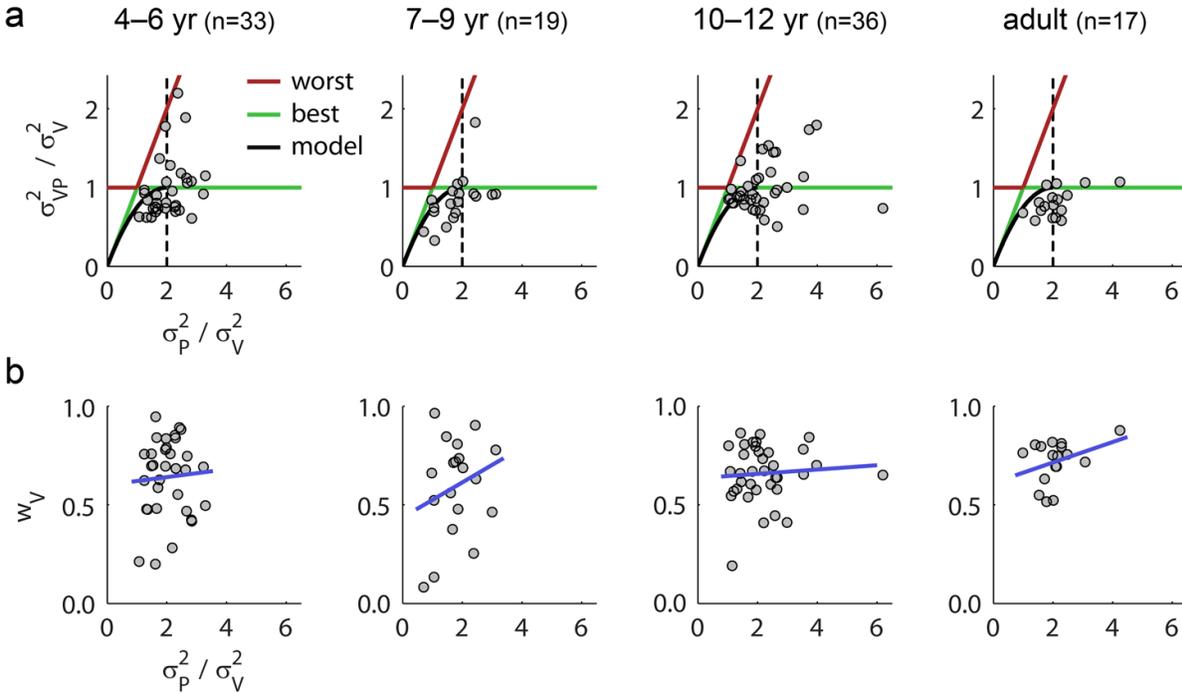
Figure 4. a. *Relationship in each age group between ratios of single-cue variances (P/V) and combined-to-single-cue variances (VP/V). Performance is predicted based on use of the single worst cue, the single best cue, or integration of cues according to the Bayesian model. High ratios along the x axis correspond to much more reliable vision than proprioception. Ratios below 1 along the y axis corresponds to an improvement given VP compared with V alone.* **b.** *Relationship in each age group between ratios of single-cue variances (P / V) and weighting for vision in the combined-cue condition VP (regression lines).*

In this analysis our aim is to understand the whole data set, not just the part fit by the model, which is clearly incomplete. However, the relationship between model predictions and data in Figure 4a provides clues as to how the model might be improved. In the 7-9 year and adult groups, who do achieve overall variance reduction given VP *vs.* V alone, points fall largely to the right of the model-predicted curves. It is possible that P/V ratios are being consistently overestimated. A systematic reduction in participants' estimated P/V ratios would shift points to the left, and so both bring larger numbers of points into the model-able P/V < 2 range, and provide a better fit of data to the model-predicted curves. One reason why P/V ratios might be overestimated is if P variance is being overestimated (van Beers *et al.*, 1996; see Discussion,

below). Note that for the 4-6 and 10-12 year groups, a downward revision of P/V estimates would not be sufficient to correct for deviations from the model.

