Optimal inference explains the perceptual coherence of visual motion stimuli

James H. Hedges

Howard Hughes Medical Institute, New York University, New York, NY, USA, & Center for Neural Science, New York University, New York, NY, USA

Alan A. Stocker

Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA, & Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA

Eero P. Simoncelli

Howard Hughes Medical Institute, New York University, New York, NY, USA, Center for Neural Science, New York University, New York, NY, USA, & Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

The local spatiotemporal pattern of light on the retina is often consistent with a single translational velocity which may also be interpreted as a superposition of spatial patterns translating with different velocities. Human perception reflects such interpretations, as can be demonstrated using stimuli constructed from a superposition of two drifting gratings. Depending on a variety of parameters, these stimuli may be perceived as a coherently moving plaid pattern or as two transparent gratings moving in different directions. Here, we propose a quantitative model that explains how and why such interpretations are selected. An observer's percept corresponds to the most probable interpretation of noisy measurements of local image motion, based on separate prior beliefs about the speed and singularity of visual motion. This model accounts for human perceptual interpretations across a broad range of angles and speeds. With optimized parameters, its components are consistent with previous results in motion perception.

Introduction

Visual perception has been described as a process by which uncertain and ambiguous measurements are combined with internal knowledge of the world in order to arrive at a consistent interpretation of one's surroundings (Alhazen, 1040/2005; Helmholtz, 1866/2000). For some visual stimuli, such as the well-known Necker Cube (Necker, 1832), the ambiguities cannot be uniquely resolved with internal knowledge or assumptions. Such stimuli may appear to switch back and forth between two bistable states (Figure 1a). The visual system may also exhibit a preference for a particular solution, based on a combination of stimulus cues and internal knowledge. A widely studied example in the motion literature is a drifting "grating" that has a spatial intensity pattern that varies along a single direction and is constant in the orthogonal direction. To a visual observer, the translational velocity of such stimuli is not uniquely constrained, because the retinal image contains no information about the component of velocity parallel to the spatial stripes (Wallach, 1935; Wuerger, Shapley, & Rubin, 1996), an ambiguity widely known as the "aperture problem" (Fennema & Thompson, 1979; Marr & Ullman, 1981). In the absence of additional contextual cues, the visual system seems to assume that the unobservable component has a value of zero and, thus, that the perceived velocity of the stimulus is generally in the direction normal to the spatial stripes.

When two drifting gratings are superimposed (Figure 1b), they either appear to fuse into a coherently moving "plaid" pattern or appear to move independently, sliding *transparently* past each other (Adelson & Movshon, 1982; Wallach, 1935). Clearly, both interpretations are physically plausible scenarios. How does the visual system choose? The answer depends on a variety of different attributes of the stimulus (Adelson & Movshon, 1982; Cropper, Mullen, & Badcock, 1996; Hupé & Rubin, 2003;
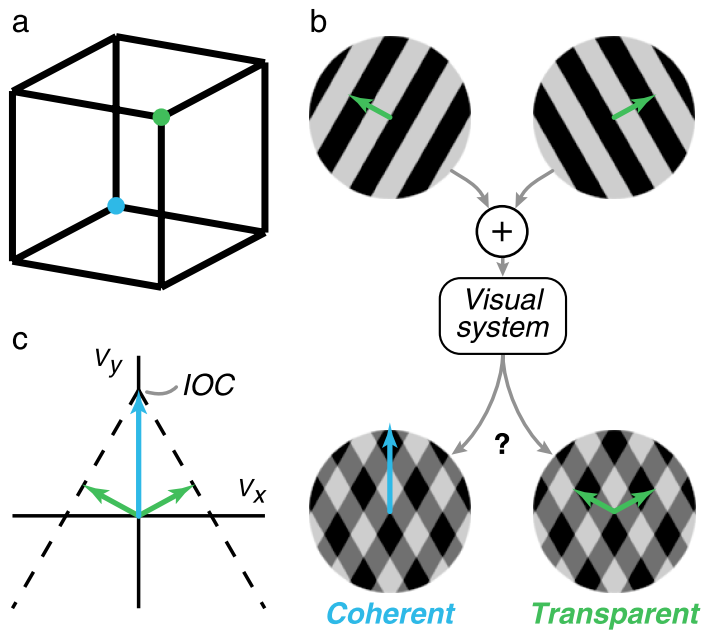
Figure 1. Perceptual ambiguity of visual stimuli. (a) A Necker Cube, which can be interpreted in one of two ways: either the green vertex appears closer, as if the cube is seen from above, or the blue vertex appears closer, as if the cube is seen from below. At any moment in time, human observers seem to perceive only one of these, but this percept switches spontaneously over time. (b) A plaid pattern, formed from the superposition of two drifting square-wave gratings, also admits two interpretations. The plaid can appear to be a singular rigidly moving pattern (blue vector) or two transparent gratings sliding over one another (green vectors). For any given plaid, humans tend to perceive one or the other initially, but this percept is also known to spontaneously switch over time (Hupé & Rubin, 2003; Wallach, 1935). (c) Illustration of the velocity–space constraints on stimulus interpretation. The two oblique vectors (green) represent the normal velocities of the two gratings shown in (b). The motion of each component grating, when viewed in isolation, is consistent with a set of translational velocities that lie along a constraint line (dashed black), and the point at which these lines intersect (blue—the "intersection of constraints," or IOC (Adelson & Movshon, 1982)) is the unique velocity that is consistent with the ambiguous motion of *both* grating components and, thus, consistent with the rigid motion of the combined plaid pattern.

Kim & Wilson, 1993; Kooi, Valois, Switkes, & Grosof, 1992; Krauskopf & Farell, 1990; Krauskopf, Wu, & Farell, 1996; Movshon, Adelson, Gizzi, & Newsome, 1986; Smith, 1992; Stoner & Albright, 1992; Stoner, Albright, & Ramachandran, 1990; Victor & Conte, 1992; Welch & Bowne, 1990). In general, the transparent interpretation becomes more likely with faster component speeds, broader angles, and longer presentation times, as well as with greater differences between the components' attributes, including in speed, spatial frequency, contrast, depth, or hue. The effects of these parameters have generally been examined one at a time, and most proposed models are tied to the specifics of these plaid stimuli and are thus difficult to generalize to natural vision.

Here, we develop a normative model that operates by combining three fundamental components using principles of statistical estimation theory. The first component is a set of orientation- and motion-selective measurements. We assume that these measurements are noisy but provide sufficient information to estimate the normal velocity of a grating at any orientation and speed. The second component is the distribution of visual speeds encountered during normal vision, expressed as a prior probability density (e.g., Simoncelli, 1993; Simoncelli & Heeger, 1992; Weiss, Simoncelli, & Adelson, 2002; Yuille & Grzywacz, 1988). Lastly, the third component is the frequency of occurrence of coherent motions relative to incoherent motions, which is also expressed as a probability. The model combines these three elements to compute the probability of a coherent percept, as well as an estimate of the velocity of that percept.

We performed a simple psychophysical experiment to validate this model, in which subjects reported whether briefly presented plaid stimuli were perceived as coherent or transparent. We measured the probability of coherent percepts as a function of both grating speed and pattern speed (adjusted by altering the angle between the plaid directions). Although the data are lawful and reasonably consistent across observers, they reveal somewhat unexpected behaviors with no apparent intuitive explanation. We then fit the parameters of the model so as to best explain the data of each subject. We find that the model provides a remarkably accurate account of the data, offering a significant improvement over previously proposed explanations.

Furthermore, we find that the fitted components of our model are in agreement with previous experimental and modeling studies. Specifically, the recovered noise characteristics and speed priors are consistent with previous measurements of speed discrimination (Bruyn & Orbán, 1988; McKee, 1981; Orbán, de Wolf, & Maes, 1984; Stocker & Simoncelli, 2006; Welch, 1989) and low-contrast biases toward slower speeds (Stocker & Simoncelli, 2006; Stone & Thompson, 1992; Thompson, 1982; Thompson, Brooks, & Hammett, 2006). The speed priors exhibit the preference for slow speeds, which has been previously suggested as an explanation for the effects of contrast on speed perception (Hürlimann, Kiper, & Carandini, 2002; Stocker & Simoncelli, 2006) and the perceived speed and direction of coherent plaids (Langley, 1999; Simoncelli, 1993; Simoncelli & Heeger, 1992; Stocker, 2006; Weiss et al., 2002) in the context of a Bayesian observer model. Finally, the recovered prior on coherent motion suggests that the visual system prefers a single coherent velocity interpretation over one with multiple velocities (Hildreth, 1984; Langley, 1999). Thus, our study not only introduces and experimentally validates an optimal observer model for the perception of a drifting plaid but also provides compelling quantitative evidence for the ability of the

visual system to apply and generalize prior probabilities across different perceptual tasks.

Bayesian observer model

We constructed an encoder–decoder model for the selection of a visual motion interpretation, as illustrated in Figure 2. The output of the encoder stage consists of noisy measurements of the normal velocities of the two gratings that constitute the plaid, denoted as $\{\bar{m}_1, \bar{m}_2\}$. Physiologically, we assume that these correspond to the responses of two distinct subsets of noisy orientation- and speed-selective neurons, such as those found in primary visual cortex. Given these noisy measurements, the decoding portion of the model then uses the rules of statistical decision theory to select one of the two hypotheses, $\{H_{\text{coh}}, H_{\text{tran}}\}$, corresponding to coherent/transparent percepts, respectively. Specifically, the model selects the most probable of the two percepts by comparing $p(H_{\text{coh}}|\bar{m}_1, \bar{m}_2)$ and $p(H_{\text{tran}}|\bar{m}_1, \bar{m}_2)$.

