

Unifying account of visual motion and position perception

Oh-Sang Kwon^{a,1}, Duje Tadin^{a,b,2}, and David C. Knill^{a,3}

^aCenter for Visual Science and Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627; and ^bDepartment of Ophthalmology, University of Rochester School of Medicine, Rochester, NY 14642

Edited by David J. Heeger, New York University, New York, NY, and approved May 11, 2015 (received for review January 11, 2015)

Despite growing evidence for perceptual interactions between motion and position, no unifying framework exists to account for these two key features of our visual experience. We show that percepts of both object position and motion derive from a common object-tracking system—a system that optimally integrates sensory signals with a realistic model of motion dynamics, effectively inferring their generative causes. The object-tracking model provides an excellent fit to both position and motion judgments in simple stimuli. With no changes in model parameters, the same model also accounts for subjects' novel illusory percepts in more complex moving stimuli. The resulting framework is characterized by a strong bidirectional coupling between position and motion estimates and provides a rational, unifying account of a number of motion and position phenomena that are currently thought to arise from independent mechanisms. This includes motion-induced shifts in perceived position, perceptual slow-speed biases, slowing of motions shown in visual periphery, and the well-known curveball illusion. These results reveal that motion perception cannot be isolated from position signals. Even in the simplest displays with no changes in object position, our perception is driven by the output of an object-tracking system that rationally infers different generative causes of motion signals. Taken together, we show that object tracking plays a fundamental role in perception of visual motion and position.

visual motion perception | Kalman filter | object tracking | causal inference | motion-induced position shift

Research into the basic mechanisms of visual motion processing has largely focused on simple cases in which motion signals are fixed in space and constant over time (e.g., moving patterns presented in static windows) (1). Although this approach has resulted in considerable advances in our understanding of low-level motion mechanisms, it leaves open the question of how the brain integrates changing motion and position signals; when objects move in the world, motion generally co-occurs with changes in object position. The process of generating coherent estimates of object motion and position is known in the engineering and computer vision literature as “tracking” (e.g., as used by the Global Positioning System) (2). Conceptualizing motion and position perception in the broader context of object tracking suggests an alternative conceptual framework—one that we show provides a unifying account for a number of perceptual phenomena.

An optimal tracking system would integrate incoming position and motion signals with predictive information from the recent past to continuously update perceptual estimates of both an object's position and its motion. Were such a system to underlie perception, position and motion should be perceptually coupled in predictable ways. Signatures of such a coupling appear in a number of known phenomena. On one hand, local motion signals can predictively bias position percepts (3–8). On the other hand, we can perceive motion solely from changes in object position (9–12). For example, motion can be perceived in stimuli with no directional motion signal by tracking position changes along a specific direction (10). These phenomena, however, are currently regarded as arising from independent mechanisms (11–14).

Given the interdependency of motion and position and the inherent noisiness of sensory signals, it is advantageous for vision to exploit the redundancy between motion and position signals and integrate them into coupled perceptual estimates. This is complicated by the fact that local motion signals can result from a combination of motions (of which object translations are only one) (15, 16). A flying, rotating soccer ball provides a prototypical example of this problem (Fig. 1A). Because the ball rotates as it flies through the air, the local retinal motion signals created by ball texture are sums of two world motions: translation and rotation of the ball. Relating local motion signals to object motion requires the solution of the “source attribution” problem (17, 18)—determining what part of a local retinal motion pattern is due to object translation and what part is due to object-relative motion of the texture pattern. To solve this attribution problem, the brain can exploit the redundant information provided by the changing stimulus position. Moreover, integrating motion and position information over time with an internal model of motion dynamics can mitigate both the uncertainty created by ubiquitous sensory noise (19) and that created by the motion source attribution problem. Although object-relative pattern motion is not a property of all moving objects, understanding how pattern motion interacts with object motion and position can help elucidate how the brain integrates motion and position signals into coherent perceptual estimates—a problem associated with all moving objects.

Here, we propose and test a computational framework in which motion and position perception derive from a common mechanism

Significance

In our visual environment, object motion is strongly correlated with changes in object position. However, there is no unifying computational framework that can account for both motion and position perception and their interactions. Here, we propose and test an object-tracking model that optimally integrates sensory signals with a realistic model of motion dynamics. The model accounts for several well-known visual illusions, including motion-induced position shifts, slow speed biases, and the curveball illusion. Moreover, the model also makes several novel and, in some cases, counterintuitive predictions about interdependencies between position and motion. In summary, we provide a unifying framework that reconceptualizes how the human visual system constructs coherent percepts from noisy position and motion signals.

Author contributions: O.-S.K., D.T., and D.C.K. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹Present address: Department of Human and Systems Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea 689-805.

²To whom correspondence should be addressed. Email: duje@cvs.rochester.edu.

³Deceased October 6, 2014.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1500361112/-DCSupplemental.

that integrates sensory signals over time to track objects and infer their generative causes. The consequence of this process is a strong, bidirectional coupling between motion and position perception that provides a unifying account for a range of perceptual phenomena. These include motion-induced shifts in perceived position (3–6), perceptual speed biases (20), slowing of motions shown in visual periphery (21, 22), and the curveball illusion (16). The presented model also makes novel predictions about interactions between position and motion perception—predictions confirmed here. Importantly, we do not fit the model to each experiment, but fit the parameters to data from experiment 1 and show that the resulting model accurately predicts subjects' performance in qualitatively different and more complex tasks (experiments 2 and 3).

