nature neuroscience

Human representation of visuo-motor uncertainty as mixtures of orthogonal basis distributions

Hang Zhang[1–6], Nathaniel D Daw[4–6] & Laurence T Maloney[4–6]

In many laboratory visuo-motor decision tasks, subjects compensate for their own visuo-motor error, earning close to the maximum reward possible. To do so, they must combine information about the distribution of possible error with values associated with different movement outcomes. The optimal solution is a potentially difficult computation that presupposes knowledge of the probability density function (pdf) of visuo-motor error associated with each possible planned movement. It is unclear how the brain represents such pdfs or computes with them. In three experiments, we used a forced-choice method to reveal subjects' internal representations of their spatial visuo-motor error in a speeded reaching movement. Although subjects' objective distributions were unimodal, close to Gaussian, their estimated internal pdfs were typically multimodal and were better described as mixtures of a small number of distributions differing only in location and scale. Mixtures of a small number of uniform distributions outperformed other mixture distributions, including mixtures of Gaussians.

Think about a baseball game. The batter has to decide whether and how to hit the incoming pitch. He needs to judge the position and speed of the ball, given his own visual uncertainty, and to estimate the probability of a successful swing, given his own visuo-motor uncertainty.

Visuo-motor decisions such as this are common in everyday life and have been studied in a rich and increasing body of laboratory tasks[1–3]. Human subjects are frequently found to compensate for their own sensorimotor uncertainty in ways that approximate an ideal Bayesian observer who maximizes expected reward[4–11]. Although plausible neural representations have been proposed for the combination of probabilistic information[12,13], little is known about representations of pdfs that capture visuo-motor error[14].

In the framework of Bayesian decision theory[2], the visuo-motor uncertainty associated with a possible reaching movement is summarized as a pdf on possible movement outcomes in space or time. The pdf is often close to Gaussian in form (**Fig. 1a**) and is centered on the point that the subject aims for. Suppose that the subject can gain a reward if she reaches to and hits a small target. A plot of the reward associated with each possible outcome is called the gain function $G(x)$, and here is either 0 (outside the target) or the promised reward (inside the target). If the subject aims at location a, then her expected gain on each attempt would be $EG = \int G(x)f(x - a)dx$, the integral of the product of the pdf with a gain function[2]. In **Figure 1b**, we illustrate the computation of expected gain when the subject aims at the center of the target.

The computation involved is potentially demanding, and a possible way to reduce the computational load is to use additive weighted mixtures of a fixed set of basis distributions $b_1(x), \ldots, b_n(x)$ to approximate the objective pdfs[15]

$$f(x) \approx \sum_{i=1}^{n} w_i b_i(x) \quad (1)$$

Two examples of a discrete mixture distribution are shown in **Figure 1a**, the first based on non-overlapping uniform basis distributions and the second on Gaussian distributions that all share a common variance, but differ in location. The Gaussian basis functions overlap, but—if they are sufficiently widely separated—they are effectively orthogonal for our purposes. We refer to such mixtures of a finite number of orthogonal or nearly orthogonal functions as discrete mixture distributions.

In three experiments, we estimated the internal pdfs used by human subjects in planning speeded reaching movements and compared them with their objective pdfs. We found that subjects' choice behavior was better described by (Bayesian optimal) decisions based on a mixture of discrete distributions than by single Gaussian distributions or other unimodal distributions, even though their actual motor error distributions were close to Gaussian, and that the mixture of non-overlapping uniform distributions (U-mix; **Fig. 1a**) outperformed other discrete mixture distributions, including mixture of Gaussians. We found that the number of basis functions in the discrete mixture representation needed to account for human performance is small, roughly 2–6.

Discrete weighted mixture representations can speed computation of expected gain: if the expected gain for each basis function can be

[1]Department of Psychology and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China. [2]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China. [3]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [4]Department of Psychology, New York University, New York, New York, USA. [5]Center for Neural Science, New York University, New York, New York, USA. [6]Institute for the Interdisciplinary Study of Decision Making, New York University, New York, New York, USA. Correspondence should be addressed to H.Z. (hang.zhang@pku.edu.cn).

Received 18 March; accepted 5 June; published online 29 June 2015; doi:10.1038/nn.4055

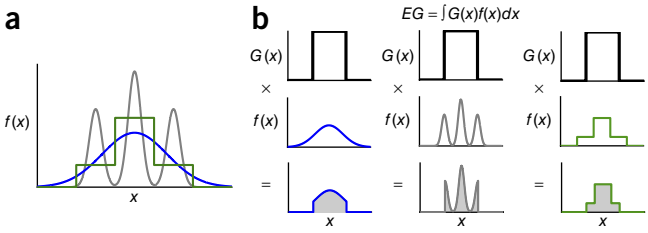


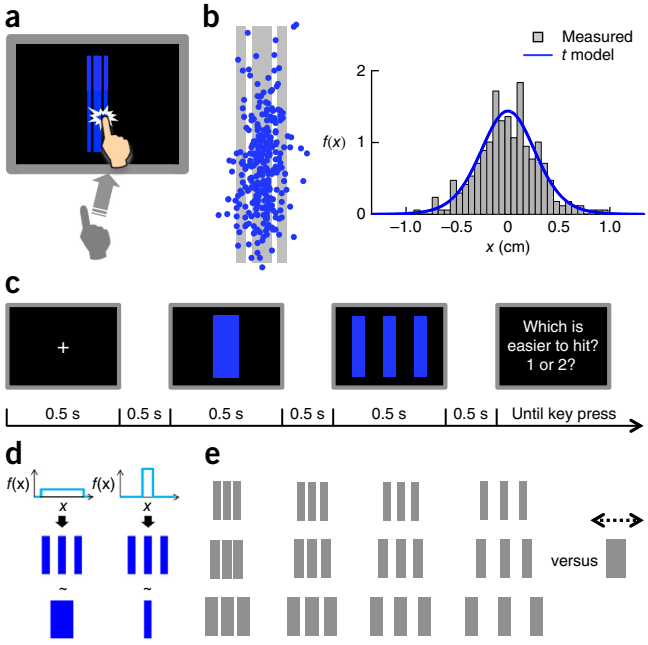
Figure 1 Computation of expected gain. (**a**) Three examples of pdfs: Gaussian distribution (blue), an mG-mix mixture distribution (gray) and a U-mix mixture distribution (green). (**b**) The computation of expected gain (EG) for each. $G(x)$ denotes the gain function and $f(x)$ denotes the pdf. The areas of the shaded regions in the plots of $G(x) \times f(x)$ correspond to $EG = \int G(x) f(x) dx$.

computed, the expected gain associated with alternative targets or movement plans will be reduced to weighted linear summation of the contributions from each basis distribution. The use of mixtures of distributions is also relevant to Bayesian model averaging, models based on mixtures of experts and hierarchical formulations of motor control (for example, hierarchical MOSAIC[16]).

Each experiment consisted of two phases, training and choice. The results of the training phase allowed us to estimate subjects' objective pdfs and the results of the choice phase allowed us to estimate their internal pdfs.

Figure 2 illustrates the task and design of experiment 1. Human subjects were first trained to repeatedly reach to a specific target on the screen within a time limit (**Fig. 2a**). Typically, their endpoints had a distribution close to bivariate Gaussian (**Fig. 2b**).

In a second phase, subjects viewed two virtual targets differing in width and configuration and chose the target they preferred to try later for monetary rewards (**Fig. 2c**). We assumed subjects' choice was generated by a softmax function based on their estimates of the expected utilities of the two targets (Online Methods), which, under our reward structure (hit = fixed reward, miss = nothing), were reduced to the probabilities of hit. Thus, subjects' choices were determined by their internal pdfs, integrated over the target regions (see **Fig. 2d** for an illustration).



Conversely, we could reconstruct approximations of subjects' internal pdfs from their choices[17].

