

LETTERS

A new perceptual illusion reveals mechanisms of sensory decoding

Mehrdad Jazayeri¹ & J. Anthony Movshon¹

Perceptual illusions are usually thought to arise from the way sensory signals are encoded by the brain, and indeed are often used to infer the mechanisms of sensory encoding¹. But perceptual illusions might also result from the way the brain decodes sensory information², reflecting the strategies that optimize performance in particular tasks. In a fine discrimination task, the most accurate information comes from neurons tuned away from the discrimination boundary^{3,4}, and observers seem to use signals from these ‘displaced’ neurons to optimize their performance^{5,6,7}. We wondered whether using signals from these neurons might also bias perception. In a fine direction discrimination task using moving random-dot stimuli, we found that observers’ perception of the direction of motion is indeed biased away from the boundary. This misperception can be accurately described by a decoding model that preferentially weights signals from neurons whose responses best discriminate those directions. In a coarse discrimination task, to which a different decoding rule applies⁴, the same stimulus is not misperceived, suggesting that the illusion is a direct consequence of the decoding strategy that observers use to make fine perceptual judgments. The subjective experience of motion is therefore not mediated directly by the responses of sensory neurons, but is only developed after the responses of these neurons are decoded.

Subjects viewed a field of moving dots within a circular aperture around a fixation point for 1 s and reported whether the direction of motion was clockwise (CW) or counterclockwise (CCW) of a decision boundary indicated by a bar outside the edge of the dot-field (Fig. 1a). On each trial, the boundary took a random position around the dot field, and a percentage of dots (3%, 6% or 12%) moved coherently in a randomly chosen direction within 22 degrees

of the boundary; the other dots moved randomly. After each trial, subjects pressed one of two keys to indicate their choice (CW or CCW). For 70% of trials, they were given feedback. On the remaining

