The development of Bayesian integration in sensorimotor estimation

Claire Chambers [1,2,3]

Taegh Sokhey [1,4]

Deborah Gaebler-Spira [1,3]

Konrad Kording [1,2,3]

1. Sensory Motor Performance Program, Shirley Ryan Abilitylab, Chicago, IL, USA
2. Biomedical Engineering, Robert R. McCormick School of Engineering and Applied Sciences, Northwestern University, Chicago, IL, USA
3. Department of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
4. Department Of Biological Sciences, Northwestern University, Evanston, IL, USA

If the brain is inherently Bayesian, then behavior should show the signatures of Bayesian computation from an early stage in life without the need for learning. Children should integrate probabilistic information from prior and likelihood distributions to reach decisions and should be as statistically efficient as adults. To test this idea, we examined the integration of prior and likelihood information in a simple position estimation task comparing children aged 6-11 years and adults. During development, estimation performance became closer to the statistical optimum. Children use likelihood information as well as adults but are limited in their use of priors. This finding suggests that Bayesian behavior is not inherent but learnt over the course of development.

The behavior of adults under uncertainty is well described by Bayesian inference, in that adult humans weigh different sources of information according to their relative uncertainty. Behavior is consistent with Bayesian computations in sensorimotor behavior (Berniker, Voss, & Kording, 2010; Kording & Wolpert, 2004), perception (Knill & Richards, 1996; Mamassian & Goutcher, 2001), cognition and reasoning tasks (Battaglia, Hamrick, & Tenenbaum, 2013; Tenenbaum & Griffiths, 2001), and cue combination across and within sensory modalities (Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004). Adult humans seem to integrate information in a way predicted by Bayesian statistics.

These numerous findings of Bayesian behavior have led to the theory that the underlying neural computations are inherently Bayesian. For example, it has been argued that the activity of neural populations reflects probabilistic population codes that directly implement Bayesian computations (Beck et al., 2008; Ma, Beck, Latham, & Pouget, 2006; Ma, Beck, & Pouget, 2008; Pitkow & Angelaki, 2017). However, findings of Bayesian behavior are not sufficient to support this claim. Bayesian behavior simply represents optimal behavior under uncertainty and there are ways of generating optimal behavior that do not explicitly implement Bayesian computation (Mandt, Hoffman, & Blei, 2017; Rao, 2004). Therefore, previous research has not fully established whether the neural code is inherently Bayesian.

If neural circuits are evolved to implement Bayesian computations, then behavior should always show Bayesian signatures, including during development. Therefore, children too should act in accordance with the rules of Bayesian integration. Specifically, they should weigh information according to its relative uncertainty in simple tasks.

Indeed, a good number papers ask how Bayesian children are. Some work on looking times in infants is consistent with early optimal integration of information (Téglás, Tenenbaum, & Bonatti, 2011) and it has been shown that young children are able to use probabilistic information to infer causality in order to perform actions (Gopnik & Wellman, 2013; Kushnir & Gopnik, 2007; Sobel, Tenenbaum, & Gopnik, 2004). However, in the former case measurements are indirect and in the latter case predictions can only be qualitative. Work on the development of cross-modal cue combination the shows that older children do not integrate information, but instead process information from each modality separately up to the age of approximately 9-11 years (Gori, Del Viva, Sandini, & Burr, 2008; Nardini, Bedford, & Mareschal, 2010; Nardini, Jones, Bedford, & Braddick, 2008). Therefore, based on previous research, it is unclear whether the behavior of children is consistent with use of Bayesian inference.

Here we investigate whether integration of current with past information to perform actions under uncertainty is present in young children or is acquired over the course of development. In our paradigm, we examine the use of probabilistic information to perform a simple sensorimotor estimation task, previously used in adults to examine integration under uncertainty (Berniker et al., 2010; Kording & Wolpert, 2004; Vilares, Howard, Fernandes, Gottfried, & Kording, 2012). Visual targets were drawn from a prior distribution and participants were shown uncertain sensory information about each target. We found that all age groups learned to use the uncertainty of sensory information. However, children did not exploit the uncertainty of the prior to perform estimations, as adults did, with this gradually emerging during development.

Methods

Experimental details

We aimed to examine probabilistic inference during sensorimotor estimation in a child population. Our task was designed to examine use of probabilistic information during sensorimotor estimation (Fig. 1a, Acuna, Berniker, Fernandes, & Kording, 2015; Berniker et al., 2010; Vilares et al., 2012). Previous findings indicate that adults weigh information according to its reliability and learn priors in a manner which resembles Bayesian integration during sensorimotor estimation. For the purposes of the current study, we adapted the experimental protocol for child participants, by using a concept that was engaging to children, using simplified instructions, and by reducing the number of trials.

Participants were 16 children (8 males) aged 6-8 years (M=6.94, SD=0.77), 17 children (8 males) aged 9-11 years (M=10.06, SD=0.75), and 11 adults (5 males) aged over 18 years (M=27.27, SD=5.31). The data of four participants aged 5 years were excluded due to difficulty in using a computer mouse. The data of two additional participants were excluded due to looking away from the screen during the experiment.

In a quiet room, participants sat in front of a 52 cm wide, 32.5 cm high computer monitor. Before starting the experiment, we presented participants with the instructions that someone behind them was throwing candy into a pond, represented by the screen; and that their aim was to estimate where the candy target landed and catch as many candy as possible over the course of the experiment. Candy targets were drawn from a Gaussian distribution centered at the middle of the screen, $N(\mu, \sigma_s^2)$. On each trial, they were presented with an uncertain "splash" stimulus for one second and were told that the splash was caused by a hidden candy target. The splash was *n*=4 samples from a Gaussian likelihood distribution that was centered on target location $N(s, \sigma_l^2)$. Participants provided an estimate of the candy target's location on the horizontal axis using a vertical bar that extended from the top to the bottom of the screen. The net appeared at the same time as the splash at a random location on screen. Participants had 6 seconds to respond. After providing a response, they were shown the true candy location.