Results

The Tracking Model. To understand the computational principles that underlie human object tracking, we modeled it using a Bayesian system that optimally inverts a generative model of sensory signals to infer the causes of motion and position signals in the world (Fig. 1 *A* and *B*). The model is fully specified by the generative model that it inverts—a statistical model of world motions and corresponding sensory measurements. The world model includes two sources of motion (Fig. 1*A*): (*i*) translation of an object through a scene and (*ii*) motion of the texture within the object, which we will refer to as “pattern motion.” Because motions in the world generally vary smoothly over time, both object and pattern velocities are correlated over time. We modeled this by assuming that both velocities follow a stationary Gaussian process (discrete Ornstein–Uhlenbeck process; Eq. 1 and *Supporting Information*). The observer is assumed to have noisy sensory measurements of (*i*) object position, (*ii*) retinal object velocity, and (*iii*) retinal velocity of the texture pattern within the object boundaries (which is equal to the sum of object and pattern velocities) (Fig. 1*A*). A simple stochastic model that embodies these constraints (Fig. 1*B*) is given in discrete form by the time update equations as follows:

$$\begin{array}{ll}
 \text{Model of motions in the world} & \text{Observation model} \\
 x_t = x_{t-1} + \Delta t \cdot v_{t-1}^{obj} & y_t^x = x_t + \eta^x \Omega_t^{yx} \\
 v_t^{obj} = \alpha v_{t-1}^{obj} + \sigma^{v-obj} \Omega_t^{vo}; & 0 \leq \alpha < 1 \\
 v_t^{pattern} = \beta v_{t-1}^{pattern} + \sigma^{v-pattern} \Omega_t^{vp} & 0 \leq \beta < 1 \\
 & y_t^{v-obj} = v_t^{obj} + \eta^{v-obj} \Omega_t^{vo} \\
 & y_t^{v-pattern} = v_t^{obj} + v_t^{pattern} + \eta^{v-pattern} \Omega_t^{vp}
 \end{array} \quad [1]$$

The first term in the motion model simply updates object position by integrating object velocity over time. The other two terms represent stationary Gaussian processes that characterize object and pattern velocities. α and β determine velocity correlations over time (see *Discussion* for more details). σ^{v-obj} and $\sigma^{v-pattern}$ specify SDs of changes in object and object-centered pattern velocities (Ω_t represents unit variance, Gaussian white noise). These parameters modulate how strongly the model weights new sensory signals vs. predictions derived from previous signals and the world motion model. y_t^x , y_t^{v-obj} , and $y_t^{v-pattern}$ represent sensory measurements of retinal stimulus position on the retina, retinal object velocity, and retinal velocity of the texture pattern within the object's boundary. These are corrupted by Gaussian noise with SDs given by η^x , η^{v-obj} , and $\eta^{v-pattern}$. Because we assume the sensory noise to be Gaussian, the optimal tracker, which continuously updates its estimates of the objects' state variables (x_t , v_t^{obj} , and $v_t^{pattern}$) from the noisy sensory measurements (y_t^x , y_t^{v-obj} , and $y_t^{v-pattern}$), is known to be a Kalman filter (19, 23).

Here, the Kalman filter addresses two estimation problems. First, to optimally track current object states (e.g., position), it

provides optimal weights on the current and past sensory signals. When sensory signal uncertainty is low, estimates of objects' states are largely determined by the current signals, showing little dependence on past signals. Simply stated, sensory history is of less use when the current sensory signal is accurate. However, when signal uncertainty is high, the estimates of objects' states are strongly influenced by past signals. In this case, integrating over past signals reduces the effects of sensory noise (19, 23). Second, the Kalman filter optimally attributes the measured retinal velocity of texture pattern to (*i*) object motion and (*ii*) pattern motion (Fig. 1*A*). The optimal solution of the attribution problem again depends on signal uncertainty. When positional uncertainty is low, the system can accurately attribute the retinal texture motion to the actual source(s) of the motion signal. However, when positional uncertainty is high, pattern motion and object motion should strongly interact (as detailed next).

To illustrate the model's behavior, we simulated it on a standard motion perception stimulus: a translating texture pattern presented within a stationary contrast envelope. This stimulus is also used to demonstrate striking perceptual effects of motion on perceived position: such stimuli appear as shifted in the direction of the pattern motion [motion-induced position shift (MIPS); *Movie S1*] (3–6). The model's estimates evolve over time from initial stimulus onset but asymptote to stable values in ~ 200 ms (Fig. S1; also see Fig. 3). The model distributes responsibility for the measured texture motion (Fig. 1*A*) to a combination of object and pattern motion. Notably, the model attributes more of the measured texture motion to the object motion when position uncertainty is high (Fig. 1*C, Right*)—a reasonable outcome given the predominance of object motions in the natural world. Consequently, this leads to estimates of object position that are gradually shifted in the direction of pattern motion as position uncertainty increases, i.e., the well-known MIPS phenomenon. This model also provides a novel account of perceptual slowing of pattern motion. In stimuli with weakly specified boundaries, most of the perceptual

slowing results from attributing some of the pattern motion to object motion (Fig. 1*C, Right*) rather than from a stationary slow speed prior (20) (see *Discussion* for more details).