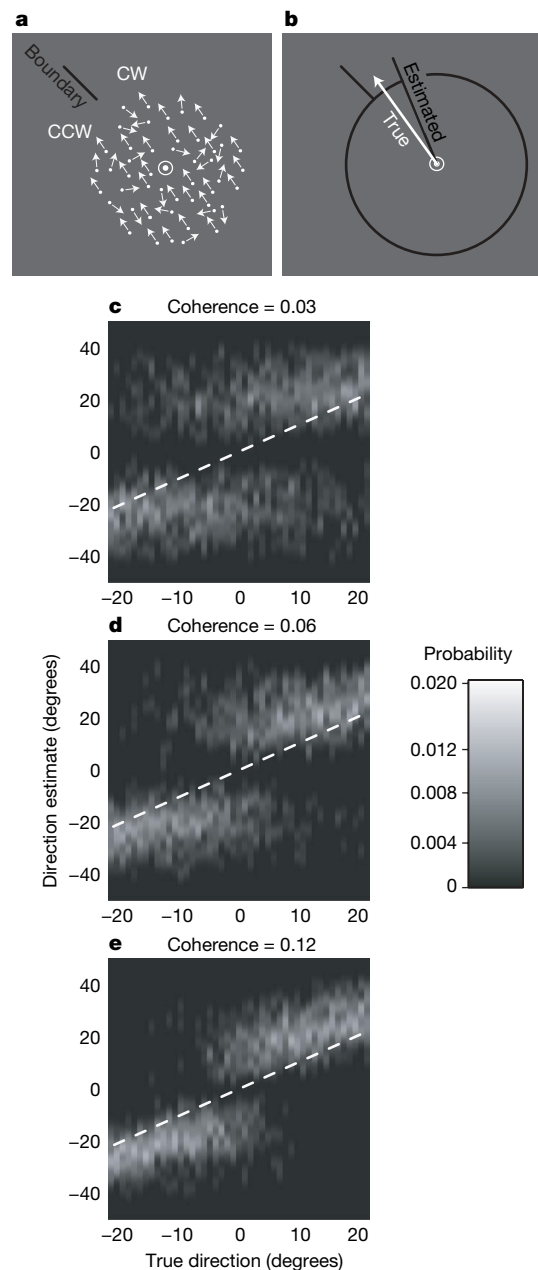


Figure 1 | The combined discrimination-estimation experiment. **a**, The discrimination phase. Subjects viewed a field of moving random dots and indicated whether its direction was clockwise (CW) or counter-clockwise (CCW) with respect to an indicated discrimination boundary that varied randomly from trial to trial. **b**, The estimation phase. On an unpredictable 30% of trials, after discriminating the direction of motion, subjects reported their estimate of the direction of motion by extending a dark line from the centre of the display with the computer mouse. **c–e**, Image maps representing the distribution of estimation responses for one subject at the three coherence levels. Each column of each plot represents the distribution of estimates for a particular true direction of motion, using a nonlinear lightness scale for probability (right). The observed values have been smoothed parallel to the ordinate with a gaussian (s.d. = 2 degrees) for clarity. The black dashed line is the locus of veridical estimates. Responses in the top-left and bottom-right quadrants of each map correspond to error trials, whereas those in the top-right and bottom-left quadrants were for correct trials; the discrimination and estimation responses were concordant throughout. Judgement errors (top-left and bottom-right quadrants) decrease with increasing coherence and as direction becomes more different from the discrimination boundary.

¹Center for Neural Science, New York University, 4 Washington Place, New York, New York 10003, USA.

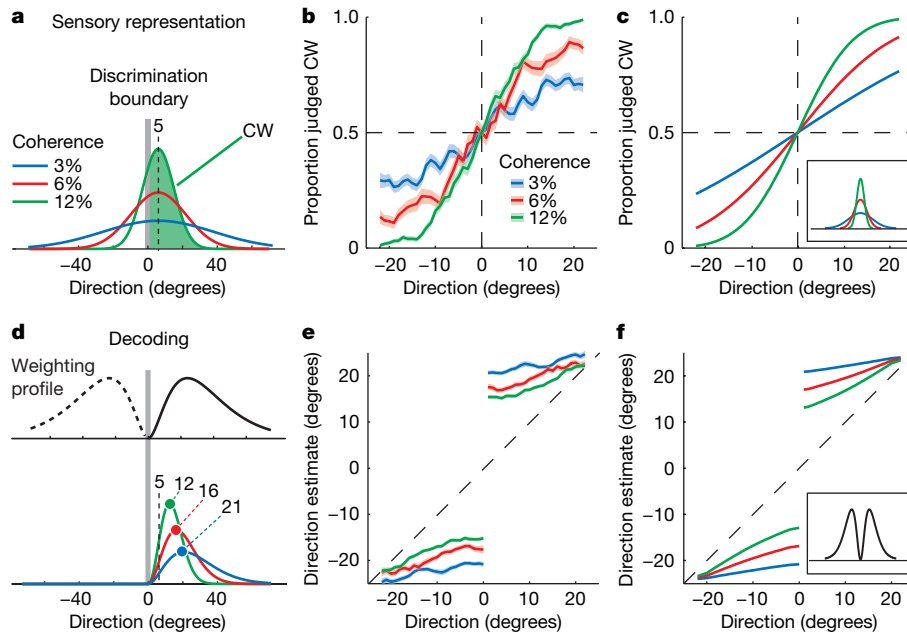


Figure 2 | Discrimination and estimation responses. **a**, The sensory representations evoked by a dot field moving 5° CW vary from one trial to the next. The plot is a cartoon of the distribution of these noise-perturbed sensory representations for different coherences. The distribution becomes more variable with weaker signals. The proportion of CW judgements is the area under the CW part of this distribution, shown for 12% coherence by the shaded green area. **b**, Proportion of CW judgements (thick lines) and their standard errors (shading) as a function of direction of motion for all coherence values for one subject. The CW and CCW portions of the data from Fig. 1b have been pooled and smoothed with a 3 degree boxcar filter. **c**, Fits to the discrimination data in **b**, using the model drawn in **a**; the inset shows the inferred sensory representations. **d**, The decoding model. The

sensory representations from **a** are multiplied by a displaced weighting profile that is optimal for discriminating CW from CCW alternatives. As a result, the peaks of the distributions shift away from the boundary. The plot shows schematically how this model predicts larger shifts for lower coherence values—the peaks for coherences of 3%, 6% and 12% fall at 12, 16 and 21 degrees respectively, even though the peak of the underlying sensory representation (from **a**) remains at 5 degrees. **e**, Subjective estimates as a function of direction of motion for trials on which motion direction was correctly discriminated. The CW and CCW portions of the data (Fig. 1d) have been pooled and smoothed with a 3 degree boxcar filter. **f**, The model fits for the subjective estimates after estimating and applying the single weighting profile that best matches the data.