One simple strategy for performing sensorimotor estimation under uncertainty is to consistently judge target location at the center of the splash – i.e. full reliance on the likelihood. This strategy works well when the likelihood distribution is narrow, because the closely-spaced points of the splash are an accurate indicator of target location (Fig. 1b, left). However, full reliance on the likelihood would cause a participant to miss targets more frequently as the likelihood distribution widens (Fig. 1b, right). When sensory information is unreliable, rather than relying on the likelihood completely, we maximize performance by giving more weight to our prior belief on target location. More generally, the best or optimal strategy involves weighing sources of information according to their relative precision.

Formally, weighing sources of information according to their relative precision corresponds to Bayesian inference. An optimal Bayesian observer combines noisy sensory information from the likelihood, $N(s, \sigma_l^2/n)$ with their learned prior, $N(\mu, \sigma_s^2)$, resulting in a posterior distribution over target location, $N\left(\left(\frac{\mu}{\sigma_s^2} + \frac{c}{\sigma_l^2/n}\right)\Big/\left(\frac{1}{\sigma_s^2} + \frac{1}{\sigma_l^2/n}\right), 1\Big/\left(\frac{1}{\sigma_s^2} + \frac{1}{\sigma_l^2/n}\right)\right)$. The mean of the posterior is a mean of the prior and centroid of the likelihood, *c*, weighted by their precisions. From this posterior distribution, an estimation, $\hat{s}$, is computed. Therefore, the optimal reliance on the likelihood is a function of prior and likelihood uncertainties, $\sigma_s^2/(\sigma_s^2 + \sigma_l^2/n)$. We can manipulate the prior and likelihood variance and measure their influence on participants' reliance on the likelihood, in order to investigate probabilistic information during sensorimotor estimation.

To investigate how children use probabilistic information during sensorimotor estimation, we manipulated the variances of prior distribution and likelihood distributions. We used a Gaussian prior distribution with a mean at the center of the screen and standard deviation of .03 (Narrow Prior) or .1 (Wide Prior) in units of screen width. The likelihood distribution was centered on target location and could have a standard deviation of .05 (Narrow Likelihood), .1 (Medium Likelihood), or .25 (Wide Likelihood) in units of screen width. There were six conditions: Narrow Prior – Narrow Likelihood, Narrow Prior – Medium Likelihood, Narrow Prior – Wide Likelihood, Wide Prior – Narrow Likelihood, Wide Prior – Medium Likelihood, and Wide Prior – Wide Likelihood.

The experiment consisted of four blocks, each lasting 120 trials, preceded by a practice block lasting 10 trials. Trials were blocked by prior condition, with all likelihood conditions being presented in randomized order within one block. The prior over target location switched from block to block with a randomly chosen starting condition for each participant (i.e., narrow-wide-narrow-wide or wide-narrow-wide-narrow).

We introduced a number of modifications to engage child participants in the task. Participants were shown how much candy they had won on screen and participants won a "bonus" piece of candy for every ten candy they caught. Sounds were presented to signal successfully catching a target and missed responses when they did not respond within the 6-second time window. Step-by-step instructions were shown to participants on screen before the experiment, to ensure that all participants received the same instructions. Participants were told that their payment was based on the number of candy that they caught.

Ethical approval was provided by the NU IRB #20142500001072 (Northwestern University, USA). Participants signed a consent form before participation. For participants aged under 18 years, a parent provided consent for their child to take part and completed the Developmental Coordination Disorder questionnaire (Wilson et al., 2009), a modified Vanderbilt questionnaire to assess for ADHD (Wolraich et al., 2003), and the Behavior Assessment System for Children, BASC-3, parent rating scales (Reynolds, 2004). After the participant had completed the game, we administered the child mini-mental state evaluation to obtain an approximate assessment of cognitive ability (Ouvrier, Goldsmith, Ouvrier, & Williams, 1993). No data was excluded on the basis of neuropsychological test results.