These observations lead to several testable predictions. First, because MIPS magnitude depends on the degree to which the model assigns responsibility for the texture motion to the object itself, any change in the stimulus that increases MIPS magnitude should also decrease the perceived pattern speed. That is, MIPS magnitude should be negatively correlated with perceived pattern speed (Fig. 1*C*; prediction tested in experiment 1). Second, the model predicts a counterintuitive inconsistency between perceived position and motion speed. The model's estimate of object speed remains nonzero even after the object position estimate has stabilized (prediction demonstrated and tested in experiment 2). This violation of classical physics derives from the iterative structure of the model. The tracker uses its estimates of object motion and position at time $t - 1$ to predict the object position at time t . It then integrates the predictive estimate with new sensory position signals measured at time t to derive the new position estimate. When the tracker's speed estimate is nonzero, the predicted

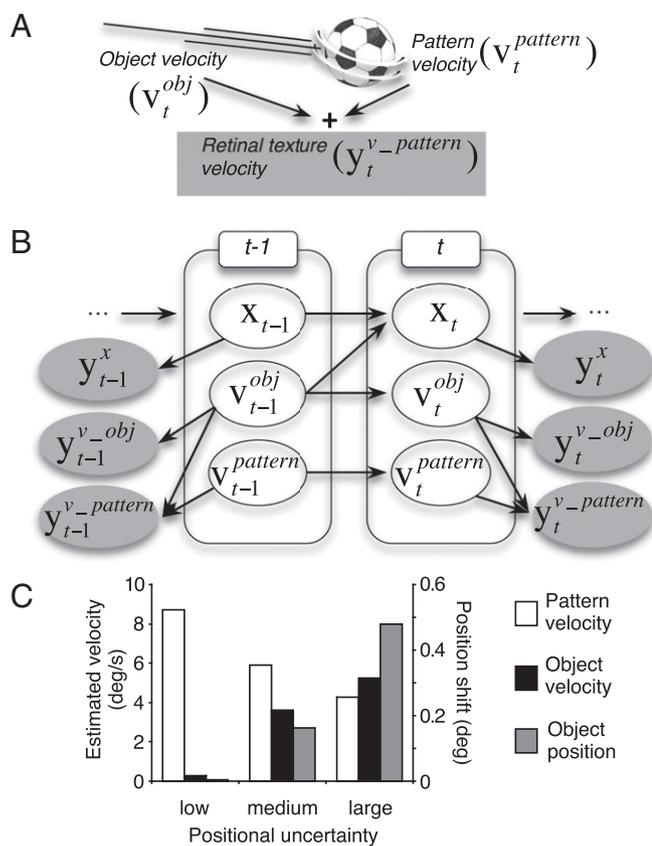


Fig. 1. Schematic illustration of the object-tracking model and its behavior. (A) An example of an object with both object boundary motion and pattern motion. (B) A generative model of the Bayesian observer. White nodes indicate hidden variables and gray nodes indicate observable variables that are noisy measurements of the connected hidden variables. Arrows indicate causal links. (C) Model behavior for a typical MIPS stimulus containing a moving pattern within a static envelope. The steady-state estimates of the three object states (position, object velocity, and pattern velocity) are plotted for different positional uncertainties. At low positional uncertainty, most of the retinal texture motion is correctly attributed to the pattern motion. Consequently, illusory object motion and MIPS are negligible. At high positional uncertainty, much of the texture motion is attributed to object motion (reflecting a prior that object motion is more likely than pattern motion). This results in relatively low estimated pattern velocity and large MIPS.

position is shifted in the direction of motion, but is then “pulled back” by the sensory position signal. The position estimate stabilizes when the two effects exactly balance each other.

To provide a strong test of the model, we used model parameters fit to the subjects’ perceptual judgments for simple MIPS stimuli (experiment 1) to predict perceptual judgments for the more dynamic stimuli where either stimulus position changed over time (experiment 2), the pattern motion changed over time (experiment 3; see Fig. 4) or both types of signals changed (see Fig. 5).

Experiment 1: The Interdependency of Motion-Induced Position Shifts and Speed Perception.

To test the prediction that the MIPS magnitude should be negatively correlated with perceived pattern speed, we measured psychometric functions for subjects’ judgments of position and pattern speed in conventional MIPS stimuli. Test stimuli contained translational motion moving either up or down (10°/s) within a stationary envelope. We manipulated the uncertainty of contour position by (i) using either sharp or fuzzy contrast envelopes (pillbox or Gaussian) and (ii) presenting stimuli at different eccentricities (5°, 10°, 15°) (Fig. 2 A–C). Subjects’ position biases (Fig. 2D) qualitatively replicate previous results (3)—

MIPS magnitude increased with eccentricity. It was also greater for Gaussian envelopes than for pillbox envelopes. In contrast, perceived pattern speed was lower for Gaussian envelopes and decreased with eccentricity (Fig. 2E). Moreover, there was a strong negative correlation between MIPS magnitude and perceived pattern speed in all subjects (Fig. 2F; all $r > 0.89$, all $P < 0.016$). The best-fitting tracking model (see Supporting Information for model fitting) exhibited a highly similar pattern of perceptual biases (Fig. 2 D–F; blue data points) and also provided a good fit to subjects’ psychometric functions (Fig. S2). More importantly, we will use model parameters from this experiment to directly predict perception for more complex stimuli in experiments 2 and 3. As a notable aside, the tracker model provides a novel, computationally grounded account of the well-known phenomenon of peripheral slowing (prominent in Fig. 2E), in which motion stimuli appear slower in the periphery than near the fovea (21, 22) (Movie S2).