The probabilities for each percept are computed using Bayes' rule. The conditional probability of the coherent percept may be written as:

$$p(H_{\text{coh}}|\bar{m}_1, \bar{m}_2) \propto p(H_{\text{coh}})p(\bar{m}_1, \bar{m}_2|H_{\text{coh}})$$

$$= p(H_{\text{coh}})\int p(\bar{m}_1, \bar{m}_2|\vec{v}, H_{\text{coh}})p(\vec{v}|H_{\text{coh}})d\vec{v}$$

$$= p(H_{\text{coh}})\int p(\bar{m}_1, \bar{m}_2|\vec{v})p(\vec{v})d\vec{v}$$

$$= p(H_{\text{coh}})\int p(\bar{m}_1|\vec{v})p(\bar{m}_2|\vec{v})p(\vec{v})d\vec{v},$$

$$(1)$$

where we have assumed that the noise in the two measurements is independent when conditioned on the true velocity. The expression includes an integral, over all possible two-dimensional velocities $\vec{v}$, of the probability that the two normal velocity measurements could have arisen from a single coherent velocity $\vec{v}$, multiplied by the prior probability $p(\vec{v})$.

In analogous fashion, we write the conditional probability for the transparent percept as:

$$p(H_{\text{tran}}|\bar{m}_1, \bar{m}_2) \propto p(H_{\text{tran}})p(\bar{m}_1, \bar{m}_2|H_{\text{tran}})$$

$$= (1 - p(H_{\text{coh}}))\iint p(\bar{m}_1, \bar{m}_2|\vec{v}_1, \vec{v}_2)p(\vec{v}_1, \vec{v}_2)d\vec{v}_1 d\vec{v}_2$$

$$= (1 - p(H_{\text{coh}}))\int p(\bar{m}_1|\vec{v}_1)p(\vec{v}_1)d\vec{v}_1 \int p(\bar{m}_2|\vec{v}_2)p(\vec{v}_2)d\vec{v}_2,$$

$$(2)$$

where we have used the fact that the prior probability of the two percepts must sum to one (i.e., we assume that these are the only two possible percepts). The expression
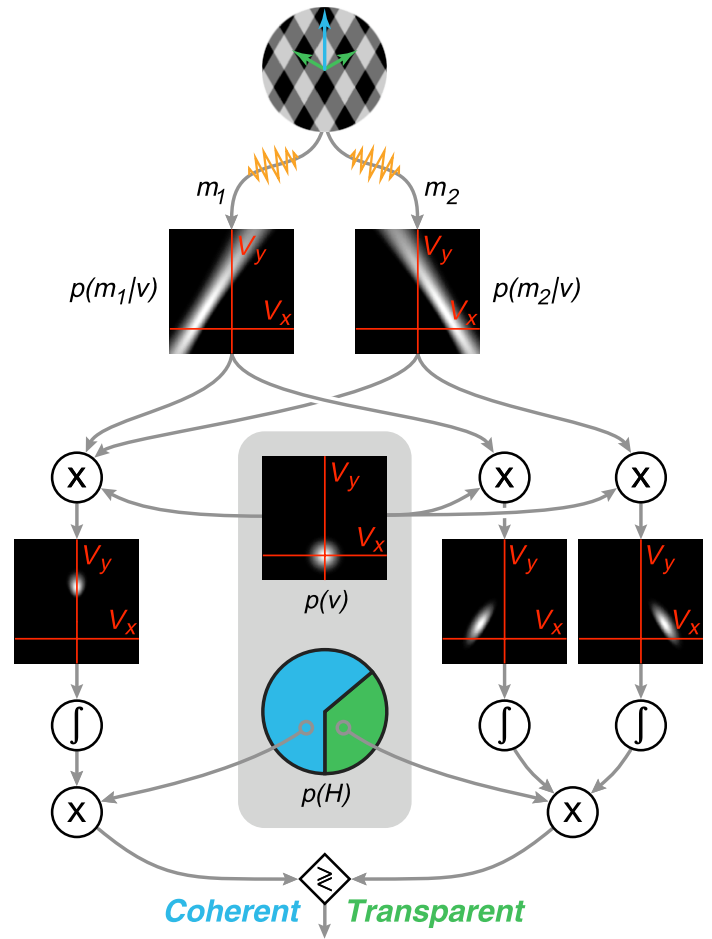

Figure 2. Illustration of the Bayesian observer model, responding to a single presentation of a plaid. In the encoding stage (not shown), an observer makes noisy measurements $\{\bar{m}_1, \bar{m}_2\}$ of the normal velocities of the two gratings. In the decoding stage, the observer forms separate likelihood functions for the two component motions based on their associated measurements. These are combined with prior preferences (internal to the observer's visual system) in order to arrive at a percept. Internal preferences are contained within the gray region and include a prior distribution over velocity, $p(\vec{v})$, and a prior probability of coherent motion, $p(H)$. For the coherent motion percept, both likelihoods are multiplied together with the velocity prior, and this posterior distribution is then integrated (left). For the transparent percept, each likelihood is individually multiplied by the velocity prior and integrated (right). The resulting scalar values are then multiplied by the internal prior for coherence/transparency, yielding posterior probabilities for each of the percepts. Finally, these are compared, and the larger one is selected as the percept.

includes integrals over the two-dimensional velocities consistent with the normal velocity measurements of each grating individually, which are then multiplied (since they represent independent probabilities). Note that we do not need to include the normalizing factor, $p(\bar{m}_1, \bar{m}_2)$, when comparing these two percept probabilities, since it is the same for both of them.

The computation of Equations 1 and 2 relies on the *likelihood function*, $p(\vec{m}|\vec{v})$, which expresses the probability that the observed normal velocity measurement $\vec{m}$ could have arisen from a stimulus moving with two-dimensional velocity $\vec{v}$. Note that the likelihood is a function of the conditioning variable $\vec{v}$ and is, thus, *not* the same function as the measurement noise distribution, which is a function of the normal velocity $\vec{m}$ (see Appendix B for derivation). For a single grating, the likelihood function lies along a ridge parallel to the measured orientation of the grating (Figure 2), indicative of the fact that the component of motion in this direction cannot be inferred from the visual input (the aperture problem, as described in the Introduction). The thickness of this ridge corresponds to the uncertainty in the measurements regarding the speed and the orientation of the grating; as one follows the ridge from its center (i.e., the point closest to the origin) outward, it fans out, corresponding to the uncertainty in the measurements regarding grating direction.

The model is built from three fundamental components. First, we must specify the distribution of the measurement noise, which is used to generate the noisy measurements and is also inverted to form the likelihood functions (see Appendix B for details). This noise limits the ability of an observer to discriminate both the speed and direction of moving gratings. In previous literature, speed discrimination thresholds at low speeds have been reported as approximately constant, whereas thresholds for moderate speeds (1 deg/s to 10 deg/s) are proportional to grating speed (Bruyn & Orbán, 1988; McKee, 1981; Orbán et al., 1984; Stocker & Simoncelli, 2006; Welch, 1989). Direction discrimination is approximately constant (when expressed in angular units; Nakayama, 1985). We capture both of these discrimination behaviors with an additive Gaussian noise model in the two-dimensional vector space of normal velocities (see Appendix B and Figure B1). We assume that this noise is separable (independent) in speed and direction and that the variance in both of these attributes grows proportional to the squared speed plus a constant. Previous measurements also suggest that the ratio of the two proportionality factors (for direction and speed, respectively) is typically 1:3 (Nakayama, 1985). In summary, the measurement noise is governed by three parameters: a constant and a proportionality factor that determine speed discriminability and a proportionality factor between speed and direction discrimination.

Second, the prior probability density over velocity is intended to represent the distribution of retinal velocities that occur during normal vision. We assume that this is circularly symmetric (i.e., a function only of speed), flat for very slow speeds, and falls as a power-law function for higher speeds:

$$p(\vec{v}) = \frac{1}{(|\vec{v}|^2 + c_4^2)^{c_5}}, \qquad (3)$$

where c_4 determines the speed at which the prior transitions from a constant regime to a power-law regime, and c_5 is an exponent that controls the rate of decay. This parametric description is consistent with previous theoretical proposals (Dong & Atick, 1995), with simulations of graphical environments (Roth & Black, 2007) and with perceptual prior models reverse-engineered from human speed discrimination data (Stocker & Simoncelli, 2006).

Finally, the model includes a (scalar-valued) prior probability that local motion on the retina is coherent (i.e., arises from a single moving source). In addition to the superposition of two transparent surfaces, there are a variety of real-world situations in which retinal motion can fail to be coherent. Examples include the boundaries of shadows (in which the shadow can move independently from the underlying surface), occlusion boundaries (in which the motion on each side of the boundary can be different), and non-translational motions (such as the surface of water or swarms of insects). We might expect that visual scenes are dominated by coherent motions, but we know of no measurements of this probability, so we treat this scalar values as a free parameter of the model.

Perceptual experiment

To validate our model, we performed a psychophysical experiment to measure human perception of plaid coherence. We constructed a set of 100 moving plaid stimuli by additively superimposing pairs of drifting square-wave gratings. The two gratings in each pair moved at the same normal speed, s_g, in directions deviating from vertical by equal (but opposite) amounts. Thus, each stimulus is determined by the grating speed and the angle between the grating normal velocities and the vertical axis, θ_g (an angle of zero corresponds to an upward moving horizontal grating). For presentation purposes, we reparameterize the stimuli in terms of the grating speed and the *pattern speed* corresponding to the unique translational velocity that is consistent with the motion of the two gratings: $s_p = s_g / \cos(\theta_g)$. Figure 3a shows the collection of stimuli that we used, plotted in terms of these two speeds.