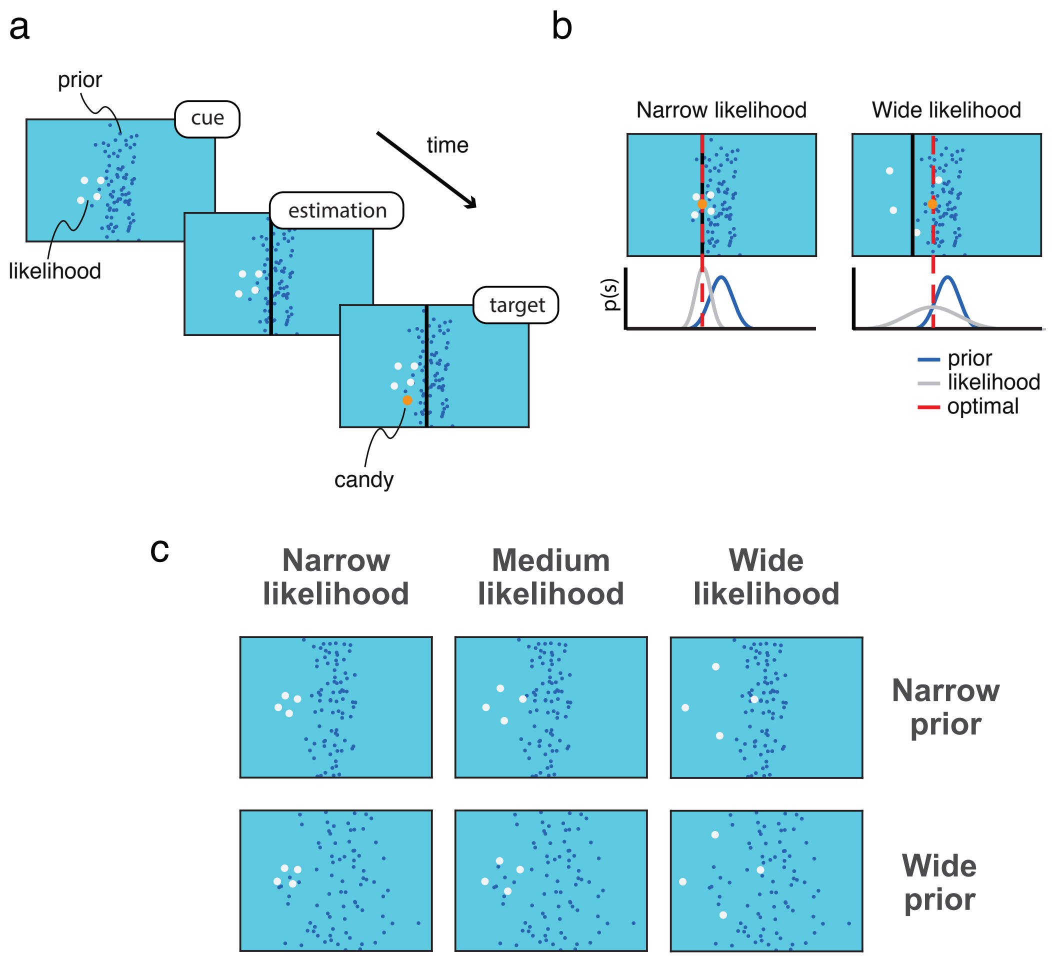

Figure 1. (**a**) Experimental protocol. Participants were shown a visual cue (likelihood) with experimentally controlled uncertainty (splash), created by a hidden target (candy) drawn from a prior distribution. Participants were told that the splash was created by candy falling into a pond. Participants were prompted to place a vertical bar (net) where the hidden target fell, and were then shown feedback on target location. (**b**) Relying on the likelihood. A simple strategy would be to rely entirely on likelihood information by pointing at its centroid on each trial. While this strategy is close to optimal when the likelihood is precise or narrow (black bar, left panel), this strategy is less successful when the likelihood is wider (black bar, right panel), as samples from the likelihood become a less reliable indicator of target location and the optimal estimate shifts closer to the prior mean. The optimal strategy involves weighing prior and likelihood information according to their relative uncertainties (**c**) Experimental design. In order to quantify integration of the prior and likelihood, we measured reliance on the likelihood (*Estimation slope*) under different conditions of prior variance and likelihood variance. The prior could be narrow or wide, and the likelihood could be narrow, medium, or wide.

Data analysis

We were interested in the integration of probabilistic information from a prior distribution and sensory information from the likelihood in sensorimotor estimations. To investigate this, we examined whether samples from the likelihood distribution, $X = \{x_1, x_2, x_3, x_4\}$, were combined with information about the prior distribution, $N(\mu, \sigma_s^2)$, when producing an estimate of target location, $\hat{s}$. We quantified this for each condition using the extent to which participants relied on the likelihood, given by the linear relationship between the centroid of the splash, $c = \sum_i^N x_i/n$, and their estimate on each trial, $\hat{s}$. We performed a linear regression with estimations, $\hat{s}$, as dependent variable and the likelihood centroid, c, as the independent variable, which resulted in a measure of reliance on the likelihood, which we term the *Estimation slope* (*ES*). We assumed that participants accurately learned the mean of the prior, and set the intercept to 0. If participants relied only on the likelihood to generate their estimate, then they should point close to the centroid of the splash, c, on all trials, leading to an *Estimation slope* ≈ 1. If instead participants ignore the likelihood and instead only use their learnt prior, then their estimates should not depend on the c, leading to an *Estimation slope* ≈ 0. Therefore, from participants' estimations we obtain a measure of their reliance on the likelihood or prior.

The theoretical variance of the likelihood and prior used in the experiment provide optimal values for the *Estimation slope*, i.e. how much participants should rely on the likelihood. For an optimal Bayesian observer, sources of information are weighed according to their relative reliabilities, $ES_{opt} = \sigma_s^2/(\sigma_s^2 + \frac{\sigma_l^2}{n})$. Therefore, in order to quantify how optimal participants were, we can compare *Estimation slope* quantified from participant's data with the optimal *Estimation slope*, ES_{opt}, by computing the absolute difference, $|ES - ES_{opt}|$ in each condition, and then averaging this across conditions, which provided a *Distance to the optimal* score for each participant.

We wanted to know how sensitive participants were to the prior and likelihood variance. We therefore devised separate measures to quantify how much participants distinguished between prior conditions and between likelihood conditions. The sensitivity to the prior was simply the difference between the *Estimation slopes* across prior conditions, which was then summed across likelihood conditions:

$$Sensitivity\ to\ prior = (ES_{\mathrm{WP\ NL}} - ES_{\mathrm{NP\ NL}}) + (ES_{\mathrm{WP\ ML}} - ES_{\mathrm{NP\ ML}}) + (ES_{\mathrm{WP\ WL}} - ES_{\mathrm{NP\ WL}})$$

Similarly, the sensitivity to the likelihood was the difference between *Estimation slopes* across likelihood conditions for a fixed prior condition, summed across prior conditions:

$$Sensitivity\ to\ likelihood = (ES_{\mathrm{WP\ NL}} - ES_{\mathrm{WP\ ML}}) + (ES_{\mathrm{WP\ ML}} - ES_{\mathrm{WP\ WL}}) + (ES_{\mathrm{NP\ NL}} - ES_{\mathrm{NP\ ML}}) + (ES_{\mathrm{NP\ ML}} - ES_{\mathrm{NP\ WL}})$$

To examine subject-specific biases, in the form of an overall tendency to use the likelihood only or prior only regardless of the experimental condition, we computed a *Bias* score for each participant, which was simply the *Estimation slope* averaged across all conditions. This allowed us to examine use of simple response strategies.