30%, feedback was withheld and subjects estimated the direction of motion they had seen by aligning a bar extending from the fixation point to the direction of their estimate (Fig. 1b).

For all subjects, discrimination performance was lawfully related to motion coherence and direction: performance improved for higher coherences and for directions of motion farther from

the boundary, and there was no systematic bias in the choice behaviour (Fig. 1c–e). However, when subjects were asked to report the direction of motion, their estimates deviated from the direction of motion in the stimulus, and were biased in register with their discrimination choice (Fig. 1c–e). The magnitude of these deviations depended on both the coherence and the direction of motion, being

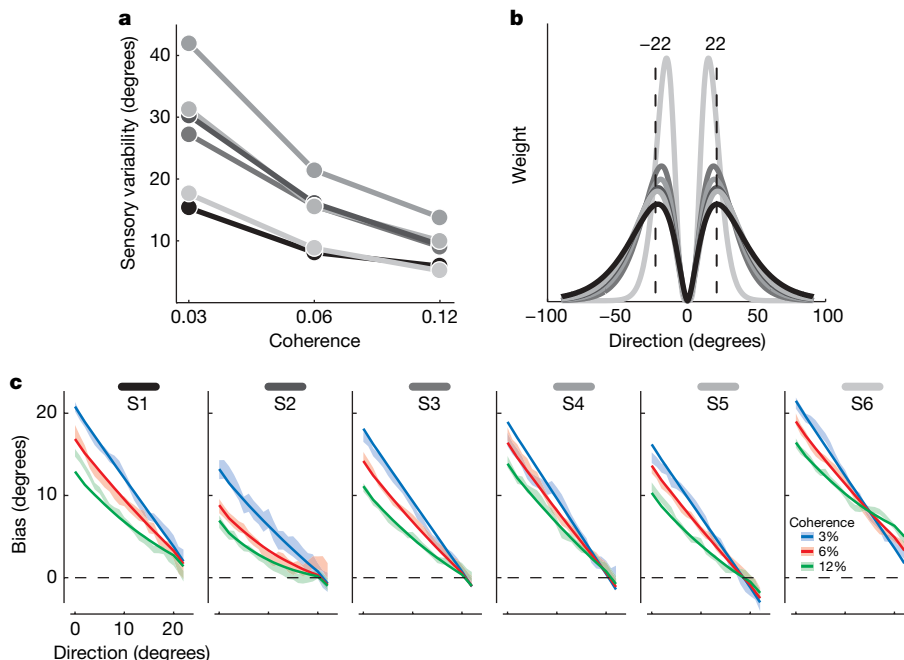


Figure 3 | Summary data for all six subjects. **a**, The variability of the sensory representations as a function of motion coherence for all subjects (computed as the standard deviation of the gaussian fits, for example, Fig. 2c, inset) are shown with different shades of grey (S1 to S6 in **c**). **b**, Recovered weighting functions for all six subjects. The dotted lines delimit the range of directions of motion (–22 to 22 degrees) that were used in the experiment. **c**, The mean bias (the difference between the true and estimated directions) ± one standard error (shading) and the model fits (thick line) for all subjects and all coherence values.

larger for more uncertain conditions when either coherence was low or the direction was close to the boundary—the conditions in which discrimination performance was worst.

The sensory representation evoked by the random dot stimulus is perturbed by noise^{8,9,10}, and we therefore expect it to be more variable from trial to trial for weak motion signals than for stronger ones (Fig. 2a). How well observers discriminate the alternatives depends on the strength of the motion signal and its direction with respect to the boundary. Assuming that the variability in the sensory representation can be described by a gaussian, we fitted the discrimination performance of each subject (Fig. 2b) with a cumulative gaussian to estimate the spread of the sensory representation for each level of coherence. As expected, the variance of this distribution decreased with increasing coherence (Fig. 2c, inset).