The targets were vertically elongated so that only horizontal error affected reward. On each trial, one target was a triple (three identical, equally gapped rectangles) and the other was a single (one rectangle). We chose the triple target as a convenient way to explore the distribution of probability mass in the tails of the internal pdfs by varying the gap between the outer rectangles and the inner (**Fig. 2e**).

RESULTS

Experiment 1: objective pdf

We ignore the irrelevant vertical direction and describe only the horizontal statistics. Subjects' endpoints in the reaching task (**Fig. 2b**) had a Gaussian-like distribution that was symmetric around the target center. The distribution of all but one subject's visuo-motor error (endpoints' deviation from the mean endpoint) had a kurtosis higher than that of Gaussian (by 0.04–1.78, median of 0.44), indicating a more peaked center or heavier tail. We modeled each subject's visuo-motor error as a scaled Student's t distribution with a scale parameter and a shape parameter (Online Methods), for which the Gaussian distribution is a limiting case. The t model captured individual subjects' visuo-motor error in s.d. (Pearson's $r = 1.0$, $P < 0.001$) and kurtosis (Pearson's $r = 0.82$, $P = 0.004$). We refer to the t distribution estimated in a subject's reaching task as the subject's objective visuo-motor error distribution, or objective pdf.

Experiment 1: internal pdf

We first visualized subjects' internal pdfs from their choices, assuming a Gaussian-process prior (Online Methods), which results in a smoothed estimate of the pdf; it assumes only a weak correlation between adjacent locations on the pdf and allows for the possibility that the underlying distribution is multimodal.

It should be noted that the resulting fits are effectively smoothed and any abrupt changes in the pdf may be lost in this analysis. In addition, because of the choice of stimuli in experiment 1, we could not reliably estimate the probability density in a small interval at the center of the pdf (we removed this limitation in experiment 2, Online Methods) and we replaced it by a horizontal bar.

For the single subject's data shown in **Figure 3a**, there appeared to be multiple discrete modes (peaks) or possibly steps. The presence of such modes is inconsistent with the unimodal form of the objective pdf (that is, t distribution). Results for all subjects are shown in **Supplementary Figure 1**.

Figure 2 Task and design of experiment 1. (**a**) The reaching task. Subjects were required to hit the target (blue regions) at the center of the screen within 400 ms for each of 300 trials. (**b**) One subject's visuo-motor error distribution in the reaching task. Left, the endpoint of each trial (blue dot) is marked on the target (gray regions). Right, the distribution of horizontal visuo-motor error fits to a scaled Student's t distribution. (**c**) Time course of the choice task. On each trial, subjects chose between a triple (an array of three rectangles) and a single (one rectangle) as to which target that they perceived to be easier to hit. The subject did not reach to hit the target; she only chose a preferred target. (**d**) Rationale of the choice task. We illustrate two extreme cases. Left, when the subject' internal pdf of visuo-motor error is a uniform distribution that is wide enough to contain the whole triple, the subject would be indifferent (denoted by ~) between the two targets when the width of the single equals the total width of the three rectangles in the triple. Right, in contrast, when the subject's model is a uniform distribution that covers only the central rectangle of the triple, the equivalent single would be merely as wide as the central rectangle. (**e**) Design of the choice task. 12 different triples were used, for each of which, the width of its paired single was adjusted by an adaptive procedure for 70 trials.

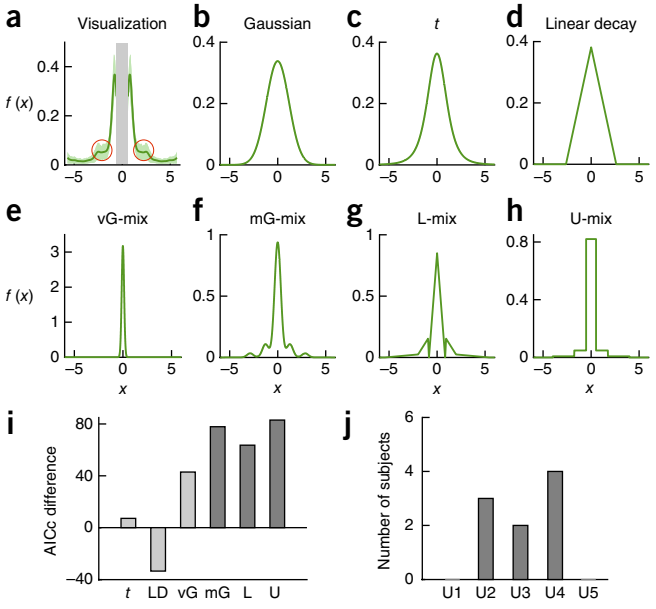


Figure 3 Internal pdfs in the choice task of experiment 1. (**a**) Non-parametric visualization of the internal pdf for one subject. Green-shaded regions denote ±s.e.m. *x* is in the unit of the subject's horizontal s.d. estimated from the reaching task. The gray-shaded central range of [−0.6, 0.6] could not be reliably estimated in experiment 1 (Online Methods) and the visualization therefore gives information about the pdf only away from the origin. Two regions of interest are marked by red circles. The visualizations for all subjects are shown in **Supplementary Figure 1**. (**b**–**h**) Internal pdfs estimated from different models for the same subject. (**i**) AICc difference between the Gaussian model and the other six models summed over the nine subjects. The unimodal models (including vG-mix) and mixture models are coded in light gray and dark gray, respectively. Positive difference indicates better fit. LD denotes linear decay. (**j**) Number of subjects best fit by each U-mix model.

The results of this analysis suggest, but do not demonstrate, that subjects' internal pdfs are multimodal. We used a model comparison procedure to further explore the form of the internal pdf.

We compared seven different classes of models of the internal pdf: three unimodal distributions, a mixture distribution that is always unimodal and three mixture distributions that could be multimodal (Online Methods). The Akaike information criterion with a correction for sample sizes (AICc)[18,19] was used for model selection.

Unimodal distributions. The first and the baseline model was the Gaussian model, whose variance was fitted as a free parameter. The second model was the *t* model, whose scale and shape parameters were free. In a third model, the linear-decay model, we assumed that the probability density functions in subjects' internal pdfs were continuous, but took the simple linear form of a triangular distribution, with variance as a free parameter.

Mixture distributions. We next considered four classes of mixture distributions (including the unimodal mixture distribution). One class is the linear combination of multiple uniform distributions (that is, U-mix) whose ranges abut one another. We assumed symmetry for the current problem: a U-mix model with *n* components was composed of *n* pairs of uniform distributions symmetric around 0 (that is, two symmetric uniform distributions were counted as one component). The U-mix distributions shown in **Figure 1**, for example, had two components. For each subject, we constructed five levels of U-mix models with increasing number of components, denoted U1–U5, and fit them to the subject's choices. The ranges (spatial extent) and weights (heights of the components) of the uniform components were free parameters.

Two classes of mixtures of Gaussian distributions were modeled: the vG-mix model is a linear combination of Gaussian distributions with the same mean but different variances and the mG-mix is a linear combination of Gaussian distributions with the same variance, but different means. The mean(s), variance(s) and weights of the Gaussian components were fitted as free parameters. The vG model, described among mixture distributions for convenience, was classified as a unimodal model.

A vG-mix or mG-mix with *n* components had the same number of free parameters as a U-mix with *n* components. Similar to U-mix, we constructed five levels of vG-mix and mG-mix and obtained the best-fit vG-mix and mG-mix. Last, we considered a mixture model composed of piecewise linear components, denoted as L-mix.

The pdf estimated from each model is plotted for one subject in **Figure 3b–h**. The AICc differences between the baseline (Gaussian) and the other six models (summed over the nine subjects) are shown in **Figure 3i**. Two conclusions can be reached. First, all of the mixture distribution models fitted better to subjects' choice patterns than any of the unimodal distributions. The unimodal distribution models tended to smooth over any abrupt changes in subjects' choice patterns, whereas the mixture distribution models, with their multiple discrete modes or steps, were able to capture them (**Supplementary Fig. 2**).