To quantify the uncertainty of our estimation of *Estimation slope*, *Distance to optimal*, *Sensitivity to prior*, *Sensitivity to likelihood*, and *Bias*, we performed bootstrapped estimation by resampling the data with replacement 1000 times and performing the fit for each resampled data set. Since the data did not meet the requirements for parametric statistical tests, we performed non-parametric tests on the data: Kruskal-Wallis tests to examine main effects, Mann-Whitney U tests to examine differences between groups, and Wilcoxon signed-rank tests for comparison of individual samples with chance level, Bonferroni-corrected for the number of comparisons.

Results

We wanted to investigate the development of Bayesian integration. To do so, we examined whether children aged 6-11 years old and adults could learn to use uncertainty of different sources of information (prior and likelihood) during sensorimotor estimation (Fig. 1). We examined use of probabilistic information during development by quantifying participants' task performance, reliance on the likelihood, the degree to which they deviated from the statistical optimum and their sensitivity to the prior variance and likelihood variance condition.

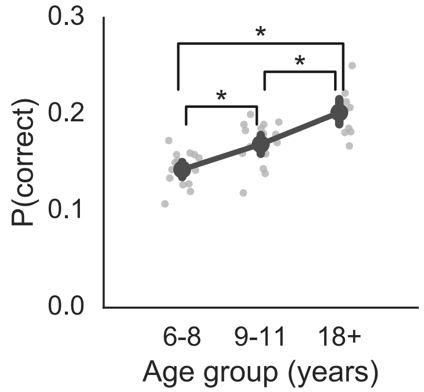

Figure 2. Performance of candy-catching task. The proportion of correct responses (*p(correct)*), where the net overlapped with the target, is shown as a function of age group. *p(correct)* for individual participants is shown in gray overlaid with the mean for each age group (error bars = 95% CI) in blue. *p(correct)* increases with age with significant differences between age groups (Table 1).

It was first important to establish that the all age groups understood and carried out the task, with performance above chance level of 2%. We therefore compared the proportion of candy targets caught, *p(correct)*, to chance level for each age group (Fig. 2). The performance of all age groups exceeded chance level, as tested by a non-parametric Wilcoxon signed rank tests with Bonferroni-corrected p-values (6-8 y: median score = .14, W = 0, p = .0013; 9-11 y: median score = .17, W = 0, p = .0008; 18+ y: median score = .20, W = 0, p = .0100). This shows that all age groups understood and carried out the candy-catching task. Therefore, differences between age groups cannot be attributed to a lack of understanding of the task.

Next, we asked if performance improves during development. Performance increases significantly with age (Kruskal-Wallis test, H(2) = 25.97, p < .0001), with significant differences between age groups (Bonferroni-corrected Mann-Whitney U tests, Table 1). Possible contributors to the improvement in performance include a systematic difference between participants' behavior and optimal behavior that decreases during development, and motor or decisional noise. Here, we were primarily interested in the integration of prior and likelihood information into participants' decisions, and therefore focused on the former possibility.