Experiment 2: A Conflict Between Estimated Object Speed and Changes in Object Position.

Velocity is classically defined as a change in position over time. It seems intuitive, therefore, that a tracker’s estimates of object velocity should match the rate at which its estimate of object position changes over time. As noted above, a rational tracker using noisy sensory signals violates this simple physical intuition; when simulated on the MIPS stimulus, even when the tracker’s object position estimate stabilizes (i.e., the derivative of position estimates is 0), its estimate of object speed is nonzero.

Using the MIPS stimulus to test this prediction is problematic because to measure perceived object speed a moving reference stimulus should be used. This creates a matching ambiguity—should subjects match perceived object speed or changes in object

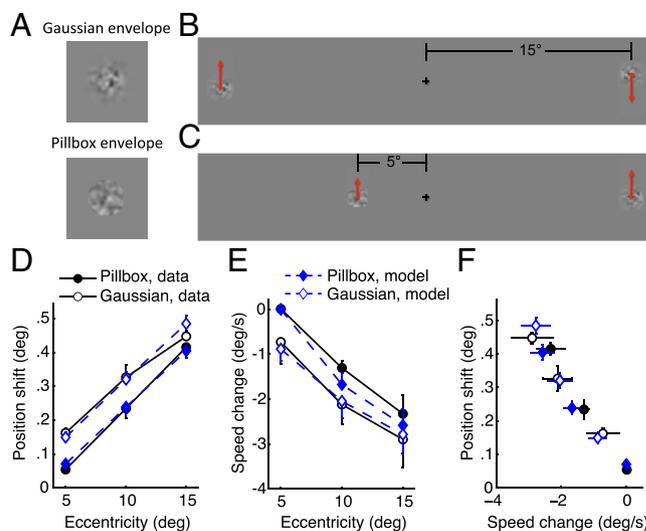


Fig. 2. Experiment 1 stimuli and results. (A) Two types of spatial envelopes. (B) In the position task, subjects judged whether the right stimulus was above or below the left stimulus. For the two stimuli to be perceived at the same vertical locations, the physical location of the downward moving stimulus has to be higher than the position of upward moving stimulus (as depicted). (C) In the speed task, subjects judged whether the test stimulus pattern motion was faster or slower than the reference speed. For the two pattern speeds to be perceived as equal, the physical speed of the stimulus in the far periphery has to be faster than the reference stimulus (as depicted). (D) The position shift increases with increasing eccentricity [$F_{(2,8)} = 147.3$, $P < 10^{-15}$] or when the envelope boundary is blurred [$F_{(1,4)} = 68.3$, $P = 0.001$]. (E) In contrast, the perceived speed decreases with increasing eccentricity [$F_{(2,8)} = 17.1$, $P = 0.001$] or the envelope boundary is blurred [$F_{(1,4)} = 10.4$, $P = 0.032$]. (F) The strong negative relationship between MIPS magnitude and changes in the perceived speed. The best-fitting model (blue symbols and lines in D, E, and F) closely fits the data (see Fig. S2 for models fits to the entire dataset).

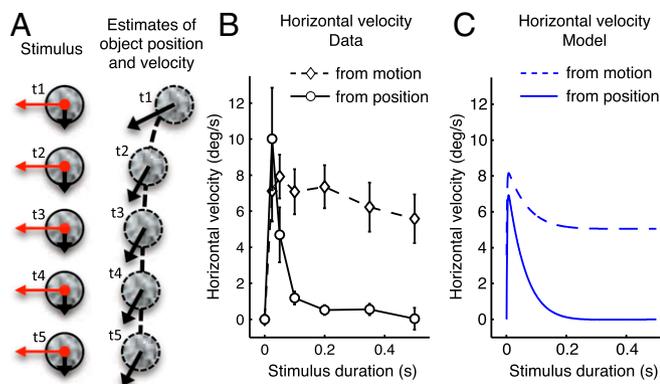


Fig. 3. Experiment 2 stimuli and results. (A) Left panel is a schematic illustration of the curveball illusion stimulus. Thin red line represents pattern motion, and bold black line represents object motion. Right panel shows the temporal dynamics of the model's estimates of object position and velocity (bold black line). (B) Subjects' perceived horizontal object velocity as indicated by two different estimates: by object motion direction (dashed line) and by the change in object position over time (full line). The results reveal a large conflict between the perceived horizontal velocity and the change in perceived position. (C) Predictions of the object-tracking model closely matches subjects' estimates of perceived horizontal velocity. The model predictions are derived from subjects' data in experiment 1 (i.e., the model has no free parameters).