Second, the U-mix model outperformed the other mixture models. According to the

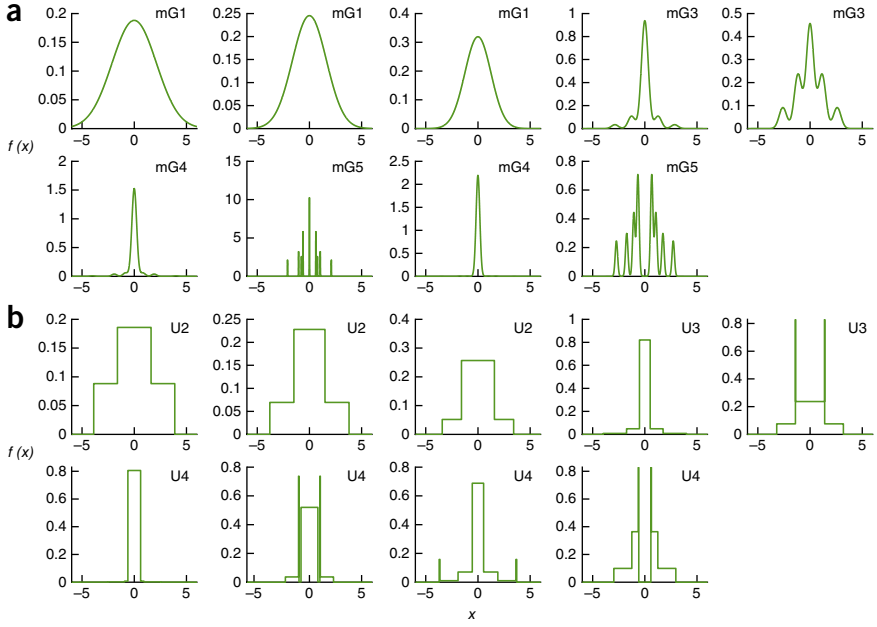


Figure 4 Model fits of all subjects' internal pdfs in experiment 1. (**a**) mG-mix model. (**b**) U-mix model. Each panel shows the pdf of one subject. *x* is presented in the unit of the subject's horizontal s.d. estimated from the reaching task. Subjects are shown in the same order as in **Supplementary Figure 1**.

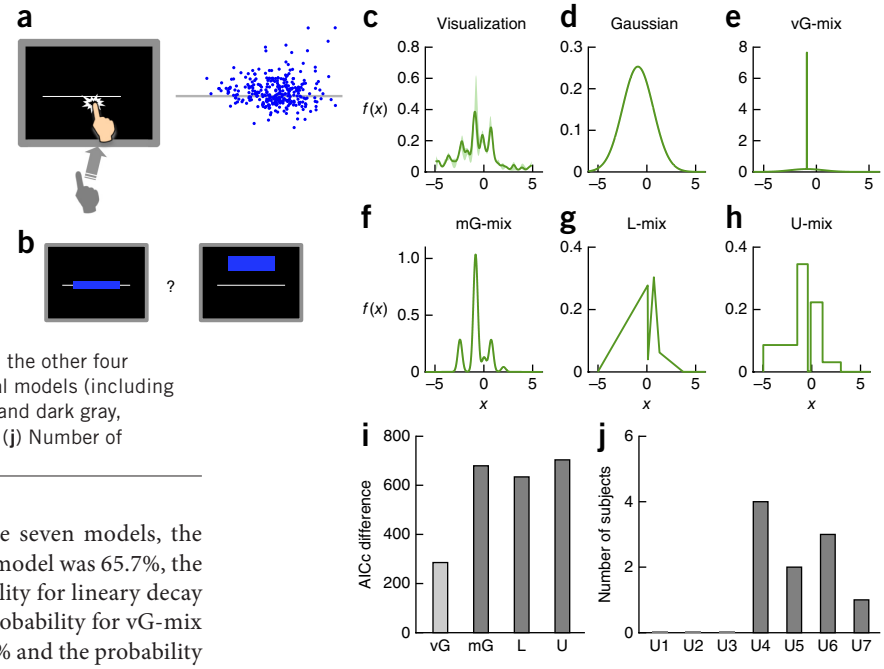


Figure 5 Experiment 2. (**a**) The reaching task. Left, the task. The task for experiment 2 was the same as that for experiment 1, except that a horizontal line was used as the target. Right, the endpoints for one subject. (**b**) The choice task. The task for experiment 2 was similar to that for experiment 1, but each pair of targets consisted of a rectangle on the line and a rectangle off the line. (**c**) Non-parametric visualization of the internal pdf for one subject. Shaded regions denote ±s.e.m. x is in the unit of the subject's vertical s.d. estimated from the reaching task. (**d**–**h**) Internal pdfs estimated from different models for the same subject. (**i**) AICc difference between the Gaussian model and the other four models summed over the ten subjects. The unimodal models (including vG-mix) and mixture models are coded in light gray and dark gray, respectively. Positive difference indicates better fit. (**j**) Number of subjects best fit by each U-mix model.

group-level Bayesian model selection[20], of the seven models, the probability for the U-mix model to be the best model was 65.7%, the probability for Gaussian was 1.3%, the probability for lineary decay was 1.0%, the probability for t was 1.0%, the probability for vG-mix was 1.3%, the probability for mG-mix was 28.0% and the probability for L-mix was 1.7%.

The underlying Bayes factors provide only weak evidence favoring the uniform mixture model over the Gaussian mixture model. However, in our third experiment (when considering two-dimensional basis functions), we found stronger evidence for uniform over Gaussian mixtures (see below).

Model fits of all subjects' internal pdfs are shown in **Figure 4** for the mG-mix and U-mix models. The U-mix models were evidently discrete mixture representations, as defined in above. The 'runner up' mG-mix models were also discrete mixture representations: the basis pdfs in the mG-mix models had, on average, an overlap of only 8.0% in probability density (Online Methods); that is, they were close to non-overlapping, orthogonal.

According to AICc comparisons, the U-mix models that best fit subjects' choices contained only a small number of components (**Fig. 3j**): all of the subjects were best fit by U2–U4. For example, the best-fit U-mix model for one subject shown in **Figure 3h** was U3.

Experiment 2

We wanted to show that the discrete mixture representation of visuo-motor uncertainty was not somehow a result of the particular stimuli we used in experiment 1. Could the triples, in particular, have somehow led the subject to choose a discrete mixture representation that she might otherwise never have made use of?

In a second experiment, we trained subjects to touch a line within a time limit (**Fig. 5a**). The distribution of each subject's endpoints was close to Gaussian. Instead of an all-or-none reward, the amount subjects received on each trial could be any integer between 0 and 100, decreasing with the distance of their endpoint to the target line. In the subsequent choice phase, subjects chose between a central rectangular region and a peripheral rectangular region (**Fig. 5b** and Online Methods).

The analyses of experiment 2 were similar to those of experiment 1, except that the design of the experiment allowed us to include asymmetric distributions among the candidates for subjects' internal pdfs and there was no restriction on estimating the pdf near its center as there was in experiment 1 (Online Methods). The visualization and model fits for one subject are shown in **Figure 5c**–**h** (also see **Supplementary Figs. 3** and **4**). Again, all the mixture models were superior to the Gaussian and the U-mix model was superior to the other mixture models in AICc (**Fig. 5i**). Not only the U-mix models, but also the mG-mix models, were discrete mixture representations: the basis pdfs in the mG-mix models had, on average, an overlap of only 9.2% in probability density. They were close to non-overlapping, orthogonal. According to group-level Bayesian model selection[20], among the five models, the probability for the U-mix model to be the best model was 63.4%; the probability for Gaussian was 0.7%, the probability for vG-mix was 22.3%, the probability for mG-mix was 12.7% and the probability for for L-mix was 0.9%. The best-fit U-mix models for most subjects were U4–U6 (**Fig. 5j**).