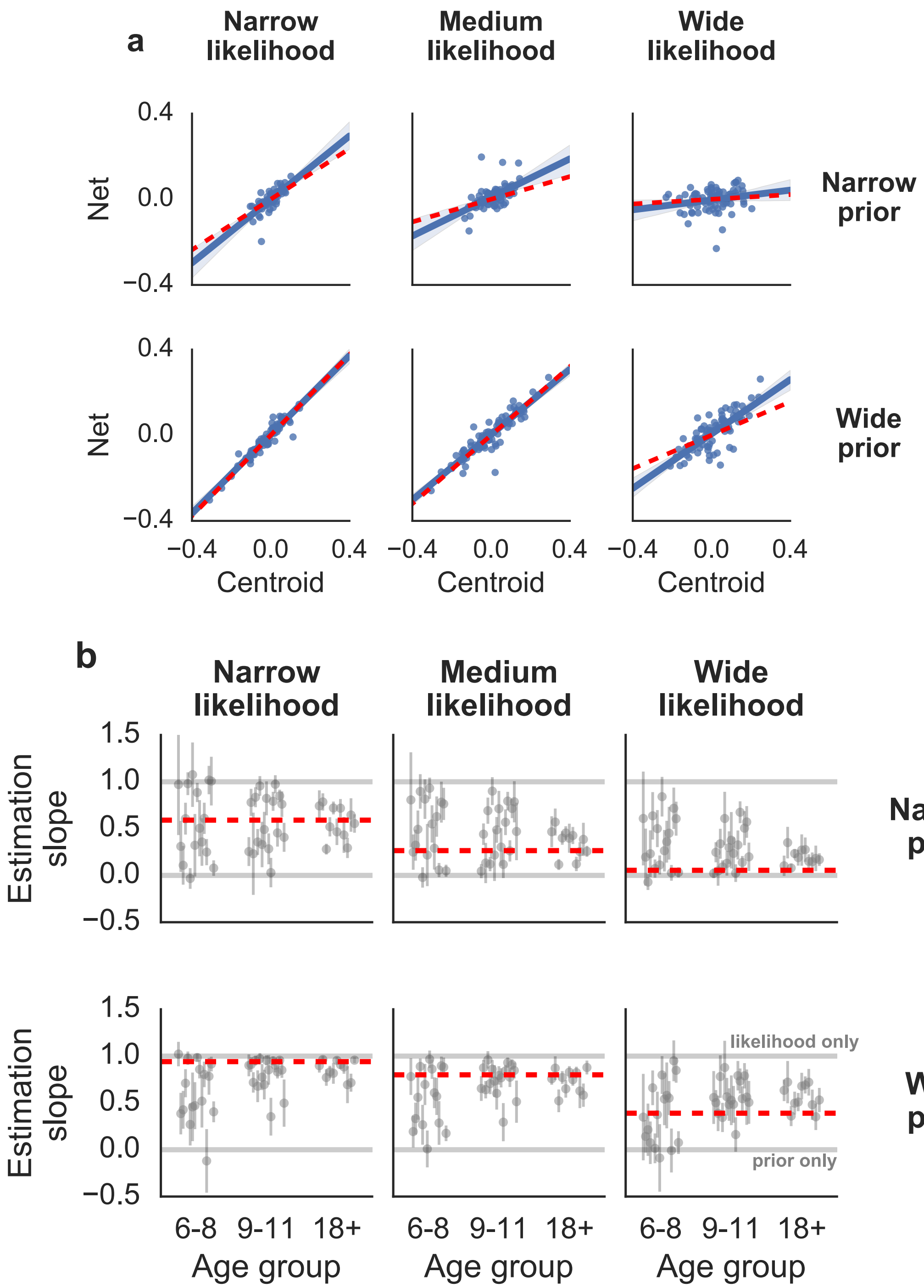

Figure 3. Estimation data. (**a**) Estimation data overlaid with linear fit for a representative participant aged 11 years old. The net position as a function of the centroid of the likelihood is shown for each trial (points). The fitted *Estimation slope* (blue line) and the optimal *Estimation slope* (dashed red line) are displayed. Each panel displays estimation data for one condition, as defined by prior and likelihood width. (**b**) The median bootstrapped *Estimation slope* is shown as a function of age group (error bars = 95% confidence intervals). The optimal *Estimation slope* values are shown (dashed red line).

In order to investigate sensorimotor estimation under uncertainty, we manipulated the variance of the prior and likelihood and measured the *Estimation slope* in each condition. In order to quantify the *Estimation slope*, we estimated the relationship between the centroid of the likelihood on each individual trial, c, and estimation, $\hat{s}$. Full reliance on the likelihood indicates a close relationship between estimations and the likelihood, *Estimation slope* = 1. Full reliance on the prior indicates a lack of relationship between estimations and the likelihood, *Estimation slope* = 0. To estimate the *Estimation slope* from the data, we performed linear regression on the data of individual participants for each condition, as defined by the prior width and likelihood width. The fitting procedure provides reasonable fits to the data for individuals (shown for an 11-year old, Fig. 3a), and across the entire data set (Fig. 3b). This allows us to quantify the nature of prior/likelihood integration for each participant.

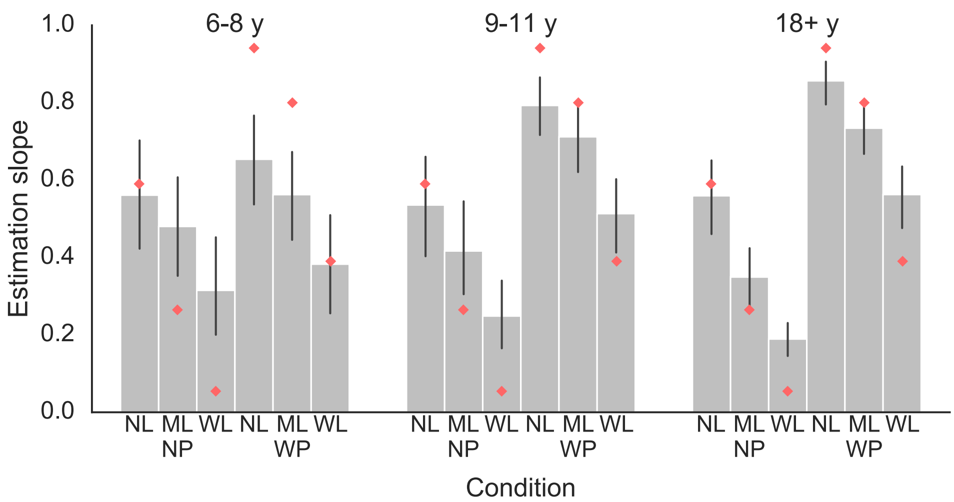

Figure 4. *Estimation slope* as a function of prior and likelihood for different age groups. The average *Estimation slope* is shown for all conditions and age groups, with error bars displaying the 95% CI and the optimal *Estimation slope* in each condition shown by red diamonds. The youngest age group distinguish between likelihood conditions (NL, ML, WL), but distinguish less between prior conditions (NP, WP). With age, there is a shift toward optimal weighing and greater use of the prior variance.

We were interested in whether sensorimotor integration of prior and likelihood improves during development. We, therefore, examined the relationship between the *Estimation slope* measured from the data and the optimal *Estimation slope*, ES_{opt}, i.e. the reliance on the likelihood that would maximize performance given the experimentally-imposed variance parameters. There is a shift toward the optimal *Estimation slope* during development in each condition (Fig. 3b). We formally examined the shift toward the statistical optimum using a distance-to-optimal score, which was the absolute distance from the statistical optimum in each condition averaged across conditions, leading to one score per participant (Fig. 5a). A Kruskal-Wallis test demonstrated a significant effect of age on *Distance to optimal*, (H(2) = 21.32, p < .0001), with significant differences between all age groups (Table 2). Therefore, sensorimotor estimation shifts toward an optimal integration of prior and likelihood during development.

Two possible contributors to the shift toward the optimal are participants' use of the prior and their use of the likelihood in computing estimations. We therefore examined the sensitivity to the prior variance, which quantifies the degree to which participants distinguished between the narrow and wide prior conditions: $Sensitivity\ to\ prior = (ES_{WP\ NL} - ES_{NP\ NL}) + (ES_{WP\ ML} - ES_{NP\ ML}) + (ES_{WP\ WL} - ES_{NP\ WL})$. We also examined the sensitivity to the likelihood, which quantifies the degree to which participants distinguished between the likelihood conditions: $Sensitivity\ to\ likelihood = (ES_{WP\ NL} - ES_{WP\ ML}) + (ES_{WP\ ML} - ES_{WP\ WL}) + (ES_{NP\ NL} - ES_{NP\ ML}) + (ES_{NP\ ML} - ES_{NP\ WL})$. Together, these two measures account for use of probabilistic information during the task.