position? To avoid this ambiguity, we adapted the well-known “curveball illusion” in which a translating object containing a moving pattern (as in a spinning baseball) appears to follow a curved trajectory (16) (Movie S3). The experimental stimulus was shown in visual periphery and contained a pattern moving horizontally and embedded in a circular envelope, the combination of which moved directly downward at a constant speed (Fig. 3A). Despite its straight motion path, this stimulus is perceived to follow a curved trajectory. It initially appears to move obliquely in a direction given by the local retinal motion but ultimately curves back closer to vertical (although remaining somewhat oblique). The tracking model accounts for the curveball illusion in much the same way as it accounts for MIPS. The downward physical motion of the stimulus trivially generates the downward component of object velocity. The horizontal pattern motion induces a horizontal shift in perceived position that grows over time but quickly stabilizes. At the same time, the model attributes some of the horizontal pattern motion to the object motion, so that the asymptotic estimate of object velocity has a nonzero horizontal component, giving it the appearance of moving obliquely. This leads to a counterintuitive prediction—subjects should see the object moving obliquely even after the object's horizontal position stops changing. This occurs because the model stabilizes at a point where the attractive pull of the sensory position signal is balanced by the horizontal position shift predicted by the model's estimate of horizontal speed.

To test this prediction, we measured subjects' final percepts of object position and motion direction for stimuli with varied durations (25–500 ms). Fig. 3B contrasts subjects' estimates of horizontal object speed derived either from (i) their judgments of motion direction or (ii) their position estimates. As predicted, the perceived horizontal component of object motion asymptotated at a positive nonzero value (i.e., subjects continued to perceive the ball moving obliquely), whereas subjects' estimates of horizontal position stabilized after ~250 ms. That is, there is a sizeable inconsistency between object speed derived from motion direction and object speed derived from changes in position. This same inconsistency is found in the behavior of the tracking model derived from experiment 1. Despite having no free parameters, the model matches subjects' perceptual judgments remarkably well (Fig. 3C).

Experiment 3: Object Tracking vs. Weighted Vector Sum. The existing explanation of the curveball illusion assumes that pattern and object motion are erroneously integrated in visual periphery (16, 24). Thus, the oblique object trajectory (Fig. 3A) is a result of a weighted vector sum of pattern and object motion signals. Indeed, the oblique object velocity at a given stimulus duration (Fig. 3B) can be predicted by the weighted vector sum as long as the perceived object motion direction is between pattern and object motion directions. The tracking model, however, produces both the oblique object motion and matches its temporal dynamics (Fig. 3C) by relying on a general process that decomposes motion signals into their physical causes.

To further differentiate these two frameworks, we developed stimuli where the predictions of the tracking model are qualitatively different from the weighted vector sum. Specifically, within a static object envelope, we showed pattern motion that changed direction at a continuous rate (rotating around the clock face; Fig. 4A, Top). Both models predict constant-speed, circular trajectory of object motion, which is consistent with human percepts (Movies S4 and S5). However, the two models give different predictions about the phase difference between the