Model simulations and data

We simulated the model on the set of symmetric plaids used in our experiments. For each combination of grating and pattern speed, we generated 50 pairs of noisy measurement samples. We then computed the probability of coherent versus transparent for each and summarized these using the fraction of trials on which transparency was more probable than coherence. This fraction is then plotted at the appropriate location in the stimulus space of
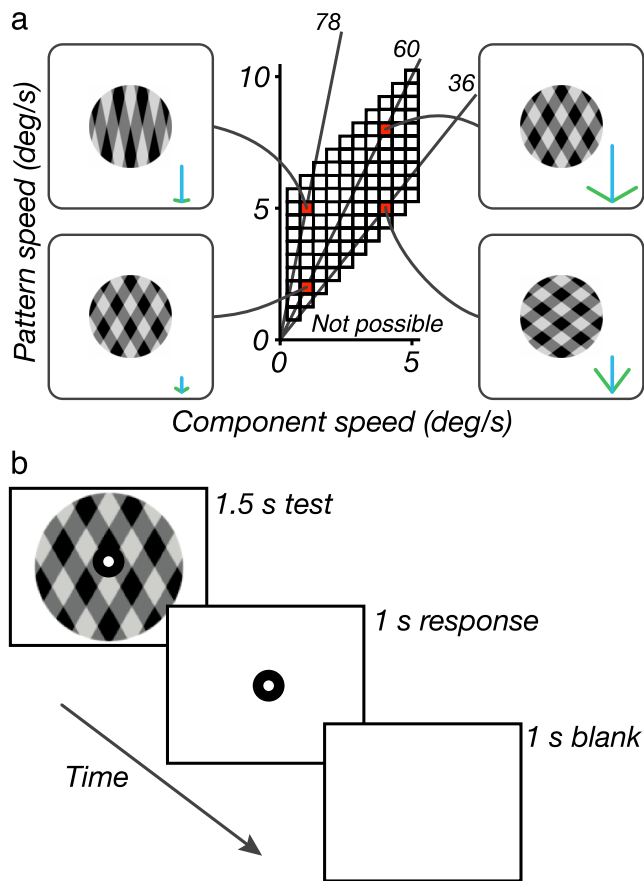
Figure 3. Parameterization of experimental stimuli and psychophysical task. (a) Plaids were generated for 100 different combinations of component and pattern speed (corresponding to small squares in the plot). Four combinations of speeds are highlighted (red squares) by showing a representative frame of the corresponding stimulus. Adjacent to each of these frames is a velocity–space diagram indicating the normal velocities of the two components (green), as well as the pattern velocity (blue). For a fixed component speed, increasing the pattern speed corresponds to increasing the angle between the two gratings. Moving along a ray emanating from the origin in the stimulus space corresponds to proportionally increasing the speed of both components and the plaid pattern, while maintaining a fixed angle. For the four sample stimuli, the plaid angle is indicated (in deg) for these four sample stimuli by numbers adjacent to gray lines extending from the origin. Note that the region in the stimulus space below the 45-deg diagonal is not physically realizable, since the pattern speed of a plaid is always faster than the component speeds. (b) Psychophysical protocol used for measuring plaid perception. Each square-wave plaid was presented for 1.5 s. This was followed by a 1-s response period during which subjects indicated whether the plaid appeared coherent or transparent by pressing a key. This was followed by a 1-s blank period, after which the sequence was repeated.

Figure 3a, with the intensity value indicating the probability of coherent perception.

Figure 4 shows a set of these plots for different model parameters. Consider the middle plot of the hexagonal array shown in Figure 4c. Plaids with small angles, slow pattern speeds, or fast component speeds are generally perceived as coherent. Plaids with very slow component speeds and large angles can also be perceived as coherent, as indicated by the dark strip running up the left side of many of the plots.

Figure 4c shows the effects of varying the three likelihood parameters, as illustrated in Figure 4a. We adjusted each parameter to lower/higher values roughly matched to the range of behaviors seen in previous literature. Reducing the likelihood width at low speeds (by either reducing the slope or reducing the constant offset) reduces the coherence of large-angle, slow-component-speed plaids (i.e., the dark strip on the left side largely disappears). Narrowing the angular extent of the likelihood also has this effect, although it is weaker (at least for the parameter set shown here). Figure 4d shows the effects of varying the three prior parameters, as illustrated in Figure 4b. The main effect seen here is that a shallower slope in the speed prior, or an increase in the prior on coherence, shifts the entire plot toward coherence.

We may understand these behaviors by considering the fundamental elements of the model and the geometry of plaid velocities. The pattern speed is always faster than the normal speed of the two grating components, and this ratio grows with increasing plaid angle. Since the prior probability on speeds imparts a preference for slower speeds, the fast-moving coherent pattern will generally be considered less probable than the slower individual components, thus favoring the transparent interpretation (Farid & Simoncelli, 1994). This effect increases for plaids with a larger ratio of pattern to grating speed ("steep" plaids). For very slow speeds, the speed prior becomes flat and no longer favors transparent percepts.

An additional influence at slow speeds is that the measurement noise variance ceases to scale proportionally with speed and becomes flat. As a result, the likelihood ridge will become broader relative to the grating speed, and it will "fan out" to cover a broader range of angles. The net result of this is that the likelihoods for slow component gratings in a plaid with a very large angle can have significant overlap at slow pattern speeds and, thus, can still appear coherent (dark strip on left side of plots).

In addition to these influences of the speed prior and likelihood, two other properties of the model affect the coherence/transparency percept. The model may be interpreted as performing a form of "Bayesian model selection," in which coherent and transparent motion models are compared in terms of their ability to explain the measurements. Such decision rules are known to naturally enforce a form of Occam's razor, by favoring solutions with smaller numbers of parameters, for which the likelihood is more "concentrated" (Jefferys & Berger, 1991). In our context,
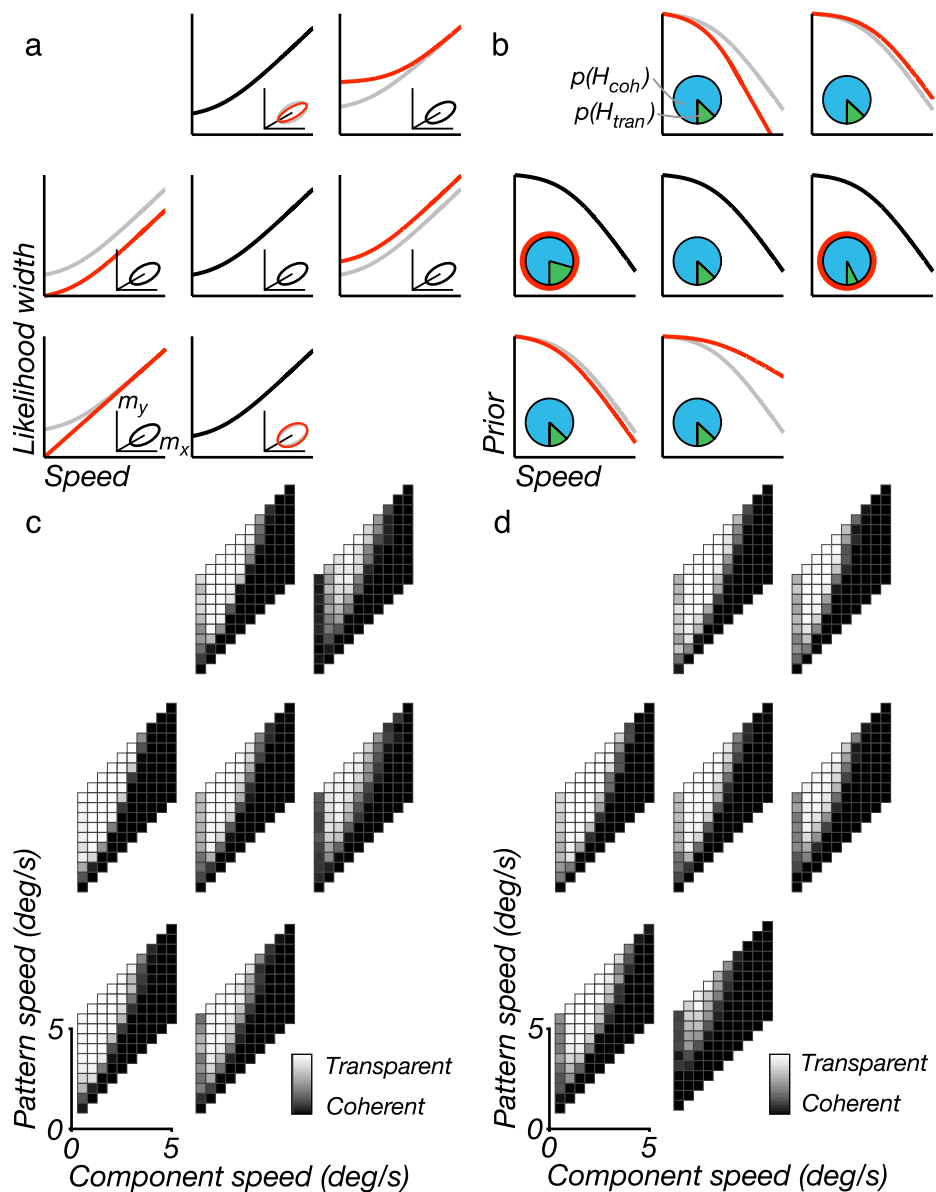
Figure 4. Model simulations for different parameter settings. (a) Illustration of seven different settings for the likelihood parameters. The black line in the central log–log plot shows the default width of the speed likelihood as a function of speed. The inset plot shows the default shape of the measurement distribution (see Figure B1a), thus illustrating the relationship between speed and direction uncertainty. Each of the surrounding six log–log plots and insets shows a variation of one of the three likelihood parameters, as highlighted in red (default values are redrawn from the center plot, in gray, for comparison). For example, the upper right and lower left plots show an increase or decrease of the speed at which the transition occurs, c_1, respectively. Left and right plots show a change in the proportionality factor at high speeds, c_2. Upper/lower plots show a change in the proportionality between the standard deviations for speed and direction, c_3. (b) Illustration of seven different settings for the prior parameters. The black line in the central log–log plot shows the speed prior. Inset pie chart shows the prior for the two hypotheses (blue for coherent, green for transparent). Each of the surrounding six log–log plots and insets shows a variation of one of the three prior parameters. The upper right and lower left plots show changes in the speed at which the prior transitions from a constant regime to a power-law regime, c_4. Upper/lower plots show a change in the rate of decay, c_5. Left/right plots show a change in the coherence prior, $p(H_{coh})$. (c, d) Simulated "percepts" of the model, corresponding to parameter values indicated in (a) and (b), respectively. For each of the grayscale plots, the individual squares correspond to plaid stimuli with different component and pattern speeds (see Figure 3a), and the intensity of each square indicates the probability that the associated plaid stimulus is perceived as transparent (black = 0%, white = 100%).
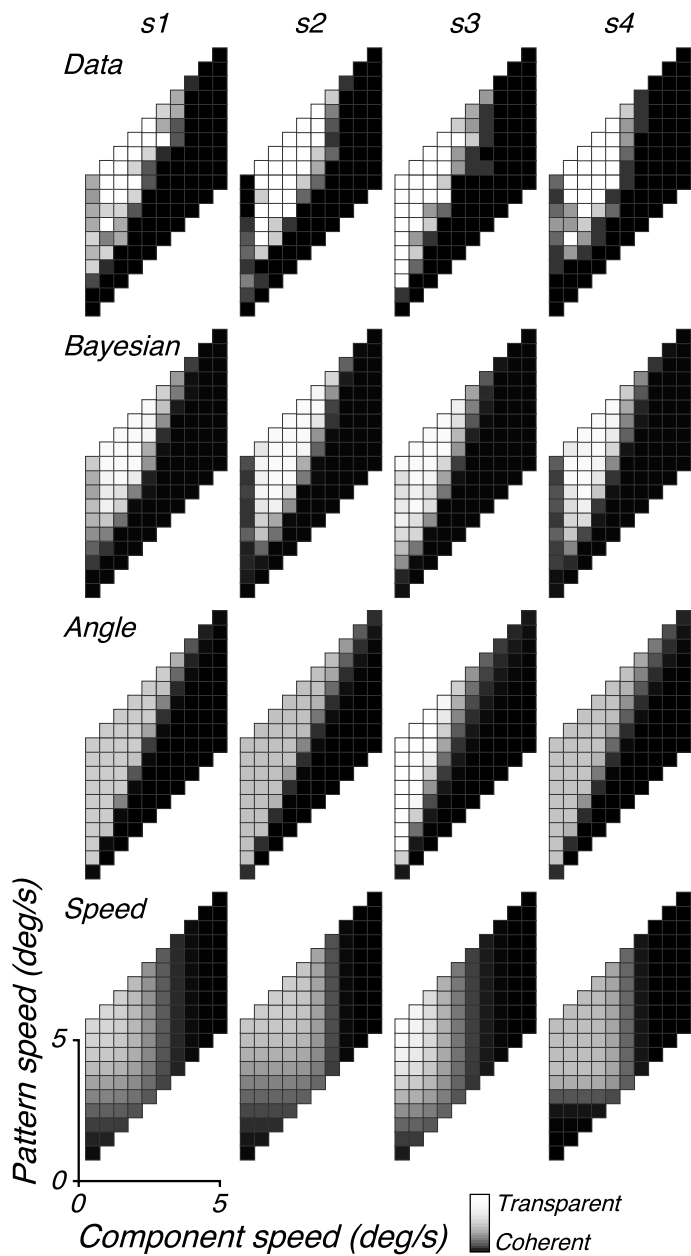
Figure 5. Perceptual data for four subjects, together with simulated percepts of three models that were fit to the data. In all plots, the intensity of each square indicates the probability that the associated plaid stimulus is perceived as transparent (see Figure 4).