Even without integrating cues, participants who are taking their differing reliabilities into account could minimise the variance of their estimate of hand position simply by using the single more reliable cue (Figure 4a, green line). The 7-9 year and adult groups are indeed generally on or below this line, while many in the 4-6 year and 10-12 year groups are well above the line, *i.e.* are much more variable when given VP information together than when given V information alone. In these age groups, the incidence of high VP/V ratios is greatest to the right of the V/P = 2 line. The lower VP/V ratios of individuals left of the line reflect the result that when we consider only these participants (Figure 3), we see evidence for a benefit for VP over single cues in all four age groups.

It seems that adult and 7-9 year old individuals whose proprioceptive estimates are highly variable compared with vision are able to down-weight or discount P and so remain close to the line predicted by relying only on V. By contrast, 4-6 and 10-12 year old individuals with high P/V ratios often perform worse than if they had relied only on V, suggesting that in these individuals the P information is influencing the VP estimate substantially, even though it is unreliable. Some individuals are close to the prediction for relying fully on P (red line), when it is the less reliable cue.

Combined-cue benefits for subsets of individuals in these groups who have relatively similar V and P variances could be explained if individuals at these ages did not choose weights appropriate to their own individual single-cue reliabilities, but adopted a mean weight that is appropriate for the situation of the two cues being relatively similar. This weight leads to successful variance reduction relative to the best single cue in participants for whom the single

cues are indeed similar in reliability, but to over-weighting of proprioception (and so a variance increase relative to the best single cue) in participants for whom proprioception is much less reliable than vision.

Our measure of cue weighting in the VP condition provides a way to investigate this possibility. Figure 4b plots the relationship between the measure of single-cue discrepancy, P/V, and the measure of weighting for vision in the VP condition. Groups in which individuals weight cues based on these single cues' reliabilities should show a positive relationship, with individuals with high P/V ratios (whose vision is much more reliable than their proprioception) giving a higher weighting for vision. As the experiment included no cue conflict conditions, the weighting measure relies on analysis of small naturally occurring differences in constant errors across conditions, and so is not as sensitive a measure of weighting as a conflict manipulation could provide.

Regression lines describing relationships between these variables are plotted in Figure 4b and analysed in Table 3. As Figure 4b and Table 3 show, there were stronger positive relationships between cue reliability and cue weighting at 7-9 years and in adults, where combined-cue benefits were evident in each group as a whole, than at 4-6 and 10-12 years, where they were not. Although the trends are in the expected direction (Table 3), no single group's analysis reached statistical significance. To increase statistical power, a post-hoc analysis combined the two groups successful at cue integration in the main (Figure 2) analysis, 7-9 year and adult. In this larger combined group there was a significant linear relationship in the predicted direction ($R^2 = 0.09$, $F_{(1, 34)} = 3.37$, $p < 0.05$, one-tailed). By contrast, in an analysis combining the two groups that were unsuccessful, 4-6 and 10-12 years, there was no significant linear relationship ($R^2 = 0.00$, $F_{(1, 67)} = 0.33$, $p = 0.284$, one-tailed).

ratio of P / V variances vs weighting for V in condition VP			
	R^2	F	d.f.	p (one-tailed)
4 – 6 yr.	0.00	0.11	1, 31	0.372
7 – 9 yr.	0.06	1.04	1, 17	0.161
10 – 12 yr.	0.01	0.21	1, 34	0.324
Adult	0.12	2.04	1, 15	0.087

Table 3. Results of linear regression analyses of P / V variance ratios vs. weighting for vision in condition VP.

These results are consistent with flexible cue weighting by individual 7-9 year olds and adults, calibrated to their individual visual and proprioceptive reliabilities, but adoption of a 'default' weighting at 4-6 and 10-12 years that does take individuals' reliabilities into account. This weighting is on average appropriate for participants with relatively similar visual and proprioceptive reliabilities, but overweights proprioception for those with relatively unreliable proprioception. Despite this inflexible weighting at 4-6 and 10-12 years, more than half of participants at all ages had a lower variance given VP than either single cue (*i.e.* are below the green line in Figure 3). Percentages of each age group with VP lower than either single cue were 65% at 4-6 years, 84% at 7-9, 63% at 10-12, and 76% in adults.