Experiment 3

For experiment 3, we applied the tests developed above to the two-dimensional choice data of ref. 21. Its task and design was the same as that of experiment 1 with the following exceptions (**Fig. 6a**–**c**). The target of the reaching task was a circle and subjects' visuo-motor error had a bivariate Gaussian distribution. The targets in the choice task were a rectangle and a circle.

We modeled subjects' internal pdfs in the horizontal and vertical directions separately and considered the Gaussian, vG-mix, mG-mix, L-mix and U-mix models (Online). As in experiments 1 and 2, the mixture models were superior to the Gaussian model and the U-mix model was superior to the other mixture models in AICc (**Fig. 6d**). Not only the U-mix models, but also the mG-mix models (**Supplementary Figs. 5** and **6**), were discrete mixture representations; the basis pdfs in the mG-mix models had, on average, an overlap of only 17% in probability density. According to group-level Bayesian model selection[20], of the five models, the probability for the U-mix model to be the best model was 97.6%, the probability for Gaussian was 0.02%, the probability for vG-mix was 0.3%, the probability for mG-mix was 0.5%, the probability for L-mix was 1.6%.

For a mixture model that had n_x components in the horizontal direction and n_y components in the vertical direction, define its number of components as $(n_x + n_y)/2$. The best-fit U-mix model for most subjects had 2–3 components (**Fig. 6e**).

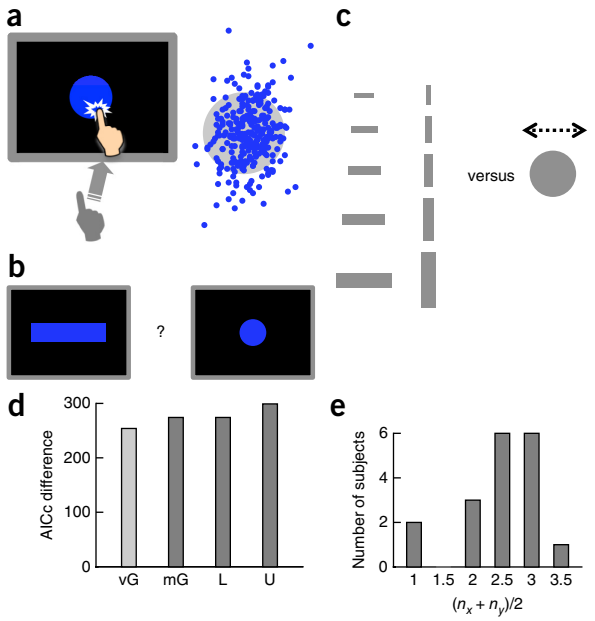


Figure 6 Experiment 3. (**a**) The reaching task. Left, the task. The task for experiment 3 was the same as that for experiment 1, except that a circular target was used. Right, the endpoints for one subject. (**b**) The choice task. The task for experiment 3 was the same as that for experiment 1, except that each pair of targets was a rectangle and a circle. (**c**) Design of the choice task. Ten different rectangles were used; for each, the radius of its paired circle was adjusted by adaptive procedures for 100 trials. (**d**) AICc difference between the Gaussian model and the other four models summed over the 18 subjects. The unimodal models (including vG-mix) and mixture models are coded in light gray and dark gray, respectively. Positive difference indicates better fit. (**e**) Number of subjects best fit by each U-mix model.

DISCUSSION

We estimated human subjects' internal model of their own visuo-motor error in a speeded reaching task and compared it with their objective visuo-motor error. Subjects' actual visuo-motor error distributions (objective pdfs) were in all cases unimodal, close to a Gaussian in many cases and close to a *t*-distribution in the remainder. However, the distributions implicit in their choices (internal pdfs) were very different from their actual distributions.

We found, first of all, that multimodal mixture distributions (for example, U-mix, mG-mix) provided a better fit to subjects' choice patterns than any single Gaussian or the alike. Second, of the mixture models tested, a model consisting of a mixture of non-overlapping uniform distributions with two to six nonzero steps performed best.

Both results are unexpected. The first and broader conclusion—that subjects' internal pdfs were mixtures of basis distributions—is well supported by our data. Moreover, the basis distributions in the mixtures either had no overlaps (for example, U-mix) or just slight overlaps (for example, mG-mix). We refer to such mixtures of local distributions as discrete mixture representations. The subjects' representations of their own visuo-motor error were discrete.

It is less clear what the form of the basis distributions was: the mG-mix fits came close to the U-mix in both goodness-of-fit and overall appearance. Although fits to the data favored uniform basis distributions over Gaussian, there could well be a third candidate that would dominate both.

There are two possible advantages of using uniform rather than Gaussian mixtures to represent probability mass. First, we can 'tile' the space of events in an orthogonal (sparse) fashion, without any bias to a particular location. Second, the probability assigned to each event (here, endpoint) depends only on the tile it is in. That is, we can estimate the constant probability density of each tile by simply counting events in the tile.

Relationship to previous measures

A few studies have reconstructed human subjects' representation of sensory probability distributions based on their decisions. One study found that subjects' internal pdf of a Gaussian prior distribution closely followed the objective prior (Fig. 2d of ref. 7), which seemingly disagrees with a discrete mixture representation of probability distributions. However, this study[7] showed is that subjects computed a weighted average of the mean of a prior distribution and the mean of a likelihood. Although such averaging is consistent with Bayesian inference based on Gaussians, it is unclear how to infer from their data that subjects actually maintained and multiplied specifically Gaussian distributions[22]. Notably, a second condition in the same study demonstrated that subjects successfully represented a bimodal prior distribution, a capability that is, broadly, consistent with our finding that the brain employs mixture distributions. Of course, a formal test of whether behavior in such a setting is best fit by particular forms of mixture, such as our U-mix, will be needed in the future.

One surprising feature of the discrete mixture representation that we observed was that it was multimodal, although the true distribution was unimodal. Although this outcome was unexpected, it is not completely unprecedented: subjects' representation of temporal prior distributions in a previous study (Figs. 7–9 in ref. 23) appeared to have more than one mode.

Discrete representation and near-optimal motor decisions

Our finding that subjects' internal pdfs of their own visuo-motor error distribution were discrete, thereby deviating systematically from the objective distribution, does not necessarily conflict with near-optimal human performance in previous studies[4–10] (see ref. 11 for an example of a binary choice task). Many tasks may simply be insensitive to systematic deviations in subjects' internal pdfs. For example, a previous study[21] found that a virtual subject with a Gaussian error distribution, but who mistakenly assumes it is a uniform distribution of the same variance, would still be able to achieve near-optimal performance in a previously described visuo-motor decision task[8]. A discrete mixture representation with two to six nonzero steps, as we found in our experiments, enables even better approximations to the objective distribution and can therefore lead to near-optimal performance as well.

Simplifying probabilistic calculation

Psychologists and neuroscientists modeling biological computation have encountered the computational problems that arise when manipulating high-dimensional or continuous distributions in many guises. Broadly speaking, tractable solutions require approximating the exact computation with some simpler, sparser form. The discrete mixture representation that we propose is one example, and shares its essential feature of sparseness with many other approaches, such as approximating distributions with a reduced rank form[24] or a kernel density estimator[25], with Monte Carlo approximations that substitute a few samples for a random variable[26–29], or with the use of linear models to approximate surface spectral reflectance density functions[30]. Given that these approaches share many essential similarities, it is possible that all arise from the same neural solution to complexity.

