doi: 10.1038/nature05739

A new perceptual illusion reveals mechanisms of sensory decoding

Mehrdad Jazayeri & J. Anthony Movshon

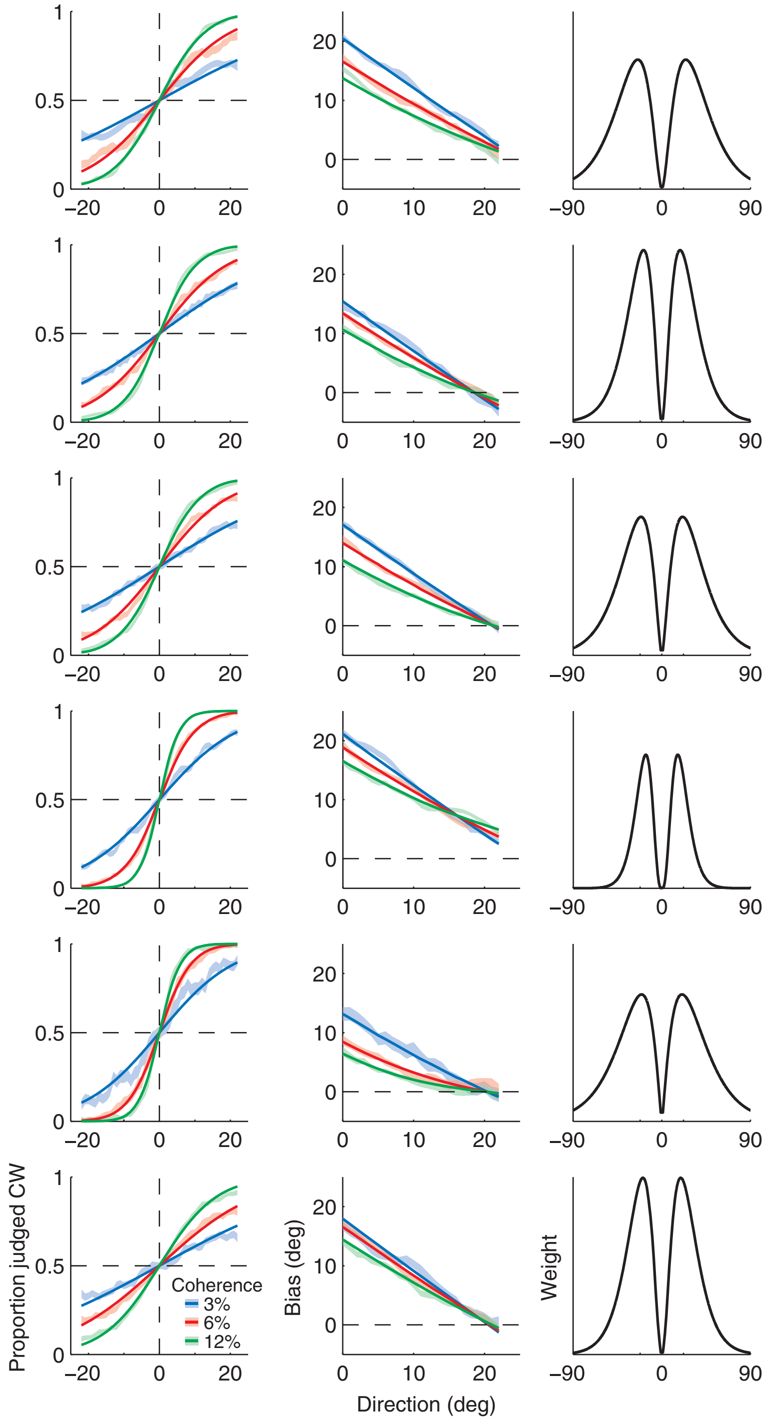
Supplementary Methods

The complete model

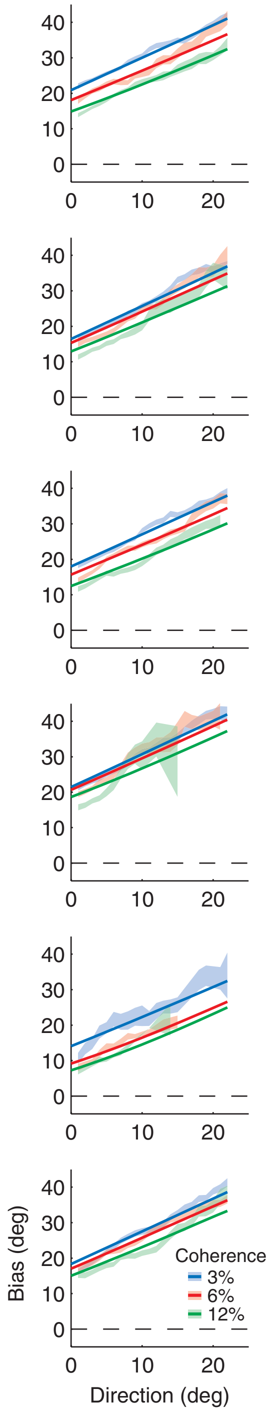
The complete model is composed of an encoding and a decoding stage. At the encoding stage, it assumes the sensory representation has a Gaussian probability density function centred at the true direction of motion. The decoding stage computes sensory likelihoods by multiplying the sensory representation with a weighting profile modelled as a gamma density function. To account for the discrimination data, we assume that a binomial process in which the underlying probabilities are taken from the area of the sensory likelihood in the clockwise (CW) and counterclockwise (CCW) parts of the resulting sensory likelihoods governs subjects' choices. The direction estimates are modelled as the peak of the sensory likelihood plus an additional constant term to account for any motor bias independent of sensory evidence. Therefore, the encoding model has 3 parameters representing the standard deviation of the Gaussian sensory