This formulation accounts simply and well for discrimination behaviour (Fig. 2c), but it does not explain why the subjective estimates deviate from the true direction of motion in the stimulus. To understand what causes the perceptual biases, we considered the events that lead to the subjective estimates of the direction of motion. On each trial, before reporting their estimate, observers make a fine perceptual judgement. To do so, they have to transform the sensory responses into a binary decision (CW or CCW). Because subjects did not receive feedback on their subjective estimates, they could only adjust their decoding strategy for the discrimination part of the combined discrimination–estimation task where they did receive feedback. As shown both in theory^{3,4} and experiment^{6,7}, in a fine discrimination task like ours, neurons with direction preferences moderately shifted to the sides of the boundary make the largest contribution, whereas neurons tuned to directions either near or very remote from the boundary are less important. Therefore, to decode the activity of sensory neurons efficiently, the brain must pool their responses with a weighting profile that has maxima moderately shifted to the sides of the boundary⁴ (Fig. 2d, top panel).

If the pattern of direction estimates is explained by such a displaced profile, there should be a weighting function which, when applied to the sensory representation of different stimuli, predicts the corresponding estimates. We computed the product of this weighting profile (Fig. 2d, top panel) with the sensory representation of the stimulus estimated from the discrimination performance (Fig. 2a), and took the peak as the direction estimate (Fig. 2d, bottom panel). For each observer, we fitted the weighting profile that, when combined with that observer's discrimination performance, best predicted the pattern of direction estimates. Remarkably, combining the sensory representation with a single weighting profile (Fig. 2f, inset) accurately captured the observed estimates for all coherence levels and all directions of motion (Fig. 2f). Though observers varied in the accuracy of their sensory representations (Fig. 3a), the inferred weighting functions were similar for all six (Fig. 3b), and the resulting model accurately predicted the estimation bias (the difference between the true and estimated directions) for all six (Fig. 3c).

The misperception of motion can be economically attributed to the decoding strategy that observers adopt to optimize fine perceptual judgements, but other interpretations are possible. For example, the misperception might reflect a change in the sensory representation evoked by the stimulus, and not the way it is decoded. To test this idea, we ran a second experiment that differed from the first only in that the fine discrimination was replaced by coarse discrimination. On every trial, we presented motion in a randomly chosen direction within 22 degrees of a bar presented in the periphery (previously used for discrimination boundary), or within 22 degrees of the direction opposite the bar. Subjects discriminated whether the direction of motion was towards or away from the bar and as before, on a subset of trials reported their estimate of the direction of motion. As shown in theory⁴ and experiment¹¹, the most accurate information now comes from neurons tuned to the two alternatives. Therefore, the bias should, if anything, change from repulsion to attraction. This is exactly the pattern of responses we observed (Fig. 4a–c). The illusion thus

depends entirely on the subject's task—it occurs during fine discrimination but not during coarse discrimination (see Supplementary Information for a more detailed discussion). Changes in the sensory representation therefore cannot explain the effect.

One other possibility is that observers did not 'truly' misperceive the motion, but when uncertain about its direction, adopted a biased response strategy to ensure that they would not disagree with their immediately preceding discrimination choice. Simple models of response bias that only depend on the preceding choices can easily be discarded because they cannot account for the systematic relationship between the subjective estimates and the strength and direction of motion (Figs 1c–e, 2e). As detailed in the Supplementary Discussion, biased response strategies that are rich enough to account for our data have to incorporate computations effectively the same as those that our decoding model employs, applying a weighting profile to the sensory representation. Should we then view these computations in the framework of sensory decoding or complex response bias? In our decoding model, all computations serve a well-grounded function that can be inferred from theoretical and experimental observations of fine discrimination^{3,4,6,7}. This model also has the