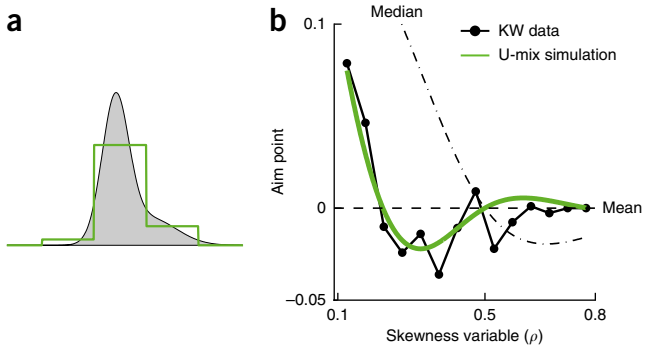


Figure 7 U-mix simulation for ref. 31. We simulated a virtual subject in the task described in ref. 31. We assumed that the subject had a quadratic loss function, but that she approximated the skewed distributions with a U-mix with three nonzero probability categories (Online Methods). (**a**) Illustration of the discrete mixture representation. The U-mix is centered on the median of the to-be-represented distribution. The half of its total width is 3.3× of the s.d. of the distribution. (**b**) Simulated aim points versus data. The simulated aim points (green curve) agreed well with their data (black curve, re-plot from their **Figure 2b**, noninverted trials, KW data). Note the U-shaped trend below zero, a deviation from the mean (dashed line) but toward the inverse direction of the median (dot-dashed line), could not be explained by the model described in ref. 31.

An important question for future work is whether the U-mix (or other mixture) distributions that we observed are the only internal representations of the distribution of visuo-motor error available to the visuo-motor system or whether they are transient representations, derived from a more accurate representation, that vary with the task imposed. That is, the neural system could maintain a high-resolution representation of visuo-motor error, but uses simplified representations to carry out specific computational tasks, just as most common programming languages use a variety of numerical representations.

Discrete representation as explanation for decision biases

The U-mix representation may be used to approximate any arbitrary probability distribution. Using a sensorimotor decision task, a previous study[31] found, for skewed error distributions, a deviation of subjects' aim point from the mean toward the median. The authors[31] estimated subjects' internal loss functions that would lead to this bias. However, in a close examination of their data, we noticed an unexplained away-from-median bias (that is, toward the longer or fatter tail) in the middle range of skewness—the U-shaped curve trends below zero. The coexistence of the toward-median and away-from-median biases could not be explained by any of the loss functions proposed previously[31] and it is not obvious whether there exists such a loss function that is non-negative.

However, if we assume subjects had a quadratic internal loss function[32–34], but employed a discrete mixture representation (U-mix) of their error distribution, we can reproduce the pattern of the previous study's data[31], both the toward-median and away-from-median biases, with satisfying precision (**Fig. 7** and Online Methods). Intuitively, the two opposite biases stem from two complementary effects. First, discretization trims the skewed tail outside the discrete range, leading to the underweighting of large errors. Second, discretization homogenizes the density in each tile, effectively moving the probability mass away from the shorter tailed side toward the longer tailed side.

An mG-mix could lead to a similar effect as U-mix (**Supplementary Fig. 7**). However, given that a mixture of two Gaussians is exactly the error distribution used previously[7,31], we had to choose a mixture of two Gaussians that differed from the actual mixture to get the pattern of biases present in the human data.

The discrete mixture representation that we propose for sensorimotor error can serve as a general framework for human representation of probability distributions and can potentially explain a range of known biases in human choices. For example, humans exhibit a skewness preference, a well-documented phenomenon in economics and finance[35–38]: they prefer reward distributions with positive skewness to those with zero or negative skewness when the mean and variance of the distributions are the same, exhibiting the pattern we found in previous data[31], which was consistent with a U-mix representation.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors would like to thank J. Tee for inspiring discussions. H.Z. and L.T.M. were supported by grant EY019889 from the US National Institutes of Health and L.T.M. by an award from the Alexander v. Humboldt Foundation. N.D.D. was supported by a Scholar Award from the McKnight Foundation and a James S. McDonnell Foundation Award in Understanding Human Cognition.

AUTHOR CONTRIBUTIONS

H.Z. designed and performed the experiments, analyzed the data and wrote the manuscript. N.D.D. and L.T.M. supervised the project and improved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Bach, D.R. & Dolan, R.J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **13**, 572–586 (2012).
2. Maloney, L.T. & Zhang, H. Decision-theoretic models of visual perception and action. *Vision Res.* **50**, 2362–2374 (2010).
3. Trommershäuser, J., Maloney, L.T. & Landy, M.S. Decision making, movement planning and statistical decision theory. *Trends Cogn. Sci.* **12**, 291–297 (2008).
4. Battaglia, P.W. & Schrater, P.R. Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *J. Neurosci.* **27**, 6984–6994 (2007).
5. Faisal, A.A. & Wolpert, D.M. Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. *J. Neurophysiol.* **101**, 1901–1912 (2009).
6. Hudson, T.E., Maloney, L.T. & Landy, M.S. Optimal compensation for temporal uncertainty in movement planning. *PLOS Comput. Biol.* **4**, e1000130 (2008).
7. Körding, K.P. & Wolpert, D.M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).
8. Trommershäuser, J., Maloney, L.T. & Landy, M.S. Statistical decision theory and trade-offs in the control of motor response. *Spat. Vis.* **16**, 255–275 (2003).
9. Jazayeri, M. & Shadlen, M.N. Temporal context calibrates interval timing. *Nat. Neurosci.* **13**, 1020–1026 (2010).
10. Wei, K. & Körding, K. Uncertainty of feedback and state estimation determines the speed of motor adaptation. *Front. Comput. Neurosci.* **4**, 11 (2010).
11. Trommershäuser, J., Landy, M.S. & Maloney, L.T. Humans rapidly estimate expected gain in movement planning. *Psychol. Sci.* **17**, 981–988 (2006).
12. Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006).
13. Huys, Q.J., Zemel, R.S., Natarajan, R. & Dayan, P. Fast population coding. *Neural Comput.* **19**, 404–441 (2007).
14. Pouget, A., Beck, J.M., Ma, W.J. & Latham, P.E. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16**, 1170–1178 (2013).
15. Maloney, L.T. Statistical decision theory and biological vision. in *Perception and the Physical World: Psychological and Philosophical Issues in Perception* (eds. D. Heyer & R. Mausfeld) 145–189 (Wiley, New York, 2002).
16. Haruno, M., Wolpert, D.M. & Kawato, M. Hierarchical MOSAIC for movement generation. *Int. Congr. Ser.* **1250**, 575–590 (2003).
17. Maloney, L.T. & Mamassian, P. Bayesian decision theory as a model of human visual perception: testing Bayesian transfer. *Vis. Neurosci.* **26**, 147–155 (2009).

18. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
19. Hurvich, C.M. & Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika* **76**, 297–307 (1989).
20. Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J. & Friston, K.J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).
21. Zhang, H., Daw, N.D. & Maloney, L.T. Testing whether humans have an accurate model of their own motor uncertainty in a speeded reaching task. *PLOS Comput. Biol.* **9**, e1003080 (2013).
22. Oruç, I., Maloney, L.T. & Landy, M.S. Weighted linear cue combination with possibly correlated error. *Vision Res.* **43**, 2451–2468 (2003).
23. Acerbi, L., Wolpert, D.M. & Vijayakumar, S. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLOS Comput. Biol.* **8**, e1002771 (2012).
24. Daw, N.D., Courville, A.C. & Dayan, P. Semi-rational models of conditioning: the case of trial order. in *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (eds. N. Chater & M. Oaksford) 431–452 (Oxford University Press, Oxford, 2008).
25. Gershman, S. & Wilson, R. The neural costs of optimal control. *Adv. Neural Inf. Process. Syst.* **23**, 712–720 (2010).
26. Vul, E., Goodman, N.D., Griffiths, T.L. & Tenenbaum, J.B. One and done? Optimal decisions from very few samples. *Cogn. Sci.* **38**, 599–637 (2014).
27. Sanborn, A.N., Griffiths, T.L. & Navarro, D.J. Rational approximations to rational models: alternative algorithms for category learning. *Psychol. Rev.* **117**, 1144 (2010).
28. Vul, E., Hanus, D. & Kanwisher, N. Attention as inference: selection is probabilistic; responses are all-or-none samples. *J. Exp. Psychol. Gen.* **138**, 546–560 (2009).
29. Daw, N.D. & Courville, A. The pigeon as particle filter. in *Advances in Neural Information Processing Systems* (ed. J.C. Platt, D. Koller, Y. Singer & S. Roweis) 369–376 (MIT Press, 2007).
30. Maloney, L.T. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A* **3**, 1673–1683 (1986).
31. Körding, K.P. & Wolpert, D.M. The loss function of sensorimotor learning. *Proc. Natl. Acad. Sci. USA* **101**, 9839–9842 (2004).
32. Todorov, E. & Jordan, M.I. Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* **5**, 1226–1235 (2002).
33. Harris, C.M. & Wolpert, D.M. Signal-dependent noise determines motor planning. *Nature* **394**, 780–784 (1998).
34. Wolpert, D.M., Ghahramani, Z. & Jordan, M.I. An internal model for sensorimotor integration. *Science* **269**, 1880–1882 (1995).
35. Hamilton, B.H. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *J. Polit. Econ.* **108**, 604–631 (2000).
36. Harvey, C.R. & Siddique, A. Conditional skewness in asset pricing tests. *J. Finance* **55**, 1263–1295 (2000).
37. Kraus, A. & Litzenberger, R.H. Skewness preference and the valuation of risk assets. *J. Finance* **31**, 1085–1100 (1976).
38. Moskowitz, T.J. & Vissing-Jørgensen, A. The returns to entrepreneurial investment: a private equity premium puzzle? *Am. Econ. Rev.* **92**, 745–778 (2002).