representation for the 3 coherence values tested (i.e. 3%, 6% and 12%), and the decoding model has 4 parameters: 3 for the gamma density function and 1 for the constant bias term.

The discrimination and estimation data were simultaneously fitted to minimize the arithmetic sum of two error terms: the negative binomial likelihood of observing the subjects' choices during the discrimination phase, and the squared error of the model's prediction for the observed mean direction estimates (Supplementary Fig. 1).

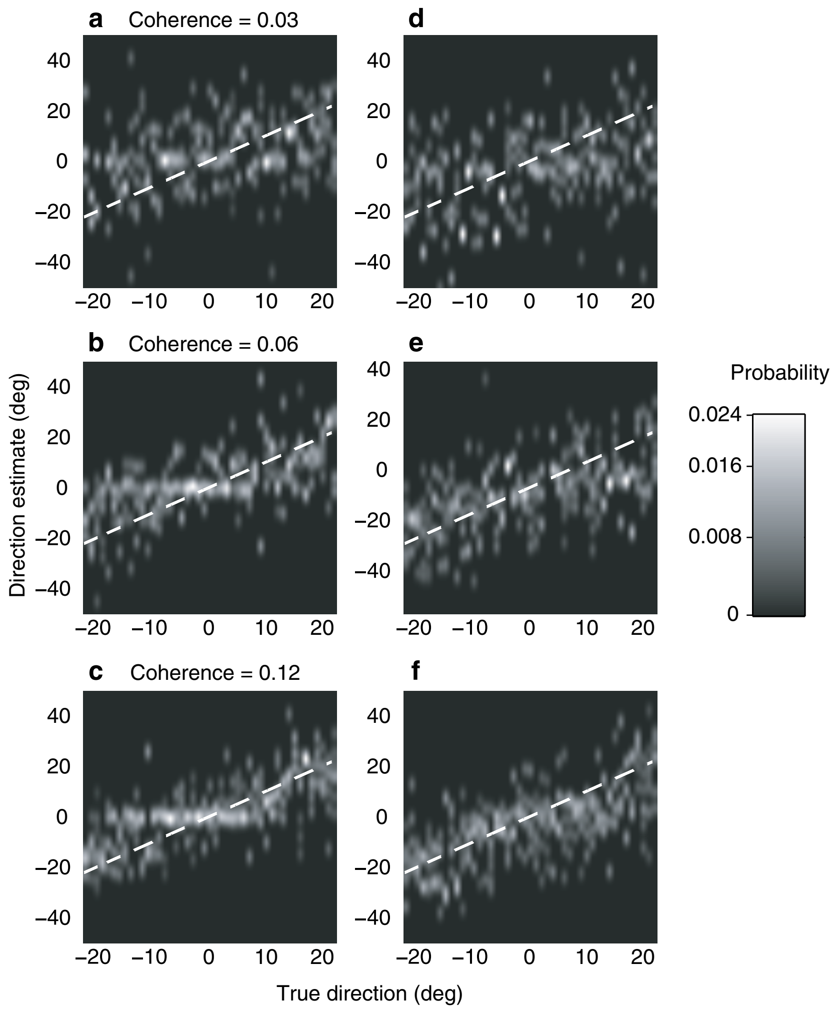
Supplementary Figures



Supplementary Figure 1. Data and corresponding fits from the complete model. Left column: the mean ± one standard error (shading) of the proportion of CW judgments and the corresponding fits of the complete model for all subjects (6 rows for 6 subjects) all coherence values (3% blue, 6% red, and 12% green). Middle column: the mean bias (the difference between the true and estimated directions) ± one standard error (shading) and the corresponding model fits (thick line). The black dashed line is the locus of veridical estimates (i.e. no bias). Right column: recovered weighting functions.