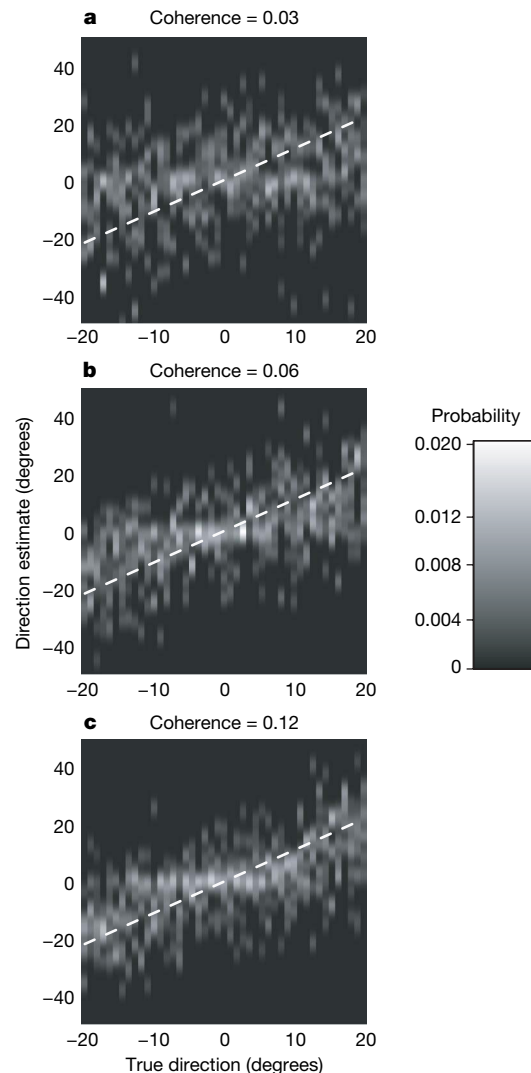


Figure 4 | Subjective estimates in a coarse direction discrimination task. a–c, The estimates are represented as in Fig. 1c–e. In the discrimination phase, the subject's task was to indicate whether the direction of motion in the random-dot was towards or away from a peripheral visual cue (the same as the discrimination boundary in Fig. 1b). The distributions for correct discriminations towards and away from the visual cue are pooled (separate plots for the two conditions are shown in Supplementary Fig. 3).

virtue of simplicity, because it accounts for observers' subjective reports using only the machinery that accounts for their objective discrimination performance. Response bias models, on the other hand, postulate two unrelated mechanisms, one to account for discrimination and the other for perceptual reports.

Bias arising from a decoding strategy may also explain some other perceptual distortions, such as those related to repulsion away from the cardinal axes¹² or from other discrimination boundaries¹³. We believe that this 'reference repulsion' phenomenon arises when subjects implicitly discriminate stimulus features against available internal or external references, such as a cardinal direction or the boundary marker in our experiment. This causes their perception of those features to be shifted away from the reference by the mechanism we have described. In other words, these incidences of misperception reveal the optimality of the system—not in perceiving, but in decoding sensory signals to make fine perceptual judgments.

Because the misperception does not seem to reflect the sensory responses to the direction of motion, the subjective experience of motion must be mediated by the machinery that decodes the responses of motion-sensitive neurons. We have argued elsewhere that areas downstream of sensory representations recode sensory responses into sensory likelihoods⁴, and the discrimination model used here is derived directly from that representation. Our results therefore suggest that the subjective experience of sensory events arises from the representation of sensory likelihoods, and not directly from the responses of sensory neuron populations.

METHODS

Eight subjects aged 19 to 35 participated in this study after giving informed consent. All had normal or corrected-to-normal vision, and all except one were naive to the purpose of the experiment. Subjects viewed all stimuli binocularly from a distance of 71 cm on an Eizo T960 monitor driven by a Macintosh G5 computer at a refresh rate of 120 Hz in a dark, quiet room.

In the main experiment, in which six of the subjects participated, each trial began with the presentation of a fixation point along with a dark bar in the periphery representing the discrimination boundary for the subsequent motion discrimination (Fig. 1a). After 0.5 s, the motion stimulus was presented for 1 s. Subjects were asked to keep fixation during the presentation of the motion stimulus. After the motion stimulus was extinguished, subjects pressed one of two keys to report whether the direction of motion was CW or CCW with respect to the boundary and received distinct auditory feedback for correct and incorrect judgements. On approximately 30% of trials chosen at random, feedback was withheld and a circular ring was presented as a cue for the subject to report the direction of motion in the stimulus (Fig. 1b). The subject reported the estimate by using a mouse to extend a dark bar from the fixation point in the direction of their estimate and terminated the trial by pressing a key. Subjects were asked to estimate accurately but did not receive feedback. The discrimination boundary and the fixation point persisted throughout the trial. Trials were separated with a 1.5 s inter-trial interval during which the screen was blank.

For the main experiment, subjects had ample time to practice and master the task contingencies. Data for the main experiment were collected only after the discrimination thresholds stabilized (changed less than 10% across consecutive sessions). After this period, subjects completed roughly 8,000 trials in 10–12 sessions, each lasting approximately 45 min.

The remaining two subjects participated in the control coarse discrimination experiment. This experiment differed from the main experiment in that motion was within 22 degrees either towards or opposite to the peripherally presented bar (that is, the discrimination boundary in the main experiment), and during the discrimination stage, subjects had to report whether motion was towards or away from the bar. On approximately 30% to 50% of the trials chosen at random, the feedback was withheld and subjects were asked to report the perceived direction of motion (same procedure as in the main experiment).