Discussion

The ability to combine multiple sensory estimates to reduce uncertainty is a pervasive feature of human perception and motor control. However, studies to date have found no evidence for this ability in childhood until after 8 or 12 years (Gori *et al.*, 2008; Nardini *et al.*, 2008; Nardini *et al.*, 2010). Here we used a simple and naturalistic task, avoiding either memory demands or two-alternative-forced-choice psychophysics, to study development of the ability to combine vision and proprioception to localise the hand.

The youngest entire age group able to demonstrate reduced pointing variability given vision and proprioception together compared with either cue alone was 7-9 years (Figure 2). This is the first evidence, to our knowledge, for a statistically significant improvement in hand localization ability in childhood when proprioception is added to vision (Smothergill, 1973; von Hofsten & Rosblad, 1988; van Beers *et al.*, 1996; King *et al.*, 2010). The improvement was modest, and the present study may have been able to measure it because we used a relatively large number of trials per condition, and estimated distributions of points using a robust procedure (Rousseeuw, 1984).

In other kinds of tasks investigating multisensory uncertainty reduction, children at a similar age did not yet integrate cues to navigate (Nardini *et al.*, 2008), or visual information to judge depth (Nardini *et al.*, 2010), and were just acquiring the ability to integrate vision and touch to judge sizes of blocks (Gori *et al.*, 2008). The 7-9 year age is thus on the early side of those shown to combine sensory cues to reduce uncertainty, but consistent with results from another manual task (Gori *et al.*, 2008).

We found two further unexpected results: first, a "u-shaped" pattern in which the older 10-12 year group failed to show an overall ability for uncertainty reduction before adults showed it

again (Figure 2), and second, evidence for uncertainty reduction in all groups, including the youngest 4-6 year old group, when considering only participants whose visual and proprioceptive variabilities were relatively similar (Figure 3).

Comparing patterns of two-cue performance with predictions based on use of either single or combined cues (Figure 4), we found that at 7-9 years and in adults (groups that on average showed improvement relative to the best single cue), individuals tended to range between showing improvement relative to the best single cue, and showing performance similar to the best single cue. By contrast, at 4-6 and 10-12 years (groups that on average showed no improvement relative to the best single cue), some individuals did show improvement, some were similar to the best single cue, and some were substantially worse and in line with use of the worst single cue. Crucially, within each group, this individual variation appeared not to be random, but to depend on the degree of discrepancy between the variabilities of the two single cues. At 4-6 and 10-12 years, observers whose proprioception was no more than two times more variable than their vision benefited on average from combining cues (Figure 3, Figure 4a), whereas others did not (Figure 4a).

We proposed that all participants tended to integrate (average) cues, and that the lack of multisensory benefit at 4-6 and 10-12 years reflects a poor ability of individuals at these ages to adopt the cue weightings appropriate for their own cue reliabilities. The weighting that the group on average adopted was suitable for observers with relatively similar single-cue reliabilities, but resulted on average in an over-weighting of proprioception, and so an increase in variability given both cues together compared with vision alone, in those whose proprioception was much more variable than their vision.

Regression analyses investigated this interpretation by examining the relationship between the relative reliability of vision and its weighting in the *VP* condition. Results were consistent with this suggestion. Linear relationships between cue reliability and cue weighting in the predicted direction were strongest at 7-9 years and in adults (although not statistically significant), and practically non-existent ($R^2 < 0.01$) at 4-6 and 10-12 years. A post-hoc analysis combining "successfully integrating" 7-9 year and adult groups and "not successfully integrating" 4-6 and 10-12 year groups found a significant linear relationship in the former combined group but not the latter. As we did not include cue conflict conditions, our measure of weighting relies only on the small naturally occurring differences in constant errors across conditions (see Figure 2a), and is therefore relatively noisy and insensitive. However, the pattern of results is consistent with the suggestion that the differences underlying successful *vs.* unsuccessful overall cue integration in the age range studied are in the ability of individuals to take their own cue reliabilities into account when combining the cues.