the coherent solution is parameterized by a single pattern velocity, whereas the transparent solution requires two component velocities, and thus, our model embeds a natural preference for coherency. Finally, the model includes a scalar-valued prior probability for coherence/transparency that can further adjust the preference for either interpretation.

We obtained data from four subjects, who were asked to indicate whether the motion of briefly presented plaids appeared coherent or transparent (Figure 3b). The top row of Figure 5 shows the data for all four subjects. All four

subjects reported primarily coherent percepts for plaids with shallow angles (although it is not clear whether this is governed by the angle or by the component speed). Three of the subjects (all but s3) showed coherent percepts for stimuli with slow pattern speeds (bottom squares in the data plots) and also for slow component gratings with a large angle (i.e., fast pattern speeds, leftmost column of squares in the data plots). The pattern of percepts seen in all four subjects is qualitatively similar to that seen in our model predictions (Figure 4).

To assess the ability of the model to explain the human data, we optimized the six model parameters so as to maximize the probability of each subject's data (see Appendix C). The second row of Figure 5 shows simulations of the best-fitting Bayesian model, which is seen to provide a close match to the individual data of each of the subjects (Figure 5, top row).

We compared these simulations to those arising from two other models that have been proposed. The "angle" model is based on the hypothesis that plaid perception is a monotonic function of the plaid angle (larger angles being more transparent; Hupé & Rubin, 2003; Kim & Wilson, 1993). We implemented this model by assuming that the proportion of "coherent" judgments for a stimulus with grating/pattern speeds $\{s_g, s_p\}$ could be expressed as

$$p(\text{"coherent"}|s_g, s_p) = w(\cos^{-1}(s_g/s_p)), \qquad (4)$$

where $w(\cdot)$ is a Weibull function with four parameters controlling the angle and slope of the transition from coherent to transparent and the saturating (min/max) response levels on either side of the transition. For each subject, we fit these four parameters by maximizing the likelihood (Figure 5, third row). For three of the four subjects (all but s3), this model does not provide a good description of data. The subjects perceive large-angle plaids with slow components as coherent, and in addition, the boundary between the transparent and coherent regions for larger component speeds does not appear to lie along a straight line emanating from the origin.

We also considered a "speed" model, based on the hypothesis that transparency arises because the visual system prefers slow-speed interpretations of moving stimuli and, thus, is most likely to occur for stimuli combining fast pattern speeds with slow component speeds (Farid & Simoncelli, 1994; Farid, Simoncelli, Bravo, & Schrater, 1995). We implemented this model by assuming that the proportion of "coherent" judgments was determined by comparing both grating and pattern speeds to an (unknown) reference speed and combining these results separably:

$$p(\text{"coherent"}|s_g, s_p) = w_g(s_g) \cdot w_p(s_p), \qquad (5)$$

where w_g and w_p are both four-parameter Weibull functions, constrained to have the same transition speed

and the same saturating level for large values (i.e., in the transparent region). We fit the six parameters of this model to each subject by maximizing the likelihood of their data (Figure 5, fourth row). For all four subjects, this model does not provide as good of a description as the Bayesian model.

We quantified the relative performance of these models using a simple metric. For each subject, we computed the mean log-probability of the psychophysical data for all trials and for all stimuli. Since each subject has a different amount of variability in their data, we normalized these log-probability values to a scale ranging from the log-probability obtained from a fully random "coin-flipping" model (i.e., one that draws responses according to a fixed probability, regardless of stimulus) to that of an "omniscient" model in which responses for each stimulus are drawn according to a stimulus-specific probability. The coin-flipping model provides a lower bound on model performance: it ignores the stimuli, flipping a biased coin on each trial to determine a percept. The bias of the coin is the only parameter of the model and is set (for each subject) to the proportion of coherent judgments averaged over all stimuli and all trials. The omniscient model provides an upper bound on performance and has 100 parameters, corresponding to the probability of a coherent percept for each of the 100 stimulus conditions. For each subject, we set these parameters to the values that maximize the likelihood of observing the data, which simply correspond to the proportion of coherent responses for each stimulus.

A comparison of the models, on this normalized log-probability scale, is shown in Figure 6a. Note that error bars are not shown, since bootstrapping our model-fitting procedure, along with the normalization of the plotted likelihoods, is computationally prohibitive. For all subjects, the Bayesian model fits the psychophysical data nearly as well as the omniscient model and better than the other two models. For subject s3, the angle model performs reasonably well, but it clearly fails to explain the data of the other subjects (consistent with the appearance of the predicted responses in Figure 5). The speed model generally performs poorly, although it is slightly better than the angle model for the fourth subject. Perhaps more importantly, it is worth noting that both the speed and angle models are merely descriptions of the data for this particular experiment, and neither has an obvious generalization to real-world stimuli.

One might be concerned that the poor fit of the angle model is due to the fact that it has fewer parameters (four, as compared to the six parameters of the Bayesian and speed models), but this seems unlikely. As mentioned previously, the shape of the subject data simply do not adhere to an angular form. We can also attempt to compensate for the differences in the number of free parameters using a standard methodology for model comparison, the Bayesian information criterion (BIC; Schwarz, 1978), computed as $n\ln(\sigma_e^2) + k\ln(n)$, where n is the number of stimulus conditions, σ_e^2 is the variance of
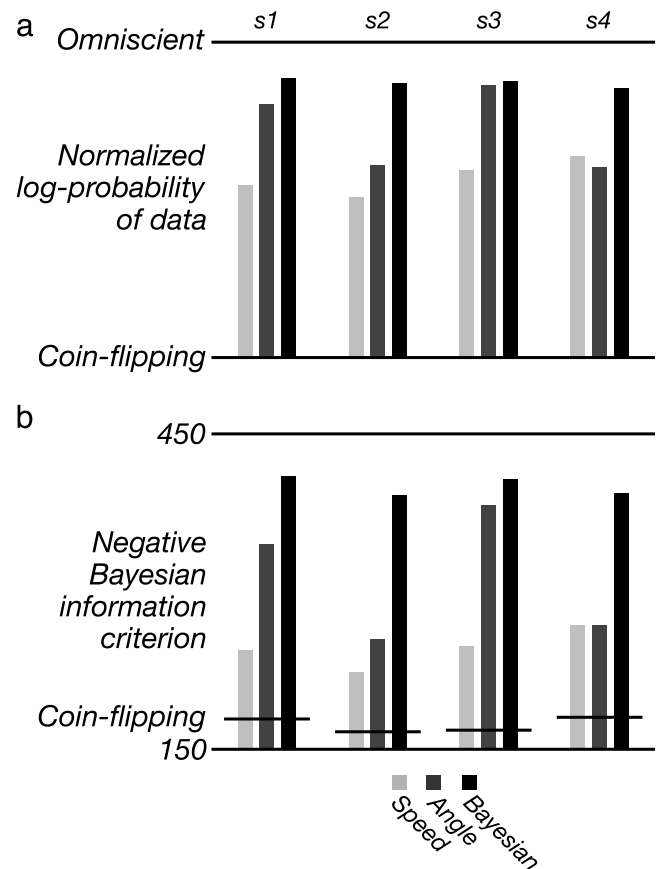