All stimuli were presented on a dark grey background of 11 cd m^{-2} . The fixation point was a central circular white point subtending 0.5 degrees with a luminance of 77 cd m^{-2} . A gap of 1 degree between the fixation point and the motion stimulus helped subjects maintain fixation. The discrimination boundary was a black bar 0.5 degrees by 0.15 degrees, 3.5 degrees from fixation. The motion stimulus was a field of dots (each 0.12 degrees in diameter with a

luminance of 77 cd m^{-2}) contained within a 5-degree circular aperture centred on the fixation point (Fig. 1). On successive video frames, some dots moved coherently in a designated direction at a speed of 4 deg s^{-1} , and the others were replotted at random locations within the aperture. On each trial, the percentage of coherently moving dots (coherence) was randomly chosen to be 3, 6 or 12%, and their direction was randomly set to a direction within 22 degrees of the discrimination boundary; in the second experiment, half the trials presented motion within 22 degrees of a direction 180 degrees away from the boundary. The dots had an average density of 40 dots $\text{deg}^{-2} \text{ s}^{-1}$. The presentation of a black circular ring with a radius of 3.3 degrees around the fixation cued the subjects to report their estimate, which they did by moving the mouse to extend and align a black bar of width 0.15 deg to the direction of their estimate.

We modelled the sensory representation with a gaussian probability density function centred at the true direction of motion. The variance of this distribution for each subject was estimated by fitting a cumulative gaussian to his/her discrimination performance to maximize the likelihood of observing the subjects' choices for each level of motion coherence. To predict the direction estimates, we multiplied this sensory representation by a weighting function, and took the peak of the result. An additional additive constant accounted for any motor bias independent of sensory evidence. We chose a gamma probability density function as a convenient parametric form for the weighting profile. For each subject, we obtained parametric fits by minimizing the squared error of the model's prediction for the observed mean direction estimates for that subject. The gamma distribution provided a good fit for our data, but our conclusions do not depend on the exact form of the weighting profile. This simple procedure of first finding the sensory likelihoods from discrimination data, and then computing the weighting profile from the estimation data, crystallizes the contrast between encoding and decoding in our model. In detail, however, it neglects the subtle effect of the weighting profile on discrimination behaviour. In Supplementary Methods, we detail a complete model that simultaneously accounts for both the discrimination choices and the direction estimates in a single step (Supplementary Fig. 1), and present the fits for the correct as well as error trials (Supplementary Fig. 2).

Received 7 November 2006; accepted 9 March 2007.

Published online 4 April 2007.

1. Eagleman, D. M. Visual illusions and neurobiology. *Nature Rev. Neurosci.* **2**, 920–926 (2001).
2. Gregory, R. L. *Eye and Brain: The Psychology of Seeing* 5th edn (Oxford Univ. Press, 1997).
3. Seung, H. S. & Sompolinsky, H. Simple models for reading neuronal population codes. *Proc. Natl Acad. Sci. USA* **90**, 10749–10753 (1993).
4. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nature Neurosci.* **9**, 690–696 (2006).
5. Patterson, R. D. Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.* **59**, 640–654 (1976).
6. Regan, D. & Beverley, K. I. Postadaptation orientation discrimination. *J. Opt. Soc. Am.* **2**, 147–155 (1985).
7. Hol, K. & Treue, S. Different populations of neurons contribute to the detection and discrimination of visual motion. *Vision Res.* **41**, 685–689 (2001).
8. Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
9. Dean, A. F. The variability of discharge of simple cells in the cat striate cortex. *Exp. Brain Res.* **44**, 437–440 (1981).
10. Tolhurst, D. J., Movshon, J. A. & Dean, A. F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785 (1983).
11. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
12. Huttenlocher, J., Hedges, L. V. & Duncan, S. Categories and particulars: Prototype effects in estimating spatial location. *Psychol. Rev.* **98**, 352–376 (1991).
13. Rauber, H. & Treue, S. Reference repulsion when judging the direction of visual motion. *Perception* **27**, 393–402 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a research grant from the NIH. We are grateful to B. Lau, E. Simoncelli, D. Heeger, M. Landy and N. Graham for advice and discussion.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J. (mjaz@cns.nyu.edu).