ONLINE METHODS

Ethics statement. The experiments were approved by the University Committee on Activities Involving Human Subjects of New York University. Informed consent was given by each subject before the experiment.

Subjects. There were, respectively, 10 (1 male), 12 (4 male) and 18 (8 male, 4 left-handed) subjects, aged 18–40, that participated in the three experiments. All subjects used the index finger of their dominant hand for the reaching movement. Subjects received $12 per hour plus a performance-related bonus.

Apparatus and stimuli. Stimuli were presented in a dimly lit room on a 17-inch (33.8 × 27 cm) Elo touch screen mounted vertically on a Unistrut frame, controlled using the Psychophysics Toolbox[39,40]. Subjects were seated at a viewing distance of 30 cm. In the reaching task, the starting position, a stabilized key, was 28 cm away in depth and 20.5 cm below the screen center. The touch screen recorded the endpoints.

Procedure and design. The three experiments had common task structures: reaching and choice. The procedure and design of experiment 1 is described below, as are the differences in experiment 2. The methods used for experiment 3 are described elsewhere[21].

Experiment 1: reaching. The reaching task served to reveal each subject's objective visuo-motor error distribution to us and to the subject. Subjects were required to touch a target at the center of the screen (with a horizontal and vertical random jitter within ±1 cm) within 400 ms. The target (**Fig. 2a**) was consisted of three vertical rectangles: one 0.4-cm-wide central rectangle and two 0.2-cm-wide flankers, separated by 0.13-cm-wide gaps. They were 5 cm high, therefore large enough to render subjects' vertical errors inconsequential (only 0–5.2%, median 0.70% of subjects' endpoints fell outside the vertical boundaries). Only the horizontal errors of the endpoints would be of interest to us and to the subjects. Subjects' constant errors (mean deviation from the center) were negligible (0.0065–0.45 [median 0.18] of the s.d.).

Subjects held down the starting key to trigger the next target. The timer started when they released the key. If they reached the screen in the time limit, a dot would appear on the target to echo the endpoint. An additional message indicated hit, miss or time-out.

There were 50 warm-up trials and 300 formal trials. Subjects received financial rewards for hitting the target. At the end of the task, eight bonus trials would be randomly drawn from the 300 reaching trials they had performed. Each bonus trial delivered $1 for hit, zero for miss or incurred a penalty of $2 for time-out.

Experiment 1: choice. The choice task was designed to estimate the visuo-motor error distribution subjects assumed in planning their own reaching movements. Subjects chose between two targets, selecting the one they judged to be easier to hit (**Fig. 2c**). Subjects were instructed to imagine hitting the targets from the same starting position and under the same time limit as they had in the earlier reaching task, but no actual reaching attempts were allowed.

On each trial, one target was a row of three equally spaced, identical rectangles (triple); the other target was a single rectangle (single). Each target was presented at the center of the screen for 0.5 s and they were separated by a duration of 0.5 s. Subjects were prompted to respond, "Which is easier to hit? 1st or 2nd?" by key press.

The heights of the targets were the same as those of the reaching task. The widths of the targets were tailored for each subject based on the s.d. of her horizontal visuo-motor error, σ_0, estimated from the reaching task (0.23–0.38 cm across subjects). There were 12 combinations of Triples (**Fig. 2e**), whose width of rectangles was σ_0, $1.5\sigma_0$ or $2\sigma_0$, and whose gap widths were 0.4, 0.6, 0.9 or 1.35× the width of the corresponding rectangles. The width of the single paired with each triple was adjusted by a 1-up/1-down staircase procedure that terminated after 70 trials. All 12 staircases were interleaved.

The 12 × 70 = 840 formal choice trials were preceded by 20 warm-up trials. Trials were self-initiated by key press. Similar to the reaching task, a monetary incentive was applied to encourage subjects to choose the target associated with a higher probability of hit. Subjects were instructed that, at the end of the experiment, eight targets would be randomly selected from those they had preferred in the formal choice trials. They would attempt to hit these bonus targets and be rewarded for hits just as in the reaching task.

Experiment 2. Three settings were different from experiment 1. First, the target in the reaching task was an 8-cm-long line (**Fig. 5a**) that delivered graded instead of all-or-none rewards. For any trial completed within the time limit and the span of the line, subjects received a reward between 100 and 0 points (5,000 points = $1), decreasing as a logistical function of the endpoint-to-line distance. Second, in the choice task subjects chose between two rectangular regions (**Fig. 5b**), one on the target line (central), the other off the target line (side). Subjects were instructed to choose the region that was more likely to catch their endpoints in the previous reaching task. Two trials would be selected at random as bonus trials at the end of the experiment, for each of which subjects could win US$5 if they were correct. Third, staircase procedures were not used. Denote σ_0 as the s.d. of the subject's vertical visuo-motor error (0.45–0.71 cm across subjects). The central had 5 possible heights, $0.2\sigma_0$, $0.4\sigma_0$, $0.6\sigma_0$, $0.8\sigma_0$ σ_0. The side had six possible heights, $0.4\sigma_0$, $0.8\sigma_0$, $1.2\sigma_0$, $1.6\sigma_0$, $2\sigma_0$ and $2.4\sigma_0$, and must be greater in height than its paired central. The side could be above or below the target line by ten possible distances, $0.25\sigma_0$, $0.5\sigma_0$, $0.75\sigma_0$, …, $2.5\sigma_0$. A full combination of these conditions times two repetitions resulted in 960 trials, presented in random order.

Exclusion of trials or subjects. *Reaching.* Time-out trials, 2.7–15% (median 4.3%) for experiment 1, 2.6–27% (median 9.2%) for experiment 2, 1.3–21% (median 9.0%) for experiment 3, and outlier trials beyond $8\sigma_0$ were excluded. In experiment 2, endpoints outside the ends of the target line (no more than one trial for each subject) were also excluded; time-out trials and outside trials were replaced during the experiment.

Choice. In experiment 1, one subject was excluded due to violation of dominance: the subject consistently preferred the single even when the single could be contained in the triple. In experiment 2, two subjects were excluded. One of them chose the central in 99% trials; the other chose the upper region in 97% trials.