Figure 6. Quantitative comparison of models. (a) Normalized mean log-probability of the psychophysical data of the four subjects for the three different estimator models. These probabilities are expressed on a scale that varies from the value obtained for a random ("coin-flipping") model to one that knows the probability of coherent responses for each stimulus condition ("omniscient"). (b) Negative of the Bayesian information criterion of the same data for the same estimator models. The values for the coin-flipping model for each subject are shown as horizontal lines.

the model error (across stimulus conditions), and k is the number of free parameters in the model. The (negative) BIC values for the models are plotted in Figure 6b. For all subjects, the Bayesian model is seen to do a better job of fitting the data, despite the penalty for having as many, or more, free parameters.

Estimates of the observer likelihood and priors

The Bayesian model is derived from a set of fundamental components, and these may be interpreted and tested beyond the confines of the plaid coherence experiment. As such, it is worth examining the form of the likelihood and prior functions that we obtained by fitting
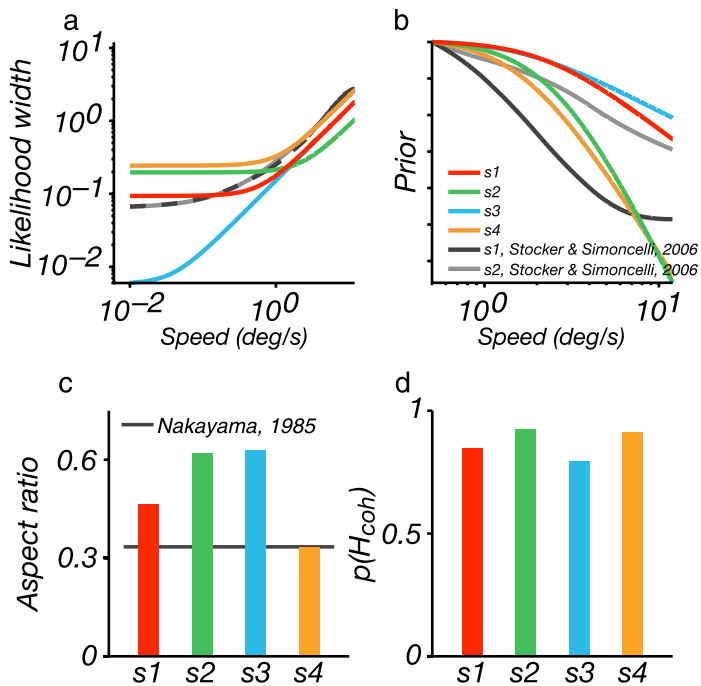
Figure 7. Model likelihoods and priors, optimized to fit each subject's coherence data. (a) Speed likelihood widths, as a function of speed, shown in a log–log plot. These are flat for low speeds and grow linearly for high speeds. (b) Speed priors, shown in a log–log plot. These are flat for low speeds and fall as a power law for high speeds. (c) The aspect ratios of the elliptical measurement distributions (which govern the ratio of discriminability in speed and direction). Horizontal line indicates a value of 0.33 typical of previous studies (Nakayama, 1985). (d) The values of the prior for the coherent interpretation, $p(H_{coh})$, are centered around 0.85, indicating that observers believe that singular motions are more common.

our subjects' data. The likelihood widths in our model are parameterized by a function that is constant at low speeds, transitioning to a linearly increasing behavior at higher speeds. The fitted values for the width at low speeds and the transition speed vary significantly across subjects (Figure 7a). The transition for most subjects is around 1 deg/s, but s3 indicates a transition at roughly 0.02 deg/s. This explains the fact that the data for this subject show no ridge of coherence at the slowest component speed. The likelihood widths of the other three subjects are remarkably similar to those obtained for two subjects from direct measurements of speed discriminability (Stocker & Simoncelli, 2006; Figure 7a, black and dashed gray curves).

The speed priors that we estimated for our subjects are shown as log–log plots in Figure 7b, with all curves shifted vertically to match a common value at the leftmost point. Our two-parameter model allows for a transition from a flat prior at slow speeds to one that falls as a power law at high speeds. All four subjects showed a transition to power-law behavior at roughly 2 deg/s. Subjects 2 and 4 had relatively steep priors, and the other two were

shallower. This grouping also corresponds to similarities in these subjects' patterns of coherent percepts, with s1 and s3 showing a similar pattern, and s2 and s4 showing the same. These prior distributions are similar to those obtained by reverse-engineering a Bayesian speed estimation model from speed discriminability data for two subjects (Stocker & Simoncelli, 2006; Figure 7b, black and dashed gray curves).

The values for the remaining two constants are plotted in Figures 7c and 7d. The likelihood aspect ratio had a mean value of 0.51 and was greatest for s3 and s2. The value of the coherence prior had a mean value of 0.87 and was relatively consistent across subjects, suggesting a significant preference for coherent percepts by our subjects' visual systems.

Discussion

We have studied human perception of the coherence of symmetric square-wave plaids over a range of different speeds and angles and developed a model for the perceptual interpretation of such stimuli. The model determines which of the two interpretations is more probable, given the sensory evidence and the observer's prior beliefs. Specifically, starting with noisy measurements of the speed and direction of the constituent gratings, the model arrives at a perceptual interpretation by combining three fundamental ingredients according to the rules of optimal statistical inference. These are: (1) a probabilistic description of the noise in the measurements (i.e., a likelihood function); (2) a prior probability over speed; and (3) a prior probability for stimulus coherence. The first two ingredients have been used in previous Bayesian models for motion perception (Hürlimann et al., 2002; Montagnini, Mamassian, Perrinet, Castet, & Masson, 2007; Simoncelli, 1993; Simoncelli & Heeger, 1992; Stocker & Simoncelli, 2006; Weiss, 1998; Weiss et al., 2002), which estimate speed or velocity from visual input. The third ingredient has been suggested in the computer vision literature (Hildreth, 1984; Weiss, 1998), and a related preference for singular interpretations was used in a model developed by Langley (1999) to account for coherence/transparency of plaids. The overall structure of our model is similar to a number of "Bayesian model selection" formulations used to describe perceptual cue-combination phenomena, in which the observer must implicitly decide whether to interpret two cues as arising from one or two sources (Körding et al., 2007; Natarajan, Murray, Shams, & Zemel, 2009; Sato, Toyoizumi, & Aihara, 2007; Yuille & Bülthoff, 1996).

We have fit parametric forms of all three ingredients to perceptual coherence data, which result in an excellent account of the data for each subject. The quality of the fits relies on use of a likelihood whose width is proportional to speed but flattens at low speeds, coupled with a speed

prior that falls as a power law (and also flattens at slow speeds), both previously proposed to explain human speed discrimination behaviors (Stocker & Simoncelli, 2006). In particular, Gaussian priors and/or speed-independent likelihood functions, as used in earlier Bayesian models for velocity (Hürlimann et al., 2002; Simoncelli, 1993; Weiss et al., 2002), result in a poor account of the data (not shown). Inter-subject differences are explained by perturbations in the shape or magnitude of the model likelihood widths or priors. Furthermore, we find that the fitted likelihood widths and priors are generally consistent with previously published estimates obtained from speed discrimination measurements for single gratings (Stocker & Simoncelli, 2006). We find this last point quite remarkable, given that our experimental data provide no direct measurement of our subjects' perception of stimulus speed.

The literature contains a number of studies of plaid coherence that are of direct relevance. Hupé and Rubin (2003) and Kim and Wilson (1993) each measured perceptual coherence of plaids and concluded that the angle was the primary determinant of the perceptual interpretation. Direct comparison to these studies is difficult because the stimuli differed from ours in many ways (grating type, window size, eccentricity, contrast), as did the subjective task. Nevertheless, we can see that the behavior of each of our subjects (Figure 5) is consistent with this description but only over a portion (or a one-dimensional slice) of the speed/angle parameter space. Three of the subjects (Figure 5, columns 1, 2, and 4) show clear violations of an angle model, with coherent regions that extend vertically on both the left side (very slow component speeds, below 1 deg/s) and right side (fast component speeds, above 3.5 deg/s) of their data plots.

Farid and Simoncelli (1994) examined the coherence of square-wave plaids and explored the same parameter space as the experiments reported here but over a different range. They found that a primary determinant of coherence was the pattern speed: plaids with pattern speeds faster than a cutoff speed of roughly 5–6 deg/s were more likely to be perceived as transparent. They also found that subjects had trouble making coherence judgments when the grating speed exceeded this speed. In a later study, they concluded that these behaviors could arise because of a preference for slower speed interpretations (Farid et al., 1995). Our data are again broadly consistent with those findings over some parameter ranges, but we see that, for the extended range of conditions measured here, a simple model based on pattern speed clearly fails to fit the data (Figure 5, row 4).

Langley (1999) developed a model for coherent/transparent motion perception based on spatiotemporal gradient measurements, motivated by concepts similar to those developed here. He incorporated two regularization constraints that express preferences for slow speeds and for a reduction in the degrees of freedom used to explain the data. The latter can be interpreted as a preference for

coherence, since a coherently moving stimulus can be explained using a single velocity. Perceptual transparency/coherence predictions were made by computing a measure of certainty that the data could be accounted for by a unique translational velocity. These reported predictions are aligned with human perception, although direct comparisons to data were not provided in the paper.