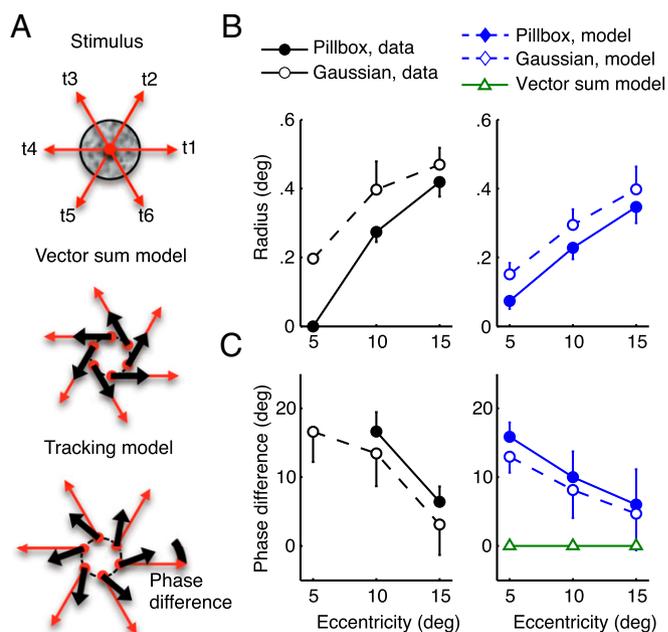


Fig. 4. Experiment 3 stimuli and results. (A) Schematic illustration of the illusory rotation stimulus and models' estimates of object motion. The upper image is the actual stimulus—a static envelope with a translating pattern motion whose direction changes (i.e., rotates) at 1 Hz (Movies S4 and S5). Red arrows show changes in pattern motion direction. The vector sum model (middle image) predicts that the estimated object motion direction (black arrows) should be identical to the actual pattern motion direction (red arrows). The tracking model (bottom image) predicts that the estimated object motion direction should lead the physical pattern motion direction. In both models, object trajectory converges to a circular path. (B) Psychophysical data (Left) and object-tracking model predictions (Right) for the radius of the illusory object motion. The perceived motion radius increases with increasing eccentricity [$F_{(2,8)} = 26.8, P = 0.0003$] or when the boundary of envelope is blurred [$F_{(1,4)} = 16.1, P = 0.016$]. The same pattern is exhibited by the object-tracking model (as fitted in experiment 1). (C) Psychophysical data (Left) and object-tracking model predictions (Right) for the phase difference between the perceived object motion and the actual pattern motion direction. The phase difference decreases as the eccentricity increases [$F_{(1,4)} = 18.1, P = 0.013$], with no significant effects of the boundary type [$F_{(1,4)} = 0.77, P = 0.43$]. The object-tracking model closely matches the human data. We could not measure the phase difference for the hard boundary condition at 5° eccentricity because observers did not perceive illusory object motion rotation [the model predicts a very small rotation radius (0.07°), which is likely below the perceptual threshold].

physical pattern motion direction and the estimated object motion direction (Fig. 4A, *Middle* and *Bottom*). In the weighted vector sum model, the estimated object motion direction is identical to the pattern motion direction. That is because the rotating pattern motion is the only physical motion stimulus, and thus solely determines the object motion direction. In contrast, the tracking model predicts that the perceived object direction leads ahead of pattern motion, and the size of this phase lead depends on the eccentricity and boundary conditions (Fig. 4C, *Right*; all model parameters unchanged from experiment 1). This is because the model's estimate of object motion is determined not only by the current pattern motion but also by the forward prediction from previous state estimates.

To test the model predictions, we presented the same stimuli used in experiment 1 to the same subjects, but with the direction of motion within the static envelope changing at 1 Hz (Fig. 4A, *Top*). Subjects adjusted the radius and phase of a circularly moving comparison disk presented at the same eccentricity in the opposite hemifield to match the perceived motion of the test stimulus. The radius of perceived illusory object rotation gradually increased with increasing eccentricity (Fig. 4B, *Left*) and was also larger for soft than for sharp boundaries. More importantly, we found a significant phase difference between comparison and test stimuli, magnitude of which decreased with eccentricity (Fig. 4C, *Left*). The tracking model, without any free parameters, reproduces the same pattern of results (Fig. 4B and C, *Right*). If the basic vector sum model is amended with three MIPS-related parameters (two for the eccentricity dependency; one for the envelope effect), it can also account for the perceived rotation radius ($R^2 = 0.93$). However, even with those extra parameters, the vector sum model cannot predict the observed phase lag (Fig. 4C, *Left*).

To further test the predictive ability of the tracking model, we modified the rotating pattern motion stimulus by rotating the entire object in the direction opposite to the pattern motion (Fig. 5A). The tracking model predicts that these two motions combine to create nontrivial motion percepts that depend on the stimulus eccentricity. When viewed near foveally, the model's estimate of the object's trajectory is like a soft-cornered diamond with the rotation direction matching the actual object motion (Fig. 5B). However, as the stimulus is moved further into the visual periphery, the model's estimate of the object trajectory transitions through a near-linear motion to a near-elliptical motion, now rotating in the opposite direction from the actual object motion. The reader can test this illusion predicted by the model by viewing [Movie S6](#).

Discussion

We present an object-tracking model that provides a unifying account of both motion and position perception and, significantly, of interactions between these two key visual features. The strongest evidence for the object-tracking framework is that a model fit to

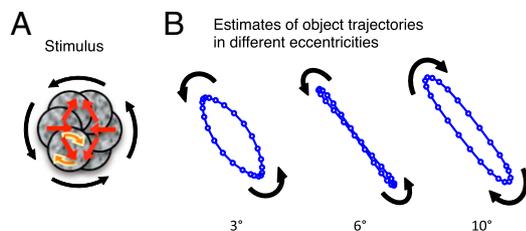


Fig. 5. A novel illusion predicted by the tracking model. (A) We added circular object motion (black arrows) to a stimulus whose pattern motion (red arrows) rotates in the opposite direction (orange arrows). (B) The tracking model predicts that the percept follows the object motion direction at near eccentricities and pattern motion direction at far eccentricities, i.e., the direction of perceived object rotation changes. Notably, at an intermediate eccentricity, the object is predicted to oscillate along a nearly linear trajectory. This prediction can be confirmed by viewing [Movie S6](#).

one aspect of perception (a conventional MIPS stimulus) easily generalizes to predict perception in qualitatively different tasks (more complex stimuli in experiments 2 and 3).

Although it is well accepted that motion influences position percepts (e.g., MIPS), computational models of underlying mechanisms have been lacking. The tracking model explains how sensory position and motion signals are integrated, an account that predicts quantitative features of MIPS. The tracking model also makes novel predictions about interactions between position and speed perception. These predictions, confirmed in experiments 1 and 2, run counter to a naïve model where subjects' percepts of object position and speed should maintain internal consistency. Namely, once one's percept of object position stabilizes, no motion should be attributed to the object (i.e., detected motion should be attributed to pattern motion within the object). Moreover, the perceived object speed should be zero, and the pattern speed should be independent of the MIPS magnitude. Instead, subjects' percepts of object motion remain biased in the direction of pattern motion even after their percepts of object position stabilize (Fig. 3) and their percepts of pattern speed decrease with increasing position biases (Fig. 2).