Discrete mixture representations. A discrete mixture representation is specified by first designating a partition of part of the real line $a_0 < a_1 < a_2 < \ldots < a_n$. Each interval of the partition $[a_{i-1}, a_i]$ is paired with a basis distribution $f_i(x)$ as follows. Let $c_i = (a_i + a_{i-1})/2$ be the location of the interval and $s_i = (a_i - a_{i-1})/2$ its scale. Then for some choice of seed distribution $f(x)$ that is nonzero outside the interval $[-1, 1]$ define

$$f_i(x) = \frac{1}{s_i} f\left(\frac{x - c_i}{s_i}\right), \quad i = 1,\ldots,n \tag{2}$$

The basis distributions $f_i(x)$ are part of a location-scale family and they are orthogonal.

Let $w_1, \ldots, w_n$ be non-negative weights with

$$\sum_{i=1}^{n} w_i = 1$$

Then, for any choice of $w_1, \ldots, w_n$, the mixture distribution

$$\varphi(x) = \sum_{i=1}^{n} w_i f_i(x) \tag{3}$$

is a discrete mixture representation based on the seed distribution $f(x)$ and the partition $a_0 < a_1 < a_2 < \ldots < a_n$.

U-mix is an example of a discrete mixture representation based on a uniform seed distribution. For convenience, we treat mixtures of Gaussians as a discrete mixture representation with the assumption that almost all of the probability density is confined to one interval of the partition: we ignore the overlap.

Data fitting and model comparison. All the data fitting procedures were conducted on the individual level using maximum likelihood estimates. We used *fminsearchbnd* (J. D'Errico), a function based on *fminsearch* in MATLAB (MathWorks) to search for the parameters that minimized minus

log likelihood. To verify that we had found the global minimum, we repeated the search process using different starting points.

Equivalent width (radius). In experiment 1 (experiment 3), for each specific triple (rectangle), subjects' probability of choosing the single (circle) was modeled as a two-parameter (location and slope) sigmoidal psychometric function of the logarithm of the width of the single (the radius of the circle). The psychometric functions of the 12 triples (10 rectangles) were assumed to have different locations but the same slope. The equivalent width (radius) corresponded to the point on the psychometric function where the probability of choosing the single (circle) was 0.5.

Linking subjects' internal pdfs to their choices. We assume a specific subject's choice between the two targets (T_1 and T_2) on a specific trial is generated as follows. First, probabilities of hit are computed for the two targets based on the subject's internal pdf, which, in the one-dimensional case, is

$$p_1 = \int_{T_1} f(x)dx$$
$$p_2 = \int_{T_2} f(x)dx \tag{4}$$

where $f(x)$ denotes the internal pdf. Subjects would receive a fixed positive value (whose utility is denoted v) for hit and 0 for miss. The expected utilities for the two targets are thus p_1v and p_2v.

The subject's choice, T_1 or T_2, is generated as a Bernoulli random variable, with the probability of choosing T_2 determined by p_1 and p_2 following the normalized expected utility model[41] in the form of a softmax function

$$\Pr(T_2) = \frac{1}{1 + e^{(p_1v - p_2v)/(\tau D)}} \tag{5}$$

where $\tau > 0$ is a temperature parameter determining the randomness of the choice, and $D = p_1(1 - p_2)v + p_2(1 - p_1)v$ is a normalization term. Note that v cancels out in the equation. For each distribution model (for example, Gaussian, U-mix), we fit $\Pr(T_2)$ to subjects' choices to estimate its free parameters and τ.

Visualization of subjects' internal pdf. For each subject, we visualized the internal pdf implicit in the subject's choices using a Bayesian inference procedure as follows. First, we generated 1,000,000 pdfs by sampling from a Gaussian-process prior[42] in log space and normalizing each sample (to guarantee the area under any pdf equals one). The length scale of the Gaussian process was $0.3\sigma_0$. In experiment 1, we required the pdfs to be symmetric around 0 and spanned the stimuli range $[-5.7\sigma_0, 5.7\sigma_0]$. We arbitrarily set the densities within the central range of $[-0.6\sigma_0, 0.6\sigma_0]$ to be constant because the central width of the smallest triple was σ_0, leaving the central densities underdetermined. In experiment 2, asymmetry was allowed and the pdfs spanned the stimuli range $[-4.9\sigma_0, 4.9\sigma_0]$. Second, for each sample pdf, we computed its likelihood of generating the subject's choices. The temperature parameter τ in equation (5) was chosen to be the same as the subject's τ fitted in the Gaussian model. Third, with each sample pdf's likelihood serving as its weight in importance sampling, we obtained the posterior distribution of the subject's internal pdf and accordingly its mean ±s.e.m. (that is, 68% confidence interval).

***t* model**. The probability density function is in the form of a scaled Student's t distribution

$$f(x \mid \kappa, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\kappa\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(\nu + \frac{x^2}{\kappa^2}\right)^{-(\nu+1)/2} \tag{6}$$

where $\kappa > 0$ is a scale parameter and $\nu \geq 4$ is a shape parameter (x/κ has a standard Student's t distribution of ν degrees of freedom), $\Gamma(.)$ is the gamma function. The Gaussian distribution is a limiting case of the scaled Student's t distribution with $\nu \to \infty$.

Linear-decay model. The probability density function is in the form of a triangular distribution

$$f(x \mid \xi) = \begin{cases} \frac{\xi + x}{\xi^2} & -\xi \leq x \leq 0 \\ \frac{\xi - x}{\xi^2} & 0 < x \leq \xi \\ 0 & x < -\xi \text{ or } x > \xi \end{cases} \tag{7}$$

where $\xi > 0$ is the free parameter, denoting the boundary of the regions with nonzero probability density.

Mixture models. Denote w_i as the weight for the i-th component (or pair of components) of an n-component mixture model, which satisfies $0 \leq w_i \leq 1$ and

$$\sum_{i=1}^{n} w_i = 1$$

The Gaussian model is a special case of mixture models vG-Mix or mG-Mix with $n = 1$.

vG-mix is a linear combination of Gaussian distributions with the same mean, but different variances

$$f(x \mid \mu, \sigma_i, w_i) = \sum_{i=1}^{n} w_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}}, \quad i = 1,2,\ldots,n \tag{8}$$

where σ_i and w_i are free parameters, $\mu = 0$ in experiment 1 and is a free parameter in experiment 2.

mG-mix is a linear combination of Gaussian distributions with the same variance but different means. In experiment 1, the Gaussian distributions are in pairs symmetric around 0

$$f(x \mid \mu_i, \sigma, w_i) = \sum_{i=1}^{n} w_i \left(\frac{1}{2\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_i)^2}{2\sigma^2}} + \frac{1}{2\sqrt{2\pi}\sigma} e^{-\frac{(x+\mu_i)^2}{2\sigma^2}} \right), \quad i = 1,2,\ldots,n \tag{9}$$

where $\mu_1 = 0$, μ_i ($i > 1$), σ and w_i are free parameters. In experiment 2, symmetry is not assumed

$$f(x \mid \mu_i, \sigma, w_i) = \sum_{i=1}^{n} w_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_i)^2}{2\sigma^2}}, \quad i = 1,2,\ldots,n \tag{10}$$

where μ_i, σ and w_i are free parameters.

L-mix is a distribution whose pdf is a piecewise linear function. In experiment 1, the distribution is symmetric around 0

$$f(x \mid b_i, h_i) = \begin{cases} \frac{(h_1 - h_0)|x| + h_0b_1 - h_1b_0}{b_1 - b_0} & \text{if } b_0 \leq |x| < b_1 \\ \frac{(h_2 - h_1)|x| + h_1b_2 - h_2b_1}{b_2 - b_1} & \text{if } b_1 \leq |x| < b_2 \\ \ldots & \\ \frac{(h_n - h_{n-1})|x| + h_{n-1}b_n - h_nb_{n-1}}{b_n - b_{n-1}} & \text{if } b_{n-1} \leq |x| < b_n \\ 0, & \text{if } |x| \geq b_n \end{cases} \quad i = 0,1,2,\ldots,n \tag{11}$$

doi:10.1038/nn.4055

where $b_i > 0$, $h_i > 0$ are free parameters except that $b_0 = 0$, $h_n = 0$. In experiment 2, symmetry is not assumed

$$f(x|b_i,h_i) = \begin{cases} \dfrac{(h_1 - h_0)x + h_0 b_1 - h_1 b_0}{b_1 - b_0} & \text{if } b_0 \le x < b_1 \\ \dfrac{(h_2 - h_1)x + h_1 b_2 - h_2 b_1}{b_2 - b_1} & \text{if } b_1 \le x < b_2 \\ \ldots & i = 0,1,2,\ldots,n \\ \dfrac{(h_n - h_{n-1})x + h_{n-1} b_n - h_n b_{n-1}}{b_n - b_{n-1}} & \text{if } b_{n-1} \le x < b_n \\ 0 & \text{if } x \ge b_n \text{ or } x < b_0 \end{cases} \tag{12}$$

where b_i, $h_i > 0$ are free parameters except that $h_0 = 0$, $h_n = 0$. b_i and h_i satisfy $\int f(x|b_i, h_i)dx = 1$.