In general, we can see that experimental support for the models proposed in these previous publications was limited by the range of stimulus parameters that were examined. Our study, while covering a larger range of a two-dimensional parameter space, is limited in this same sense, and we cannot be sure that it will accurately predict subject behavior for parameters that lie significantly beyond the range we have tested. In order to pave the way for future experimental investigation, we have simulated the model (with parameter values set to the mean of the best-fitting values for our subjects) over a substantially larger stimulus set (Figure 8a). We see that the fundamental behaviors of the original data (Figure 5) extrapolate in a natural way and continue to support our conclusions. In particular, the primary transition boundary between transparent and coherent percepts continues to curve upward, inconsistent with the angle model. In addition, the ridge of coherence along the left side of the plot continues to extend upward, with little sign of a transition to transparency. We also computed predicted percepts for *asymmetric* plaids, in which the two gratings differ in speed by a fixed scale factor. Figure 8b shows that the percepts for plaids with grating speeds in a ratio of 1:3 are nearly the same as those for symmetric plaids. Finally, we computed predictions for asymmetric plaids in which the normal velocities of the gratings lie on the same side of the pattern velocity (these are known as "type II" plaids in the literature (Kim & Wilson, 1993)). Figures 8c–8e show predictions for three different grating speed ratios and show a range of behaviors.

For each of our human subjects, the fitted likelihood and prior functions can also be used to directly predict their performance in estimating or discriminating the speed or direction of gratings. Since the prior preferences for slower speeds in our recovered models are similar to those found in Stocker and Simoncelli (2006), we would expect the predictions of speed estimation and discrimination to be consistent. Moreover, the likelihood and prior functions can be combined to directly predict the perceived speed and direction of plaids that are seen as coherent, as has been done with previous Bayesian models (Hürlimann et al., 2002; Simoncelli, 1993; Simoncelli & Heeger, 1992; Weiss, 1998; Weiss et al., 2002). In addition to perceptual predictions, the prior probability on speed could be compared to the distribution of retinal speeds encountered during natural vision. A power-law form for this distribution has been proposed based on theoretical studies (Dong & Atick, 1995) and simulations of graphical environments (Roth & Black, 2007), but it is difficult to measure directly (since it requires properly accounting for body, head, and
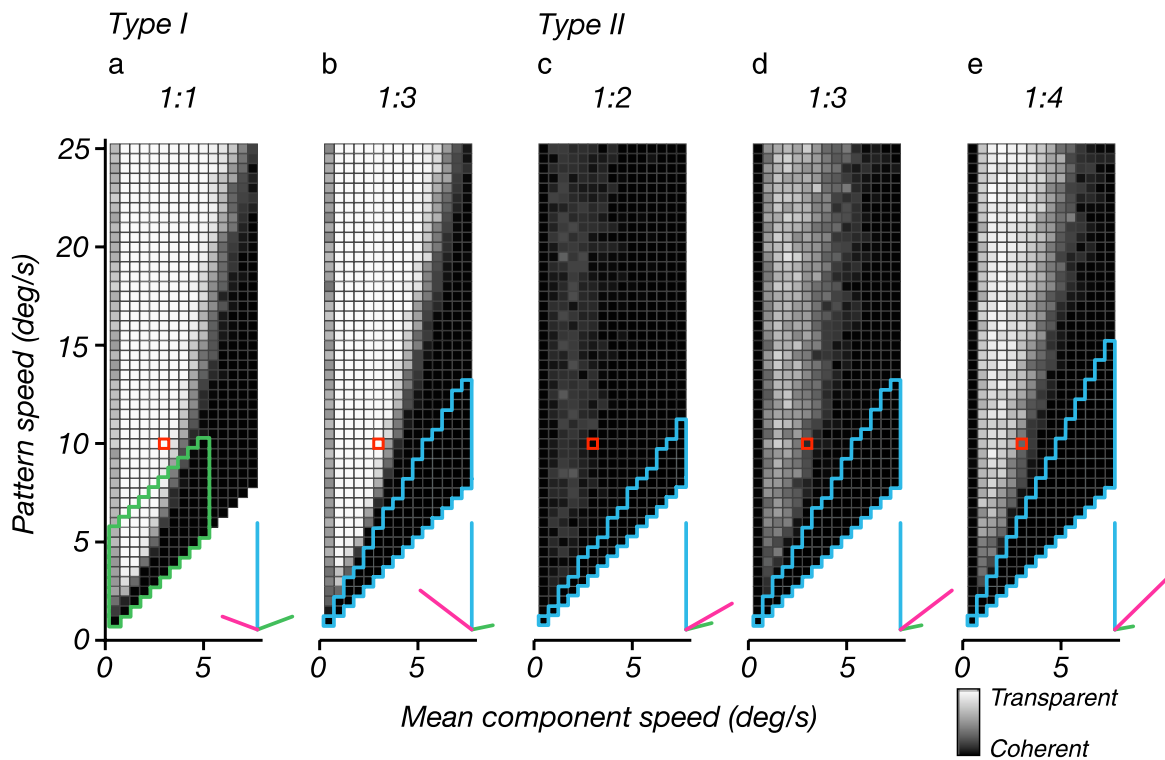
Figure 8. Model predictions, computed using parameters averaged over those used to fit our subjects' data. In all plots, the intensity of each square indicates the probability that the associated plaid stimulus is perceived as transparent (see Figure 4). Insets indicate the normal velocities and pattern velocity of a sample stimulus (corresponding to the red-outlined square in the coherence plot). (a) Coherence of symmetric plaids, over an extended region of the parameter space. The green boundary encloses the stimulus parameters covered by our experiments. (b) Coherence of asymmetric plaids, with normal speeds of the two gratings in a ratio of 1:3. The horizontal axis is the geometric mean of the two grating speeds, and the vertical axis is the pattern (IOC) speed. Region enclosed by blue boundary at bottom right is physically unrealizable. (c–e) Coherence of asymmetric one-sided plaids (also known as "Type II" plaids), in which the normal velocities are both on the same side of the pattern velocity (see insets). The three panels are computed for three different normal speed ratios.

eye movements, including during fixation and pursuit). Advances in portable eye-tracking equipment should eventually make this feasible (Eckert & Zeil, 2001). Similarly, the coherence prior could be compared with the probability of a patch of the retinal image corresponding to a single opaque surface in the world.

Our model uses simple parametric forms for likelihood widths and priors, which allow it to be fully constrained by the perceptual data. However, each of these ingredients could be further elaborated to extend the model to a wider variety of stimulus configurations. We assumed a rotation-invariant noise model, for example, such that the variability in measured grating speed and direction is independent of the true direction. We also assumed a rotation-invariant speed prior. Humans are known to exhibit anisotropies in direction judgments, and plaids with oblique IOC directions are more likely to appear transparent than those with cardinal directions (Hupé & Rubin, 2004). These inhomogeneities in direction could be incorporated into the model through a direction-dependent noise model and/or a direction-dependent prior. Additional noise dependencies on stimulus attributes such

as contrast (as in Hürlimann et al., 2002), luminance, spatial frequency (Smith, 1992), size, hue, and depth could be incorporated, based on measurements of grating speed discriminability as a function of those attributes (see Stocker & Simoncelli, 2006). Ultimately, we desire a likelihood model that predicts motion discriminability for any spatial stimulus pattern. Note that in the current model, the priors over speed and coherence should not be affected by inclusion of any variables that are statistically independent of speed and coherence in the natural environment. Finally, our experiments were performed using foveally presented stimuli. Peripherally viewed stimuli are likely to exhibit different perceptual discriminability, which could again be incorporated into the noise description of the model. However, in addition, the position of stimuli would likely require an elaboration of the velocity prior, since the velocity of the retinal image content in an active (fixating, pursuing) observer will depend on location relative to the fovea.

The perceived coherence of plaids is also known to depend on viewing duration, and with long viewing durations (more than 20 s), the percept becomes bistable

Hedges, Stocker, & Simoncelli

(Hupé & Rubin, 2003; von Grünau & Dubé, 1993). Given that our experiments were performed with short duration stimulus exposures (1.5 s), it is unlikely that our subjects would have experienced a transition from one state to the other. Models of bistable percepts commonly posit a competitive interaction between two subpopulations of cells, each receiving a stimulus-driven input that governs the relative strength of (or time spent in) each of the two perceptual states (Moreno-Bote, Rinzel, & Rubin, 2007). Thus, we can view the response of our model just before the final decision stage as providing appropriate signals for this purpose.

Finally, our model operates at an abstract computational level, specifying what is to be computed rather than how to compute it. Nevertheless, it is broadly consistent with the known physiology of the mammalian motion pathway. In particular, the measurements of the two normal motions could be replaced by responses of a population of cells selective for moving oriented components, such as those found in primary visual cortex. In addition to providing a physiological substrate for the computation, this interpretation also provides a generalization of the model to arbitrary visual stimuli.

The decoder stage must map the noisy responses of the encoder population to a decision about the coherence/transparency of the stimulus. This mapping might be achieved explicitly, by computing likelihoods (e.g., Deneve, Latham, & Pouget, 1999; Jazayeri & Movshon, 2006; Ma, Beck, Latham, & Pouget, 2006; Zemel, Dayan, & Pouget, 1998; Zhang, Ginzburg, McNaughton, & Sejnowski, 1998), multiplying them with a prior (and possibly with each other), integrating them, and comparing the outputs to arrive at a decision. Alternatively, it might be implemented more directly as a parametric mapping from noisy measurements to estimates/decisions, without an explicit physiological correlate for each probabilistic ingredient (Fischer, 2010; Simoncelli, 2009; Stocker & Simoncelli, 2006).