We examined how sensitivity to the likelihood and prior change over the course of development. We found no evidence for an effect of age on *Sensitivity to likelihood* (Fig. 5b, H(2) = 5.21, p =.0739). *Sensitivity to likelihood* was significantly above zero in all age groups (6-8 y: median = .10, W = 1, p = .0016; 9-11 y: median = .13, W = 0, p = .0009; 18+ y: median = .17, W = 0, p = .0100), showing that children aged 6-8 years had already learned to distinguish between likelihood conditions. There was a significant effect of age on *Sensitivity to prior* (Fig 5c, H(2) = 19.13, p <.0001), with significant differences between age groups, except between 9-11 year olds and adults (Table 3). For children aged 6-8 years, *Sensitivity to prior* was not significantly above 0 (median = .03, W = 35, p = .2638). *Sensitivity to prior* was significantly above 0 in children aged 9-11 years (median = .27, W = 0, p = .0009) and in adult participants (median = .33, W = 0, p = .0100). Therefore, there is an increase in participants' sensitivity to the prior condition during development, and the ability to use the prior emerges at 9-11 years. During the task, children used likelihood information when it provides precise information on target location, and also learn to use the likelihood less when this source of information is uncertain. However, use of prior information is not fully developed in the child populations tested here. The ability to fully incorporate the prior into decisions increases over the course of development.

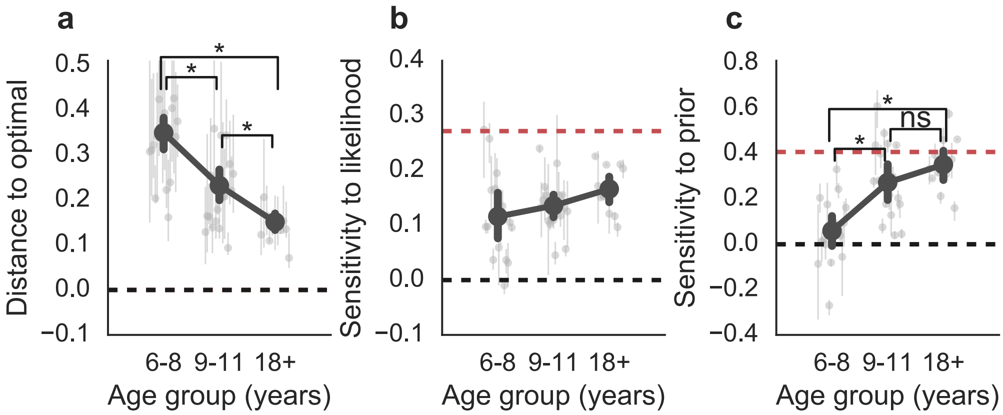

Figure 5. Relationship with statistical optimum, sensitivity to prior and likelihood as a function of age. (**a**) Distance to the optimal Estimation slope is shown for each participant. Here and in remaining plots, scores of individual participants are shown in gray overlaid with the mean for each age group (error bars = 95% CI). *Distance to optimal* decreases as a function of age group. *Distance to optimal* = 0 is shown (dashed line). (**b**) Sensitivity to the likelihood. *Sensitivity to likelihood* was computed as a difference in *Estimation slope* between likelihood conditions. There is no significant effect of age group. (**c**) Sensitivity to the prior. *Sensitivity to prior* was computed as the difference in *Estimation slope* between prior conditions. This increases as a function of age group, with significant differences between groups.

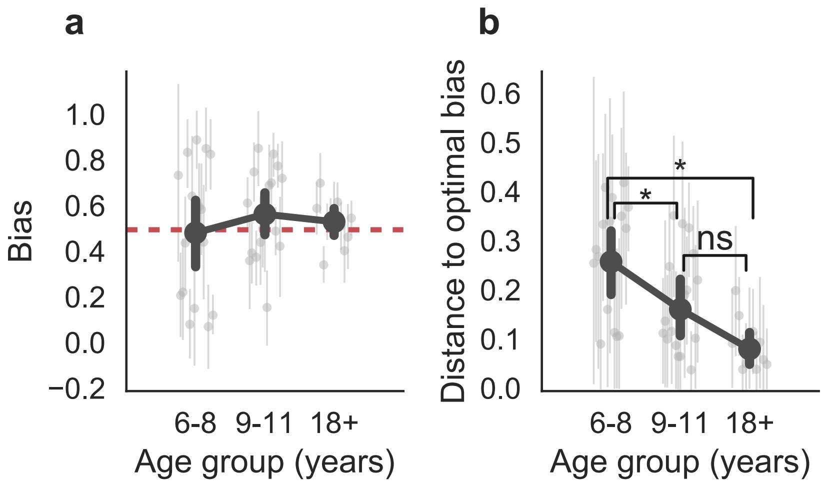

Figure 6. Overall bias. (**a**) Subject specific biases were quantified as a participant's overall tendency to rely on the likelihood or prior. *Estimation slope* was averaged across conditions for each individual to give a *Bias* score for each participant (median bootstrap with 95% CI as error bars). The mean bias is shown (error bar = 95% CI). The optimal bias is shown (dashed red line). Younger participants have a greater tendency to be biased, but biases are equally likely to be toward using the prior or likelihood. (**b**) The absolute difference between a participant's bias and the optimal is shown as a function of age. This decreases significantly with age, with significant differences between age groups (Table 4).