What is perhaps most surprising in the results is the strong influence of position signals on the perceived motion speed—a result that can be considered a mirror image of the MIPS effects. Notably, this indicates that motion perception cannot be isolated from position signals. Even in the simple displays with no changes in object position, we show perception is driven by the output of an object-tracking system.

The tracking model extends the notion of a prior toward slow speeds (20, 25) into the temporal domain. Two properties of the model contribute to slow-speed biases in perception. First, the model distributes responsibility for sensory motion signals to object and pattern motion, such that noisy position signals lead the model to attribute more of the motion signal to object motion (Fig. 1C). This results in slower pattern speed for blurred and peripherally presented stimuli (Fig. 2E). Second, for the motion model to be stationary (i.e., to reflect the fact that the distribution of velocities in the environment does not diverge over time), the parameters α and β must be less than 1. The result is a prior on motions that, on average, tend to decelerate over time (objects may accelerate or decelerate, but, on average, they eventually slow down). This slowing prior predicts perceptual slowing that increases with increasing motion uncertainty—conditions normally used to estimate the conventional slow speed prior (20, 25).

The model provides a rational account of how the visual system estimates both pattern and object motion. We addressed this question by investigating distortions in object motion percepts. These distortions are conventionally explained as the result of processes that integrate local and global motion signals (pattern and object motion, respectively). For example, accounts of the curveball illusion reason that local and global motion signals are integrated in the periphery, but not in the fovea, where the visual system can segregate conflicting motion signals (16). In the current model, the transition between integration and segregation depends not on eccentricity per se, but on a process that makes rational inferences about the causes of motion in the world—a process affected by both the uncertainty of motion signals and the uncertainty of object position. When object position signals are uncertain, the model naturally attributes more of the pattern motion to object motion. This bias is consistent with the higher prevalence of object motions in the world. When object position signals are reliable, the model resolves conflicts between position and motion signals by attributing more of the motion to object-relative pattern motion. The former appears as “integration” and the latter as “segregation” (26). Thus, the current model can be taken as a computational formulation of the transition between motion integration and segregation. One factor not considered by the model is whether these computations are taking place in retinotopic

or spatiotopic coordinates (27). For simplicity, we assumed that all computations are in a retinal reference frame. Future work will need to determine, for example, whether the tracking system state is maintained in retinotopic and/or spatiotopic reference frame.

The tracking model encompasses what researchers have alternately labeled attention-based motion (9, 12) or third-order motion (10, 11) systems. These mechanisms are thought to compute motion by tracking features, object boundaries, or salience over time. This higher-level motion system is often framed as a parallel motion pathway; however, the current model suggests a somewhat different conceptualization. This motion system is the integration locus of motion signals derived from low-level motion systems with position signals for object tracking. It only appears as a distinct system when studied with stimuli that do not engage low-level motion mechanisms (11, 28) or stimuli where motion direction is ambiguous (9, 12). In the current model, such stimuli would generate uninformative motion signals, but informative position signals. Consequently, the model infers object motion from position changes. Several lines of evidence, including lesion (29) and functional MRI studies (30), indicate that the inferior parietal lobule (IPL) is important for the perception of motion stimuli that do not activate low-level motion detectors. Thus, IPL may be critical for integrating sensory signals over time to track objects, although interactions between motion and position may also occur at earlier stages (6, 8).

In summary, we present and test a theoretical framework that provides a unifying account of our ability to perceive both object motion and object position. By conceptualizing these two important aspects of our perceptual experience in the context of object tracking, we show that the resulting model can explain a number of seemingly independent phenomena. At a computational level of description, the tracking model fits into the general class of models that implement causal inference. In perception, the causal inference framework explicitly recognizes that sensory signals come from multiple generative causes and that our percepts are shaped by processes that infer the generative causes of those signals. Although this approach has proven particularly fruitful in explaining nonlinear features of sensory cue integration (17, 18)

and sensorimotor adaptation (31), it applies to numerous aspects of visual perception that require distribution of responsibility for sensory signals across different generative causes. Lightness perception, for example, reflects the output of a process that takes into account the contributions of surface shape and illumination on local luminance gradients (32). The mechanisms by which the brain solves these problems, here modeled using a Kalman filter, are central to how humans integrate (or segregate) sensory signals to perceive the world.

Materials and Methods

Extended materials and methods may be found in *SI 2: Materials and Methods*.