U-mix is a linear combination of uniform distributions that are adjacent to each other. Let $u(a, b)$ denote the pdf of the uniform distribution on the range $[a, b]$. In experiment 1, the uniform distributions are in pairs symmetric around 0

$$f(x\,|\,\theta_i,\theta_{n+1},w_i) = \sum_{i=1}^{n} w_i \left(\frac{1}{2}u(\theta_i,\theta_{i+1}) + \frac{1}{2}u(-\theta_{i+1},-\theta_i) \right), \quad i = 1,2,\ldots,n \tag{13}$$

where $\theta_1 = 0, 0 \le \theta_2 \le \theta_3 \le \ldots \le \theta_{n+1}$ and w_i are free parameters. In experiment 2, symmetry is not assumed

$$f(x\,|\,\theta_i,\theta_{n+1},w_i) = \sum_{i=1}^{n} w_i u\left(\theta_i,\theta_{i+1}\right), \quad i = 1,2,\ldots,n \tag{14}$$

where $\theta_1 \le \theta_2 \le \ldots \le \theta_{n+1}$ and w_i are free parameters.

Bivariate Gaussian model. The probability density function of the bivariate Gaussian model has the form

$$\psi(x,y\,|\,\sigma_x,\sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)} \tag{15}$$

where σ_x and σ_y are free parameters. The probability of hitting a specific target is computed as the integral of $\Psi(.)$ over the region of the target.

One-dimensional–times–one-dimensional assumption. We considered the possibility that subjects might model their two-dimensional errors as two independent one-dimensional distributions and that they then computed the probability of hitting a two-dimensional region as the product of probabilities of hitting the two one-dimensional ranges. This one-dimensional–times–one-dimensional assumption, when applied to the Gaussian model, led to a much higher goodness-of-fit to subjects' choices in experiment 3 than the bivariate Gaussian model—the median AICc difference across subjects was 22. In the model comparison of experiment 3, the Gaussian and all mixture models were based on the one-dimensional–times–one-dimensional assumption.

Two-dimensional mixture models. A two-dimensional mixture model (vG-mix, mG-mix, L-mix or U-mix) in experiment 3 consists of two one-dimensional mixtures separately for the horizontal and vertical directions, $f(x)$ and $g(y)$, modeled in the same way as the corresponding models in experiment 1. The number of components, n, is counted as the mean of components in the two directions, $(n_x + n_y)/2$. Only models with $|n_x - n_y| \le 1$ were considered. The two-dimensional Gaussian model is a special case of two-dimensional mixture models vG-Mix or mG-Mix with $n = 1$.

The probability of hitting a rectangular target of width $2a$ and height $2b$ is computed as the product of two one-dimensional probabilities

$$\Pr(T_{rect}) = \left(\int_{-a}^{a} f(x)dx\right)\left(\int_{-b}^{b} g(y)dy\right) \tag{16}$$

The probability of hitting a circular target of radius R is also computed as the product of two one-dimensional probabilities

$$\Pr(T_{circle}) = \left(\int_{-\eta R}^{\eta R} f(x)dx\right)\left(\int_{-\eta R}^{\eta R} g(y)dy\right) \tag{17}$$

where

$$\frac{1}{\sqrt{2}} \le \eta \le 1$$

is a discounting parameter for the radius so that the probability of hitting the circle is no less than the probability of hitting its inscribed square and no greater than the probability of hitting its circumscribed square.

Overlap of basis distributions in probability density. For an mG-mix pdf with n Gaussian components ($n > 1$), we divided the spatial axis into n intervals using the middle points of the centers of adjacent components. That is, each interval was owned by one and only by one Gaussian basis distribution. The probability mass in each interval also came from the tails of other Gaussians. We computed the percentage of probability mass contributed by the owner Gaussian for each interval and averaged the percentage across intervals. 100 minus the resulting percentage was defined as the percentage of overlap.

U-mix simulation for ref. 31. The subject's task was to use her finger to control the horizontal position of a virtual pea-shooter to hit a vertical target line. The position of each pea x was drawn from a skewed distribution whose mean was determined by their aim point m and whose skewness was determined by one single parameter ρ

$$\begin{aligned} f(x\,|\,m,\rho) = {} & (1-\rho)N\left(m - 0.2, 0.2\sqrt{2}\right) \\ & + \rho N(m - 0.2 + 0.2/\rho, 0.2(1 + (1/\rho - 1)^2)^{0.5}) \end{aligned} \tag{18}$$

where $N(\mu, \sigma)$ is a Gaussian centered at μ with the s.d. σ. If a subject assumed a loss function where the cost increases quadratically with error, she should align the mean of her error distribution with the target. The previous study found the deviation of subjects' aim point from the mean varied systematically with ρ (**Fig. 7b**). They concluded that subjects' loss functions were quadratic for small errors, but less than quadratic for large errors.

For **Figure 7**, we simulated a virtual subject who had a quadratic loss function, and thus would aim for the mean, but who approximated her error distribution with a U-mix of three nonzero probability categories. For a specific ρ, the categories centered at the median of the error distribution and evenly partitioned the range of $[-3.3\sigma_\rho, 3.3\sigma_\rho]$, where σ_ρ is the s.d. of the error distribution. The probability density of each category was computed as the mean density of the error distribution within the range with the area under the u-mix scaled to be one.

For **Supplementary Figure 7**, we simulated a second virtual subject who was similar to the virtual subject above but who approximated her error distribution with an mG-mix with three components. The s.d. of each component was 0.28. For each specific ρ, the three components were centered in the same way as the U-mix components above. Their weights were estimated as free parameters to minimize the sum of squared differences between the error distribution and the mG-mix approximation.

In the experiment, the error distribution was manipulated to be right-skewed in half of the trials (noninverted trials) and was inverted in the other half. **Figure 2b** of ref. 31 shows subjects' aim point as a function of ρ for the noninverted trials and all trials. The two curves were similar, both having the toward-median and away-from-median biases. But we noticed that the all-trials curve significantly deviated from the mean when the error distribution was symmetric ($\rho = 0.5$),

an anomaly that could not be explained by any loss functions or probability distortions without introducing an asymmetry between left and right, and which was probably due to unknown visuo-motor biases subjects had with respect to pointing to left or right. For this reason, we used the curve for noninverted trials as the data to account for in our simulation.

Summary of statistical tests. No statistical methods were used to predetermine sample sizes but our choice of sample size was based on previous work, including ref. 21. AICc[18,19] and group-level Bayesian model selection[20] were used in all experiments. Pearson's correlation was used in experiment 1. We verified the assumptions of all of the statistical tests used.

A **Supplementary Methods Checklist** is available.

39. Pelli, D.G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
40. Brainard, D.H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
41. Erev, I. *et al.* A choice prediction competition: choices from experience and from description. *J. Behav. Decis. Mak.* **23**, 15–47 (2010).
42. Rasmussen, C.E. & Williams, C.K.I. *Gaussian Processes for Machine Learning* (MIT Press, 2006).

doi:10.1038/nn.4055