Since children's sensorimotor estimations were not fully explained by the integration of prior and likelihood, their responses may be partly driven by simple strategies. We investigated this possibility by quantifying the overall bias in participants' responses, by averaging *Estimation slopes* across all conditions. A *Bias* of 1 indicates that participants performed the task by always pointing to the likelihood regardless of condition and a *Bias* of 0 indicates a participant who always pointed at the prior regardless of condition. Biases appear to be more prevalent in the responding of individual participants at a younger age, but without an overall bias across participants (Fig. 6a). There is a significant effect of age on *Distance to optimal bias* (H(2) = 11.99, p =.0024, Fig. 6b). Comparisons between age groups were significant except between 9-11 year olds and adults (Table 4). Therefore, simple biases are more prominent in the responding of the youngest participant group.

Table 1: Mann-Whitney U test comparing *p(correct)* between age groups

Comparison	*Mann-Whitney U (n_1, n_2)*	*p*
6-8 y, 9-11 y	42 (16, 17)*	.0011
9-11 y, adults	25 (17, 11)*	.0021
6-8 y, adults	1 (16, 11)*	.0001

* In all tables, p<.05,
p-values are Bonferroni-corrected for # comparisons

Table 2: Mann-Whitney U test comparing *Distance to optimal* between age groups

Comparison	*Mann-Whitney U (n_1, n_2)*	*p*
6-8 y, 9-11 y	49 (16, 17)*	.0028
9-11 y, adults	47 (17, 11)*	.0457
6-8 y, adults	3 (16, 11)*	.0001

Table 3: Mann-Whitney U test comparing *Sensitivity to prior* between age groups

Comparison	*Mann-Whitney U (n_1, n_2)*	*p*
6-8 y, 9-11 y	42 (16, 17)*	.0011
9-11 y, adults	63 (17, 11)	.2373
6-8 y, adults	10 (16, 11)*	.0002

Table 4: Mann-Whitney U test comparing *Distance to optimal bias* between age groups

Comparison	*Mann-Whitney U (n_1, n_2)*	*p*
6-8 y, 9-11 y	75 (16, 17)*	.0439
9-11 y, adults	56 (17, 11)	.1227
6-8 y, adults	23 (16, 11)*	.0022

Discussion

We examined the development of Bayesian integration during sensorimotor estimation. We found that children aged 6-11 years consistently deviate from optimal use of statistical information relative to adults. Children from the age of 6 years used sensory information (likelihood), as adults did. However, use of the prior changed over the course of development. The youngest age group (6-8 years) did not distinguish between prior conditions, with use of the prior increasing with age. Children were also more likely to demonstrate overall biases in their estimations, which decreased during development. While young children used the uncertainty of an immediate source of information, they did not incorporate the distribution of targets over several trials into their judgments of target location.

We found that young children aged 6-8 years used the prior to a lesser degree than children aged 9-11 years and adults, even though they were shown target location at the end of each trial and samples from the prior were displayed on screen. A different manifestation of the experiment could have led young children to use the prior more. For example, if participants were given exposure to more trials in a longer experiment they may have eventually learned the prior, or actively engaging with the prior through trials with noiseless sensory feedback (likelihood) could have led children to incorporate the prior variance into their estimations. Nevertheless, our findings still show that older children and adults use priors in their estimations more readily than children do.

Cognitive explanations can be offered as to why young children learn to use the likelihood in their decisions, but incorporate the prior less. Using the likelihood in estimations involves learning the reliability of an immediate source of information when provided with feedback. Therefore, sensory processing of children and adults is comparable in this task. Using the prior variance involves integrating information about the target distribution over a longer timescale to learn its spatial distribution. Superior memory abilities of adults could allow them to learn the target distribution more successfully (Gathercole, 1999). More flexible decision making in adults may have allowed them to predict future target locations based on samples from the prior displayed on screen (Ernst, 2008). Children may have successfully acquired the prior but failed to integrate it with likelihood information. Our experiments do not allow us to distinguish between these possibilities.

For a theoretically-driven account of our findings we draw on the Reinforcement Learning literature. Model-based behavior leverages an understanding of the world's structure to predict successful actions (Doll, Simon, & Daw, 2012; Sutton & Barto, 1998). Using a model of the environment is not trivial. To do so, the agent must have already learned the dynamics and causal structure of the environment (Doll et al., 2012; Kording et al., 2007; Wei & Kording, 2012), and our findings, along with others, suggest that the ability to do so is not fully developed in children (Decker, Otto, Daw, & Hartley, 2016). Increase in use of the prior during development and a decrease in simple biases may reflect a progression from use of a simpler model to a more complex one. Weaker multisensory integration in children can also be understood as a failure of model-based behavior, i.e. children may have undeveloped models of the behavior of real-world objects, and therefore lack understanding of when information should be integrated or not (Ernst, 2008; Gori et al., 2008; Kording et al., 2007; Nardini et al., 2010, 2008). Further experiments are needed to fully understand the change in learning strategies in sensorimotor integration over development.

There is great interest in understanding intelligent human behavior to build artificial systems with the same degree of functionality (Lake, Ullman, Tenenbaum, & Gershman, 2016; Lecun, Bengio, & Hinton, 2015; Marblestone, Wayne, & Kording, 2016). If our aim is to understand human behavior, we may benefit from understanding the process by which human abilities are learned in the first place. The algorithmic approaches of the future may benefit from implementing a development-like process, beginning with the kind of simple biases demonstrated in young children and progressing toward the flexible and complex model-based abilities of adults (Decker et al., 2016; Spelke & Kinzler, 2007; Ullman, Harari, & Dorfman, 2012).