In either case, our current model is unrealistic in assuming a deterministic decoding stage—variability in model responses is due entirely to the measurement noise, which propagates through the decoder. Incorporating noise into the decoding stage would provide a sort of "pooling" or "decision" noise (e.g., Shadlen, Britten, Newsome, & Movshon, 1996), while sacrificing some of the optimality of the decoder. Some recent theories posit that variable responses in populations should be interpreted as samples from the posterior distribution rather than a representation of a deterministic estimate (Berkes, Orbán, Lengyel, & Fiser, 2011; Fiser, Berkes, Orbán, & Lengyel, 2010; Hoyer & Hyvärinen, 2003). Some Bayesian models for perception have also utilized posterior sampling to account for response noise (e.g., Mamassian, Landy, & Maloney, 2002), which results in "probability matching" behavior (Herrnstein, Rachlin, & Laibson, 1997). In our model, drawing a binary sample from the

posterior ($p(H_{\mathrm{coh}}|\bar{m}_1, \bar{m}_2)$) would increase variability in coherency/transparency judgments across all stimuli, which seems inconsistent with regions of the stimulus space in which subjects show no variation in interpretation (e.g., where all trials are judged coherent). In conclusion, the means by which probabilistic computations are accomplished with neurons is a topic of many recent theoretical studies. However, the full sequence of computations that underlies perceptual inference, including the representation and learning of prior information, remains a fundamental and unresolved topic for future investigation.

Appendix A

Psychophysical experiments

Four male subjects with normal or corrected-to-normal vision participated in our psychophysical experiments. Experimental procedures were approved by the New York University Committee on Activities Involving Human Subjects and all subjects signed an approved consent form. Three of the subjects (1–3) were not aware of the purpose of the study. Subjects were given brief instructions at the beginning of a block of trials. We reviewed the temporal sequence of a single trial, by showing them the diagram of Figure 3b. We explained that plaids can appear to move in two different ways, but subjects were not told anything about the stimulus parameters that would be adjusted or how these might affect their percepts. We allowed them to practice the task for a minute or two, verifying that they reported having seen both percepts. All subjects learned the task easily.

All stimuli were symmetric additive square-wave plaids with an upward pattern motion direction. Component gratings had a spatial frequency of 1.5 c/deg and contrast of 0.4. The plaids were presented within a circular aperture (5-deg diameter), with an edge profile following a raised cosine function. The spatial windowing transitioned to full contrast by the center of the aperture, meaning it was at half-value at 1.25 deg from the edge. Subjects viewed the monitor from a distance of 65 cm, such that pixels subtended 0.031 deg. Stimuli were presented at a refresh rate of 120 Hz for 1.5 s and were temporally windowed with a squared cosine function. The temporal window reached its maximum value at 15 ms after onset (approximately two frames). A fixation mark, composed of a small black circle within a white annulus, was overlaid at the center of the aperture during stimulus presentation. After each stimulus presentation, subjects were given 1 s to indicate whether the plaid appeared to move coherently or transparently by pressing a key. This was followed by a 1-s blank period. Brief tones were presented at the onsets of the test and response periods to

assist subjects in timing their responses. This sequence is illustrated in Figure 3b.

The stimulus set, as illustrated in Figure 3a, included component grating speeds ranging from 0.5 to 5 deg/s in 0.5 deg/s increments. Pattern speeds ranged from 0.5 to 5 deg/s *faster* than the corresponding component speeds, again in 0.5 deg/s increments. For example, plaids with the slowest component speed of 0.5 deg/s were presented with pattern speeds ranging from 1 to 5.5 deg/s. The angle between the normal directions of the components was twice the arccosine of the ratio of component to pattern speed. For example, for a component speed of 0.5 deg/s, the half-angles between the directions of the component motions ranged from 60 to 84.8 deg. The sequence of stimuli presented during the experiment was randomized, with each stimulus presented at least six times.

Appendix B

Derivation of model likelihood function

Consider a grating, with normal orientation θ, moving rigidly with 2D velocity $\vec{v}$. The normal speed of the grating is $\vec{v} \cdot \hat{u}_\theta$, where $\hat{u}_\theta = [\cos(\theta), \sin(\theta)]$, a unit vector in the θ direction. An observer makes a measurement, $\vec{m}$, of the 2D normal velocity of this grating that is corrupted by 2D Gaussian noise, with standard deviation $\sigma_s(\vec{v} \cdot \hat{u}_\theta)$ in speed and $\sigma_d(\vec{v} \cdot \hat{u}_\theta)$ in direction:

$$
\begin{aligned}
p(\vec{m} \,|\theta, \vec{v}) = {} & \exp[-(\vec{m} \cdot \hat{u}_\theta - \vec{v} \cdot \hat{u}_\theta)^2/(2\sigma_s^2(\vec{v} \cdot \hat{u}_\theta)^2)] \\
& \times \exp[-(\vec{m} \cdot R\hat{u}_\theta)^2/(2\sigma_d^2(\vec{v} \cdot \hat{u}_\theta)^2)] \\
& / 2\pi\sigma_s(\vec{v} \cdot \hat{u}_\theta)\sigma_d(\vec{v} \cdot \hat{u}_\theta),
\end{aligned} \tag{B1}
$$

where R is a 2×2 matrix that performs a rotation by $\pi/2$. This measurement distribution is illustrated in Figure B1a. We assume that standard deviations for speed and direction, $\sigma_s(\vec{v} \cdot \hat{u}_\theta)$ and $\sigma_d(\vec{v} \cdot \hat{u}_\theta)$, are a function of speed (Stocker & Simoncelli, 2006) and parameterized them as:

$$
\sigma_s(s) = c_2\sqrt{(s^2 + c_1^2)} \qquad \sigma_d(s) = c_3\sigma_s(s). \tag{B2}
$$

This function specifies a standard deviation that is constant at low speeds and proportional to speed at high speeds. The parameter c_1 determines the speed at which this transition occurs, and parameter c_2 determines the proportionality factor at high speeds. The standard deviation of the distribution in terms of direction is proportional to that in terms of speed, with parameter c_3 controlling this proportionality. A previous review has suggested that c_3 is approximately 0.33 (Nakayama, 1985), and we have
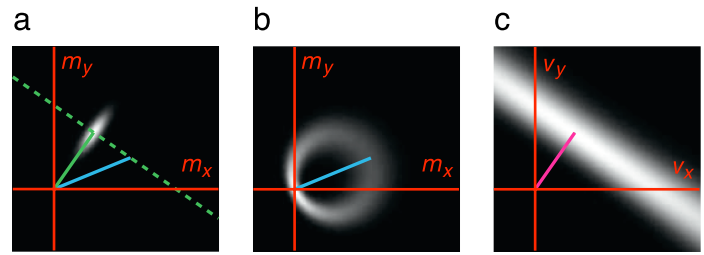

Figure B1. Derivation of the likelihood function. (a) Gaussian probability distribution of normal velocity measurements (gray-scale), $p(\vec{m}|\theta, \vec{v})$, for a grating with normal velocity specified by the green vector, but moving at physical velocity specified by the blue vector. The constraint line (dashed green) of all translational velocities consistent with the normal velocity of that grating is also shown. (b) The distribution of normal velocity measurements (grayscale), $p(\vec{m}|\vec{v})$, for an arbitrary spatial pattern moving with the specified physical velocity (blue vector). This is computed by integrating over all directions, θ, as in Equation B3. (c) The likelihood, a function of $\vec{v}$, obtained by evaluating $p(\vec{m}|\vec{v})$ for a particular normal velocity measurement $\vec{m}$ (magenta vector) drawn from the distribution in (a).

used 0.35 as an initial value from which to start the optimization.

The full measurement probability distribution for a pattern moving at velocity $\vec{v}$, with arbitrary spatial content, is obtained by integrating this expression over directions:

$$
p(\vec{m}|\vec{v}) = \int p(\theta)p(m|\theta, \vec{v})d\theta. \tag{B3}
$$

We assume that the prior distribution over orientations, $p(\theta)$, is uniform. The resulting measurement distribution is illustrated in Figure B1b, corresponding to a probabilistic version of the circle defining the set of normal velocities consistent with a given pattern velocity (see Adelson and Movshon, 1982, Figure 3). Finally, the likelihood function is obtained by evaluating this measurement distribution as a function of $\vec{v}$, for a particular measurement $\vec{m}$. The result is a "ridge," orthogonal to the measured normal velocity (Figure B1c).

Appendix C

Fitting the Bayesian model to data

The model depends on six parameters, three controlling the measurement noise (defined in Equation B2), two controlling the velocity prior (defined in Equation 3), and the last being the value of $p(H_{\mathrm{coh}})$. Let $\vec{c}$ be a vector containing these model parameters. We fit these parameters to the psychophysical data of each subject by

maximizing the likelihood. These fitted parameter values were then used to simulate trials of the experiment, in order to generate the model plots shown in Figure 5 (second row).

For each stimulus *s,* the model operates by generating two random measurements $\{\bar{m}_1, \bar{m}_2\}$ and then deciding on a percept by comparing the two posterior probabilities $p(H_{\text{coh}}|\bar{m}_1, \bar{m}_2)$ and $p(H_{\text{tran}}|\bar{m}_1, \bar{m}_2)$. Although this decision process is deterministic, the measurements are stochastic (drawn from the measurement density of Equation B1), and thus, repeated presentations of the same stimulus produces "coherent" responses with a probability denoted as $p_s(\bar{c})$. We wish to find the parameter vector $\bar{c}$ that optimizes the log likelihood of the data:

$$L(\vec{c}) = \sum_s n_s \log(p_s(\vec{c})) + \left(N_s - n_s\right)\log(1 - p_s(\vec{c})),$$

$$(C1)$$

where n_s denotes the number of "coherent" responses given by the subject over a total of N_s trials in which stimulus *s* was presented.