The interesting conclusion from this is that children are already integrating cues at 4-6 years, a younger age than previously reported. Due to immaturity in setting the correct weights the group as a whole does not benefit from combining cues; however those individuals within the group who happen to be using weights appropriate to their own abilities do benefit.

A major result still in need of an explanation is the apparent loss of the ability to benefit from multiple cues at 10-12 years. It is notable that the 10-12 year group continue the downward trend with age in overall variability (Figure 2), therefore the apparent "u-shaped" development is not in overall ability to localise the hand, but only in the more specific ability to improve precision when given combined as compared with single sensory cues.

One possibility that would explain poor integration ability at this age, particularly if its cause is inappropriate weighting of cues, would be that the body is growing, and/or sensory systems are developing, more rapidly than at younger ages. Some of those in our 10-12 year group, which has mean age 11.6 years, would be at the start of the adolescent growth spurt; for example, on average, girls reach their peak height velocity at 12 years (Rogol, Clark, & Roemmich, 2000). It is possible that at this age observers are not able to keep up with the changes in their own visual and proprioceptive functioning and so to properly calibrate their sensory systems and their relative reliabilities. An experimental test of this suggestion will require a longitudinal study relating individuals' development in cue integration ability to physical changes and changes in unisensory precision. Interestingly, recent studies of visual-auditory integration also show a temporary loss of integration ability at around 10 years (Barutchu, Crewther, & Crewther, 2009; Barutchu *et al.*, 2010). It is possible that besides bodily changes, other more central changes in sensory processing are taking place at this age.

As well as asking simply whether target localization with two cues was less variable than with one cue, we compared performance with a Bayesian (ideal observer) model (van Beers *et al.*, 1996). The model, which is to our knowledge the only quantitative model of visual-proprioceptive hand localization, did not provide a good fit to the data. First, around half of our participants could not be modelled as in these the model would predict an (impossible) negative value for the estimate of vision-only variability associated with cues above the table. Second, in those participants who could be modelled, adults and 7-9 year olds performed significantly better than the theoretically optimal model. Both these results can be explained if the experimental task systematically overestimated P variance. If P variance estimates were revised downwards, individual observers' fit to the model prediction curve at 7-9 years and in adults would also be improved (Figure 4a)

In order to present a "proprioception only" condition, we eliminated all visual information

in condition P, in a manner similar to previous studies with children (Mon-Williams *et al.*, 1999;

Smothergill, 1973; van Beers *et al.*, 1996; von Hofsten & Rosblad, 1988) and adults (van Beers,

Sittig, & Denier van der Gon JJ, 1998; van Beers *et al.*, 1999). A possibility raised by van Beers *et al.*,

1996 is that in condition P, closing the eyes has not only the intended effect of removing vision of

the target, but also has unintended consequences for proprioceptive precision. While not directly

informative for the task of localising the target, the visual frame of reference provided by the

room and table-top may help to maintain the proprioceptive sense for hand position, even when

the hand itself cannot be seen. Closing the eyes can also disrupt balance and lead to the need to

make more corrective postural adjustments (Wade & Jones, 1997), making movements less stable.

Future experiments and models should aim either to quantify the visual contribution to

proprioception, or to use more careful manipulations in which only vision of the hand and target,

but not the surrounding space, vary across conditions (e.g. King *et al.*, 2010).

All participants, whether right- or left-handed, responded using the left hand. As is

common at young ages (McManus *et al.*, 1988), the youngest group contained a relatively high

proportion of left-handed children. This could have given this group some advantage relative to

other groups; that is, 4- to 6-year-old performance may be better than performance at that age

would be if all stimuli had been presented with respect to individuals' handedness.

Although we positioned participants at a comparable height (so that the table was midway

between shoulder and elbow), because of differences in body size the geometry of the task was

not perfectly comparable across ages. In particular, in participants with shorter arms, a greater

proportion of the responding (under-table) arm was occluded. This factor could have made

conditions including vision (*V, VP*) more difficult for younger participants.