Supplementary Figure 2. Data and corresponding fits for all 6 subjects for the error trials. The mean bias (the difference between the true and estimated directions) ± one standard error (shading) and the corresponding model fits (thick line) are shown for all subjects (6 rows) and all coherence values (3% blue, 6% red, and 12% green). The quality of fits was inversely related to the number of error trials. For subjects whose performance was poor (e.g. rows 1, 3 and 6) and a large number of errors trials for all three coherences were available (as evident by the small shaded standard error), the fits were better than for those who made fewer errors.



Supplementary Figure 3. Direction estimates in the coarse discrimination task. (a-c) Image maps representing the distribution of estimation responses for one subject at the three coherence levels for all correct trials in which the direction of motion was towards the peripherally presented bar. 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 deg) for clarity. The white dashed line is the locus of veridical estimates. (d-f) Same as a-c for trials in which the coherent dots moved away from the peripherally presented bar. In both columns, for the low coherence motion, subjective estimates seem to be biased towards the peripheral bar. This pattern is the opposite of what we found in the fine discrimination task and is expected by our proposed model. In a coarse discrimination, as described in the main text, the most accurate information comes from neurons tuned to the two alternatives and as such, the appropriate weighting profile is one that has maxima at the two alternatives. The logic of our model thus, would predict a bias toward the peripheral bar. There are however, differences in the two columns. In particular, in the left column in which the motion signal was towards the peripheral bar, a proportion of subjective estimates are densely distributed near the direction of the bar as would be expected from a pure response bias. This bias is likely due to a mixed strategy in which, when uncertain of the direction of motion, subjects would simply choose the direction of the bar. The response bias is absent in the right column, showing data obtained when no reference bar was available.

Supplementary Discussion

The response bias interpretation

Our observers' reports indicate that they misperceive the direction of motion. This misperception can be economically explained by the decoding strategy that observers employ in fine direction discriminations: pooling the activity of sensory neurons with a weighting profile that has maxima moderately shifted to the sides of the discrimination boundary. It is, however, possible that observers did not misperceive the direction of motion, but when asked to report their estimate, misreported an otherwise unbiased percept. To examine this alternative hypothesis, we considered different response strategies that might give rise to the observed biases. An account based on a biased response strategy must satisfy three characteristic features of our data:

All direction estimates are consistent with the preceding discrimination choices. Direction estimates CW of the boundary were always preceded by CW discrimination choices and vice versa. This "consistency" criterion can be satisfied easily if observers avoid making direction estimates that would disagree with their preceding CW/CCW choices. For example, after making a CW or CCW choice, observers might simply estimate the direction to be somewhere in a range of directions consistent with their preceding choice. This idea can be rejected because it does not predict the systematic relationship we found between true and estimated directions of motion (Fig. 1c-e, Fig. 2e), so we must consider more elaborate response strategies that take the direction of motion in the stimulus into account.

Direction estimates are more biased for directions of motion close to the boundary. To capture this relationship, one might propose a more elaborate response strategy. Suppose that observers first extract an unbiased estimate of the direction of motion (e.g. the peak of the sensory representation), but when asked about their subjective percept, they fabricate a report that is progressively more biased when the true direction of motion is closer to the boundary. This can be thought of as a form of reference repulsion in which the magnitude of the repulsion falls for directions more distant from the discrimination boundary. An appropriate formulation of such reference repulsion could account for the observed relationship between the true direction of motion and the observers' reports. But this idea can also be rejected because it does not explain why the reports are more biased for weaker motion signals, so we must consider response strategies that, in addition to being consistent and repulsive, also vary with the profile of the sensory representation.

Direction estimates vary systematically with motion coherence. The relationship between direction of motion and subjective report is different under conditions of low and high uncertainty (Fig. 1c-e). Different levels of coherence change the profile of the sensory representation but not its mean (Fig 2a), so the appropriate response strategy must somehow take this profile into account. In other words, to satisfy all three conditions (consistency, repulsiveness and coherence dependence), the fabricated "reference repulsion" described above should operate on the whole profile of sensory representation. From a computational standpoint however, our decoding model does exactly what is needed to model such a complex response strategy: it multiplies the sensory representation by a "repulsive" weighting profile (i.e. with maxima that are shifted away from the boundary) and takes the peak of the result. In other words, the computations that characterize response strategies that are rich enough to account for the observed patterns of biases are isomorphic to those we used in our decoding model.

We are then left with the question of how to interpret the observed pattern of judgments – are they truly a measurement of misperception, or rather a reflection of a complex response bias? If we take the view that perception – in its entirety – is associated with sensory representations, and identify all subsequent computations with the generation of a response, then the observed biases can be taken as the product of some form of response strategy. If on the other hand, we assume that perceptual experience depends on how the brain decodes sensory representations, then the biased reports can be taken as a form of misperception. We favour the decoding model for reasons of parsimony – it works without invoking any additional process beyond the one that supports discrimination performance.