The model probabilities, $p_s(\vec{c})$, cannot be computed in closed form and must instead be computed through stochastic simulation. For this, we simulated 50 trials, generating new random measurements for each according to the noise parameters, then computing the probabilities of Equations 1 and 2, and comparing them to obtain an answer. Suppose that *k* of these simulated trials produced a "coherent" response. A maximum likelihood estimate of the model probability would be $\hat{p}_s(\vec{c}) = \arg\max_p[p^k(1 - p)^{(50-k)}] = k / 50$. However, this estimate is problematic when optimizing the log likelihood of Equation C1: If the simulated trials produce a value of $k = 0$ (or $k = 50$), the estimated model probability will be 0 (or 1), which can lead to an infinite log likelihood. To avoid this, we used the mean estimate, $\hat{p}_s(\vec{c}) = \int_0^1 p^k(1 - p)^{(50-k)}dp = (k + 1) / 52$, whenever the simulated trials produced $k = 0$ or $k = 50$.

Rather than performing a brute-force search over the entire six-dimensional parameter space, we performed a much more efficient and stable nested optimization. In the outermost loop, we searched over the three noise parameters (using "fminsearch," in Matlab (version R2010a)), which require the most substantial computational effort. For each setting of these three parameters, we simulated 50 trials of measurements for each grating in each stimulus condition, by drawing stochastic samples from the associated measurement distributions (Equation B1). For each of these measurements, we precomputed arrays containing values of the associated likelihood functions (Equation B3). Given these likelihood functions, we then optimized the two parameters of the speed prior. For each setting of these two parameters, we computed the prior probabilities and the integrals in Equations 1 and 2, reusing the precomputed likelihood functions (since they

depend only on the noise parameters). Finally, given settings of the noise and speed prior parameters, the values of the integrals were stored, and the last parameter, $p(H_{\text{coh}})$, was easily optimized because it appears only as a multiplicative scale factor in the last step of computation of the probabilities in Equations 1 and 2.

We used the same initial conditions for the six parameters for all subjects: $c_1 = 1$, $c_2 = 0.2$, $c_3 = 0.35$, $c_4 = 1$, $c_5 = 2.4$, $p(H_{\text{coh}}) = 0.5$. Integrals were computed numerically, over a rectangular region of the velocity plane covering $v_x \in [-16, 16]$ deg/s and $v_y \in [-12, 24]$ deg/s, sampled at increments of 0.5 deg/s. We verified that this plane was large enough, and of sufficiently fine spacing, to accurately fit the parameters. For the integration required to compute the likelihoods (Equation B3), we used 120 directional samples over an angle of π radians. We also confirmed that this sampling was sufficiently dense, so as not to significantly impact the behavior of the model.

Acknowledgments

References

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300,* 523–525.

Alhazen (2005). A philosophical perspective on Alhazen's optics. *Arabic Sciences and Philosophy, 15,* 189–218.

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science, 331,* 83–87.

Bruyn, B. D., & Orbán, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research, 28,* 1323–1335.

Cropper, S. J., Mullen, K. T., & Badcock, D. R. (1996). Motion coherence across different chromatic axes. *Vision Research, 36,* 2475–2488.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience, 2,* 740–745.

Dong, D., & Atick, J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems, 6,* 345–358.

Eckert, M., & Zeil, J. (2001). Towards an ecology of motion vision. In J. M. Zanker & J. Zeil (Eds.), *Motion vision: Computational, neural, and ecological constraints* (pp. 333–369). Berlin Heidelberg New York: Springer Verlag.

Farid, H., & Simoncelli, E. P. (1994). The perception of transparency in moving square-wave plaids. *Investigative Ophthalmology and Visual Science Supplement (ARVO), 35,* 1271.

Farid, H., Simoncelli, E. P., Bravo, M. J., & Schrater, P. R. (1995). Effect of contrast and period on perceived coherence of moving square-wave plaids. *Investigative Ophthalmology and Visual Science Supplement (ARVO), 36,* S-51.

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing, 9,* 301–315.

Fischer, B. (2010). Bayesian estimates from heterogeneous population codes. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–7). Barcelona.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Science, 14,* 119–130.

Helmholtz, H. (2000). *Treatise on physiological optics.* Bristol, UK: Thoemmes Press. (Original work published 1866)

Herrnstein, R. J., Rachlin, H., & Laibson, D. I. (1997). *The matching law: Papers in psychology and economics.* Cambridge, MA: Harvard Univ. Press.

Hildreth, E. C. (1984). *Measurement of visual motion.* Cambridge, MA: MIT Press.

Hoyer, P., & Hyvärinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15: Proceedings of the 2002 Conference* (p. 293). Cambridge, MA: The MIT Press.

Hupé, J.-M., & Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: A fresh look at plaids. *Vision Research, 43,* 531–548.

Hupé, J.-M., & Rubin, N. (2004). The oblique plaid effect. *Vision Research, 44,* 489–500.

Hürlimann, F., Kiper, D. C., & Carandini, M. (2002). Testing the Bayesian model of perceived speed. *Vision Research, 42,* 2253–2257.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9,* 690–696.

Jefferys, W., & Berger, J. (1991). Ockham's razor and Bayesian statistics. *American Scientist, 80,* 64–72.

Kim, J., & Wilson, H. R. (1993). Dependence of plaid motion coherence on component grating directions. *Vision Research, 33,* 2479–2489.

Kooi, F. L., Valois, K. K. D., Switkes, E., & Grosof, D. H. (1992). Higher-order factors influencing the perception of sliding and coherence of a plaid. *Perception, 21,* 583–598.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One, 2,* e943.

Krauskopf, J., & Farell, B. (1990). Influence of colour on the perception of coherent motion. *Nature, 348,* 328–331.

Krauskopf, J., Wu, H. J., & Farell, B. (1996). Coherence, cardinal directions and higher-order mechanisms. *Vision Research, 36,* 1235–1245.

Langley, K. (1999). Computational models of coherent and transparent plaid motion. *Vision Research, 39,* 87–108.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9,* 1432–1438.

Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (chap. 10, pp. 203–222). Cambridge, MA: MIT Press.

Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B: Biological Sciences, 211,* 151–180.

McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision Research, 21,* 491–500.

Montagnini, A., Mamassian, P., Perrinet, L., Castet, E., & Masson, G. S. (2007). Bayesian modeling of dynamic motion integration. *The Journal of Physiology, 101,* 64–77.

Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology, 98,* 1125–1139.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1986). *The analysis of moving visual patterns* (pp. 117–151). New York: Springer-Verlag.

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research, 25,* 625–660.

Hedges, Stocker, & Simoncelli

Natarajan, R., Murray, I., Shams, L., & Zemel, R. (2009). Characterizing response behavior in multi-sensory perception with conflicting cues. *Advances in Neural Information Processing Systems, 21,* 1153–1160.

Necker, L. A. (1832). Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. *The London and Edinburgh Philosophical Magazine and Journal of Science, 1,* 329–337.

Orbán, G. A., de Wolf, J., & Maes, H. (1984). Factors influencing velocity coding in the human visual system. *Vision Research, 24,* 33–39.

Roth, S., & Black, M. (2007). On the spatial statistics of optical flow. *International Journal of Computer Vision (IJCV), 74,* 33–50.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audio-visual stimuli. *Neural Computation, 19,* 3335–3355.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience, 16,* 1486–1510.

Simoncelli, E. P. (1993). *Distributed analysis and representation of visual motion.* Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

Simoncelli, E. P. (2009). Optimal estimation in sensory systems. In M. Gazzaniga (Ed.), *The cognitive neurosciences IV* (chap. 36, pp. 525–535). Cambridge, MA: MIT Press.

Simoncelli, E. P., & Heeger, D. J. (1992). A computational model for perception of two-dimensional pattern velocities. *Investigative Ophthalmology and Visual Science Supplement (ARVO), 33,* 954.

Smith, A. T. (1992). Coherence of plaids comprising components of disparate spatial frequencies. *Vision Research, 32,* 393–397.

Stocker, A. A. (2006). *Analog VLSI circuits for the perception of visual motion.* Chichester, UK: John Wiley & Sons. Number ISBN-13: 978-0-470-85491-4, 242 pp., hardcover.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9,* 578–585.

Stone, L. S., & Thompson, P. (1992). Human speed perception is contrast dependent. *Vision Research, 32,* 1535–1549.

Stoner, G. R., & Albright, T. D. (1992). Motion coherency rules are form-cue invariant. *Vision Research, 32,* 465–475.

Stoner, G. R., Albright, T. D., & Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature, 344,* 153–155.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research, 22,* 377–380.

Thompson, P., Brooks, K., & Hammett, S. T. (2006). Speed can go up as well as down at low contrast: Implications for models of motion perception. *Vision Research, 46,* 782–786.

Victor, J. D., & Conte, M. M. (1992). Coherence and transparency of moving plaids composed of Fourier and non-Fourier gratings. *Perception & Psychophysics, 52,* 403–414.

von Grünau, M., & Dubé, S. (1993). Ambiguous plaids: Switching between coherence and transparency. *Spatial Vision, 7,* 199–211.

Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychological Research, 20,* 325–380.

Weiss, Y. (1998). *Bayesian motion estimation and segmentation.* Ph.D. thesis, Massachusetts Institute of Technology.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5,* 598–604.

Welch, L. (1989). The perception of moving plaids reveals two motion-processing stages. *Nature, 337,* 734–736.

Welch, L., & Bowne, S. F. (1990). Coherence determines speed discrimination. *Perception, 19,* 425–435.

Wuerger, S., Shapley, R., & Rubin, N. (1996). 'On the visually perceived direction of motion' by Hans Wallach: 60 years later. *Perception, 25,* 1317–1367.

Yuille, A., & Bülthoff, H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge, England: Cambridge University Press.

Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature, 333,* 71–74.

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

Zhang, K., Ginzburg, I., McNaughton, B. L., & Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *Journal of Neurophysiology, 79,* 1017–1044.