We did not control or measure initiation times, which may have added some variability to our results. Visual and/or proprioceptive information was continuously available before, during and up to the response, with the aim of excluding memory demands from the design. Therefore, the measures should not have disadvantaged either younger participants, who may have been slower to respond, or conditions with more complex instructions (e.g. in condition *P* - first close eyes, then have finger placed on the now non-visible target – then respond). However there are two possible uncontrolled contributions from memory that could have differentially affected performance across ages and conditions. First, although in condition *P* the target was specified (proprioceptively) only after eyes were closed, a visual memory trace of all three targets could potentially have played some role in localisation, and such a memory trace would decay with time. Second, although in conditions *VP* and *P* static proprioceptive information about the non-responding finger-tip's location was continuously available once the finger was placed, additional movement-related proprioceptive information acquired during the preceding movement could potentially have played a role in localisation, and this information too would decay over time. Future studies should consider ways to better control or measure timing, within the constraints of what is feasible with young children.

A difference between conditions *P* and *VP* was that in *P* the experimenter placed the non-responding finger on the target (passive movement) whereas in *VP* the participant placed it (active movement) – this could have changed the type of proprioceptive information available. A difference between conditions *V* and *VP* was that in *VP* participants could see both the target and their finger, i.e. they potentially had more visual information to use. Both these factors suggest possible uncontrolled reasons for variance reduction in condition *VP* relative to others, and may help to explain participants' abilities to outperform the model, but leaves unresolved why this

should differ across ages. Future studies should aim to eliminate these differences across conditions.

We mixed trial types throughout the study, which may have unduly increased the variance across all conditions. It is possible that endpoints in the *P* condition were influenced by the presence of visual information in a preceding *V* or *VP* trial (Cheng, Luis, & Tremblay, 2008; de Lussanet, Smeets, & Brenner, 2002). In the studies by van Beers and colleagues (van Beers *et al.*, 1996; van Beers *et al.*, 1999), trials were presented in blocks of a single type. In future studies it is likely that more reliable measures would be obtained by using this approach.

A key challenge for understanding development of sensory-motor optimizations includes understanding the processes by which sensory and motor systems acquire and use information about their own reliabilities. The present results indicate that in hand localisation, development of abilities to combine sensory estimates preceded development of appropriate weighting for these. In addition, this weighting, once first acquired, was temporarily lost in later childhood. The specificity of these results to hand localisation, the learning or other mechanisms involved in acquiring optimal cue weights, and the abilities of these mechanisms to deal with physical growth, remain questions for further study. The changes in neural information processing underlying development of uncertainty reduction by cue integration are also unknown. Recent studies with animal models have begun to reveal the neuronal processes underlying optimal integration of sensory cues (Gu, Angelaki, & Deangelis, 2008; Morgan, Deangelis, & Angelaki, 2008); in the future, insights from these, together with neuroimaging techniques and mathematical models (Doya, Ishii, Pouget, & Rao, 2007; Weisswange, Rothkopf, Rodemann, & Triesch, 2011) may help to elucidate the developmental mechanisms by which humans become able to combine sensory cues to reduce uncertainty.

Acknowledgments

References

Alais, D. & Burr, D. (2004). The Ventriloquist Effect Results From Near Optimal Crossmodal Integration. *Current Biology*, 14, 257-262.

Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The Race That Precedes Coactivation: Development of Multisensory Facilitation in Children. *Dev. Sci.*, 12, 464-473.

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual Integration in Noise by Children and Adults. *J. Exp. Child Psychol.*, 105, 38-50.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational Efficiency and the Growth of Short-Term Memory Span. *Journal of Experimental Child Psychology*, 33, 386-404.

Cheng, D. T., Luis, M., & Tremblay, L. (2008). Randomizing Visual Feedback in Manual Aiming: Reminiscence of the Previous Trial Condition and Prior Knowledge of Feedback Availability. *Exp. Brain Res.*, 189, 403-410.

Clark, J. J. and Yuille, A. L. (1990). *Data Fusion for Sensory Information Systems*. Boston, MA: Kluwer Academic

de Lussanet, M. H., Smeets, J. B., & Brenner, E. (2002). The Relation Between Task History and Movement Strategy. *Behav. Brain Res.*, 129, 51-59.

Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N. (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. Cambridge: MIT Press

Ernst, M. O. (2005) . A Bayesian view on multimodal cue integration. In *Human Body Perception from the Inside Out*, ed. Knoblich, G., Thornton, I. M., Grosjean, M., & Shiffrar, M. New York: Oxford University Press

Ernst, M. O. (2008). Multisensory Integration: a Late Bloomer. *Curr. Biol.*, 18, R519-R521.

Ernst, M. O. & Banks, M. S. (2002). Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion. *Nature*, 415, 429-433.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Dev. Psychol.*, 40, 177-190.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, 18, 694-698.

Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor Haptic Orientation Discrimination in Nonsighted Children May Reflect Disruption of Cross-Sensory Calibration. *Curr. Biol.*, 20, 223-225.

Gu, Y., Angelaki, D. E., & Deangelis, G. C. (2008). Neural Correlates of Multisensory Cue Integration in Macaque MSTd. *Nat. Neurosci.*, 11, 1201-1210.

King, B. R., Pangelinan, M. M., Kagerer, F. A., & Clark, J. E. (2010). Improvements in Proprioceptive Functioning Influence Multisensory-Motor Integration in 7- to 13-Year-Old Children. *Neurosci. Lett.*, 483, 36-40.

Körding, K. P. & Wolpert, D. M. (2006). Bayesian Decision Theory in Sensorimotor Control. *Trends Cogn Sci.*, 10, 319-326.

McManus, I. C., Sik, G., Cole, D. R., Mellon, A. F., Wong, J., & Kloss, J. (1988). The Development of Handedness in Children. *British Journal of Developmental Psychology*, 6, 257-273.

Mon-Williams, M. A., Wann, J. P., & Pascal, E. (1999). Visual-Proprioceptive Mapping in Children With Developmental Coordination Disorder. *Dev. Med. Child Neurol.*, 41, 247-254.

Morgan, M. L., Deangelis, G. C., & Angelaki, D. E. (2008). Multisensory Integration in Macaque Visual Cortex Depends on Cue Reliability. *Neuron*, 59, 662-673.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of Visual Cues Is Not Mandatory in Children. *Proc. Natl. Acad. Sci. U. S. A*, 107, 17041-17046.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, 18, 689-693.

Rogol, A. D., Clark, P. A., & Roemmich, J. N. (2000). Growth and Pubertal Development in Children and Adolescents: Effects of Diet and Physical Activity. *Am. J. Clin. Nutr.*, 72, 521S-528S.

Rousseeuw, P. J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P. J. & van Driessen, K. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41, 212-223.

Smothergill, D. W. (1973). Accuracy and Variability in the Localization of Spatial Targets at Three Age Levels. *Developmental Psychology*, 8, 62-66.

Trommershäuser, J., Körding, K., and Landy, M. S. (2011). *Sensory Cue Integration*. New York: Oxford University Press

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley

van Beers, R. J., Sittig, A. C., & Denier van der Gon JJ (1998). The Precision of Proprioceptive Position Sense. *Exp. Brain Res.*, 122, 367-377.

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1996). How Humans Combine Simultaneous Proprioceptive and Visual Position Information. *Exp. Brain Res.*, 111, 253-261.

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1999). Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model. *J. Neurophysiol.*, 81, 1355-1364.

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When Feeling Is More Important Than Seeing in Sensorimotor Adaptation. *Curr. Biol.*, 12, 834-837.

Verboven, S. & Hubert, M. (2005). LIBRA: a MATLAB Library for Robust Analysis. *Chemometrics and Intelligent Laboratory Systems*, 75, 127-136.

von Hofsten, C. & Rosblad, B. (1988). The Integration of Sensory Information in the Development of Precise Manual Pointing. *Neuropsychologia*, 26, 805-821.

Wade, M. G. & Jones, G. (1997). The Role of Vision and Spatial Orientation in the Maintenance of Posture. *Phys. Ther.*, 77, 619-628.

Weisswange, T. H., Rothkopf, C. A., Rodemann, T., & Triesch, J. (2011). Bayesian Cue Integration As a Developmental Outcome of Reward Mediated Learning. *PLoS. One.*, 6, e21575.