If Bayesian computation is at the core of the neural code (Beck et al., 2008; Ma et al., 2006; Zemel, Dayan, & Pouget, 1998), behavior should show the signatures of Bayesian inference under all conditions, including during development. Our results show that children do not use probabilistic information to the same extent as adults. The finding that children gradually shift toward the statistical optimum during development suggests that the brain learns to approximate Bayesian principles by means other than explicitly implementing Bayesian computations in neural circuits (Mandt et al., 2017; Rao, 2004). Our findings fit with ideas suggested by Jean Piaget on the role of constructivism in child development, i.e. that abilities are acquired through experience by building on more basic forms of knowledge (Piaget, 1954). In that sense, learning statistics may be seen as a very basic form of knowledge. While we may be born with a general learning architecture, it seems that statistics should not be seen as core knowledge (Spelke & Kinzler, 2007), but as an acquired skill.

References

Acuna, D. E., Berniker, M., Fernandes, H. L., & Kording, K. P. (2015). Using psychophysics to ask if the brain samples or maximizes. *Journal of Vision*, *15*(3), 7. http://doi.org/10.1167/15.3.7

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18327–32. http://doi.org/10.1073/pnas.1306572110

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., …

Pouget, A. (2008). Probabilistic Population Codes for Bayesian Decision Making. *Neuron*, *60*(6), 1142–1152. http://doi.org/10.1016/j.neuron.2008.09.021

Berniker, M., Voss, M., & Kording, K. (2010). Learning priors for bayesian computations in the nervous system. *PLoS ONE*, *5*(9), 1–9. http://doi.org/10.1371/journal.pone.0012686

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. *Psychological Science*, *27*(6), 848–58. http://doi.org/10.1177/0956797616639301

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, *22*, 1–7. http://doi.org/10.1016/j.conb.2012.08.003

Ernst, M. O. (2008). Multisensory Integration: A Late Bloomer. *Current Biology*, *18*(12), 519–521. http://doi.org/10.1016/j.cub.2008.05.003

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–33. http://doi.org/10.1038/415429a

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419.

Gopnik, A., & Wellman, H. M. (2013). Reconstructing constructivism: Causal models, Bayesian learning mechanisms and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. http://doi.org/10.1037/a0028044.Reconstructing

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, *18*(9), 694–698. http://doi.org/10.1016/j.cub.2008.04.036

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues : Optimal cue combination. *Journal of Vision*, (4), 967–992. http://doi.org/10.1167/4.12.1

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference.* New York, NY: Cambridge University Press.

Kording, K., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9). http://doi.org/10.1371/journal.pone.0000943

Kording, K., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning.

Nature, *427*(6971), 244–247. http://doi.org/10.1038/nature02169

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*(1), 186–196. http://doi.org/10.1037/0012-1649.43.1.186

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). Building Machines That Learn and Think Like People. *Behavioral and Brain Sciences*, *24*, 1–101.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. http://doi.org/10.1038/nature14539

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–8. http://doi.org/10.1038/nn1790

Ma, W. J., Beck, J. M., & Pouget, A. (2008). Spiking networks for Bayesian inference and choice. *Current Opinion in Neurobiology*, *18*, 217–222. http://doi.org/10.1016/j.conb.2008.07.004

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*(1), B1-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11525484

Mandt, S., Hoffman, M. D., & Blei, D. M. (2017). Stochastic Gradient Descent as Approximate Bayesian Inference. *arXiv Preprint: ArXiv*, *1704.04289*, 1–30.

Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an Integration of Deep Learning and Neuroscience. *Frontiers in Computational Neuroscience*, *10*(September), 1–41. http://doi.org/10.3389/fncom.2016.00094

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(39), 17041–6. http://doi.org/10.1073/pnas.1001699107

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, *18*(9), 689–693. http://doi.org/10.1016/j.cub.2008.04.021

Ouvrier, R., Goldsmith, R. F., Ouvrier, S., & Williams, I. C. (1993). The value of the Mini-Mental State Examination in childhood: a preliminary study. *Journal of Child Neurology*, *8*(2), 145–148. http://doi.org/10.1177/088307389300800206

Pitkow, X., & Angelaki, D. E. (2017). Perspective How the Brain Might Work : Statistics Flowing in Redundant Population Codes. *bioRxiv*, *1702.03492*, 1–9.

Rao, R. P. N. (2004). Bayesian Computation in Recurrent Neural Circuits. *Neural Computation*, *16*(1), 1–38.

Reynolds, C. R. (2004). *Behavior assessment system for children*. John Wiley & Sons, Inc.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*, 303–333. http://doi.org/10.1016/j.cogsci.2003.11.001

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. http://doi.org/10.1111/j.1467-7687.2007.00569.x

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT press.

Téglás, E., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. *Science*, *1054*. http://doi.org/10.1126/science.1196404

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640. http://doi.org/10.1017/S0140525X01000061

Ullman, S., Harari, D., & Dorfman, N. (2012). From simple innate biases to complex visual concepts. *Proceedings of the National Academy of Sciences*, *109*(44), 18215–18220. http://doi.org/10.1073/pnas.1207690109

Vilares, I., Howard, J. D., Fernandes, H. L., Gottfried, J. A., & Kording, K. P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Current Biology*, *22*(18), 1641–1648. http://doi.org/10.1016/j.cub.2012.07.010

Wei, K., & Kording, K. (2012). Causal Inference in Sensorimotor Learning and Control. In *Sensory Cue Integration*. http://doi.org/10.1093/ISBN

Wilson, B. N., Crawford, S. G., Green, D., Roberts, G., Aylott, A., & Kaplan, B. J. (2009). Psychometric properties of the revised developmental coordination disorder questionnaire. *Physical & Occupational Therapy in Pediatrics*, *29*(2), 182–202.

Wolraich, M. L., Lambert, W., Doffing, M. A., Bickman, L., Simmons, T., & Worley,

K. (2003). Psychometric properties of the Vanderbilt ADHD diagnostic parent rating scale in a referred population. *Journal of Pediatric Psychology*, *28*(8), 559–568.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, *10*(2), 403–430.