Ten subjects participated in this study. The University of Rochester Research Subjects Review Board approved all procedures. Written informed consent was obtained from subjects. All experiments used the same base stimuli: a moving, low-pass-filtered noise pattern shown in a circular contrast envelope. Unless noted, stimulus speed was 10°/s. A 2° fixation window was enforced with an eye tracker. In Experiment 1 position task, subjects viewed pairs of stimuli presented along the horizontal midline at one of three eccentricities (5°, 10°, 15°) with pattern motion in opposite vertical directions and discriminated the vertical stimulus position (Fig. 2B). In Experiment 1 speed task, a test stimulus in a stationary envelope (pillbox or Gaussian) was presented at one of three horizontal eccentricities (5°, 10°, 15°). A reference pillbox stimulus was displayed in the opposite hemifield at 5° eccentricity (Fig. 2C; this stimulus leads to a minimal MIPS). Subjects indicated which stimulus contained faster motion. For both tasks, a method of constant stimuli was used (Fig. S2). In Experiment 2, the test stimulus (Fig. 3A) was a Gaussian envelope at 10° horizontal eccentricity that moved downward (4°/s) while the stimulus pattern moved horizontally (10°/s). Subjects judged either the direction of the "object" motion (direction task) or its horizontal position (position task) at the end of the stimulus presentation. Results were fit with cumulative Gaussian functions, from which we estimated subjects' perceptual biases. In Experiment 3, a test stimulus (Fig. 4A) and a comparison disk ($r = 0.65^\circ$) were presented at the same horizontal eccentricity (5°, 10°, or 15°) in opposite sides of visual field (Movie S5). Subjects adjusted both the radius and the phase of the comparison stimulus to match the perceived object motion of the test stimulus.

ACKNOWLEDGMENTS. We thank Greg DeAngelis, Ralf Haefner, Lorella Battelli, Alan Johnston, and Robbie Jacobs for manuscript comments. This work was supported by NIH Grants R01 EY019295 (to D.T.) and P30 EY001319. We dedicate this paper to the memory of David C. Knill, whose leadership, insight, and experimental rigor made this project possible.

- Burr D, Thompson P (2011) Motion psychophysics: 1985–2010. *Vision Res* 51(13):1431–1456.
- Challa S, Moreland MR, Musicki D, Evans RJ (2011) *Fundamentals of Object Tracking* (Cambridge Univ Press, Cambridge, UK).
- De Valois RL, De Valois KK (1991) Vernier acuity with stationary moving Gabors. *Vision Res* 31(9):1619–1626.
- Ramachandran VS, Anstis SM (1990) Illusory displacement of equiluminous kinetic edges. *Perception* 19(5):611–616.
- Whitney D (2002) The influence of visual motion on perceived position. *Trends Cogn Sci* 6(5):211–216.
- Maus GW, Fischer J, Whitney D (2013) Motion-dependent representation of space in area MT+. *Neuron* 78(3):554–562.
- Nishida S, Johnston A (1999) Influence of motion signals on the perceived position of spatial pattern. *Nature* 397(6720):610–612.
- McGraw PV, Walsh V, Barrett BT (2004) Motion-sensitive neurons in V5/MT modulate perceived spatial position. *Curr Biol* 14(12):1090–1093.
- Cavanagh P (1992) Attention-based motion perception. *Science* 257(5076):1563–1565.
- Lu Z-L, Sperling G (1995) Attention-generated apparent motion. *Nature* 377(6546):237–239.
- Lu Z-L, Sperling G (2001) Three-systems theory of human visual motion perception: Review and update. *J Opt Soc Am A Opt Image Sci Vis* 18(9):2331–2370.
- Verstraten FA, Cavanagh P, Labianca AT (2000) Limits of attentive tracking reveal temporal properties of attention. *Vision Res* 40(26):3651–3664.
- Arnold DH, Thompson M, Johnston A (2007) Motion and position coding. *Vision Res* 47(18):2403–2410.
- Fu Y-X, Shen Y, Gao H, Dan Y (2004) Asymmetry in visual cortical circuits underlying motion-induced perceptual mislocalization. *J Neurosci* 24(9):2165–2171.
- Hedges JH, et al. (2011) Dissociation of neuronal and psychophysical responses to local and global motion. *Curr Biol* 21(23):2023–2028.
- Shapiro A, Lu Z-L, Huang C-B, Knight E, Ennis R (2010) Transitions between central and peripheral vision create spatial/temporal distortions: A hypothesis concerning the perceived break of the curveball. *PLoS One* 5(10):e13296.
- Körding KP, et al. (2007) Causal inference in multisensory perception. *PLoS One* 2(9):e943.
- Shams L, Beierholm UR (2010) Causal inference in perception. *Trends Cogn Sci* 14(9):425–432.
- Faisal AA, Selen LP, Wolpert DM (2008) Noise in the nervous system. *Nat Rev Neurosci* 9(4):292–303.
- Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci* 9(4):578–585.
- Lichtenstein M (1963) Spatio-temporal factors in cessation of smooth apparent motion. *J Opt Soc Am A* 53(2):302–306.
- Tynan PD, Sekuler R (1982) Motion processing in peripheral vision: Reaction time and perceived velocity. *Vision Res* 22(1):61–68.
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng Trans ASME* 82(1):35–45.
- Tse PU, Hsieh PJJ (2006) The infinite regress illusion reveals faulty integration of local and global motion signals. *Vision Res* 46(22):3881–3885.
- Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. *Nat Neurosci* 5(6):598–604.
- Braddick O (1993) Segmentation versus integration in visual motion processing. *Trends Neurosci* 16(7):263–268.
- Turi M, Burr D (2012) Spatiotopic perceptual maps in humans: Evidence from motion adaptation. *Proc Biol Sci* 279(1740):3091–3097.
- Lu Z-L, Lesmes LA, Sperling G (1999) The mechanism of isoluminant chromatic motion perception. *Proc Natl Acad Sci USA* 96(14):8289–8294.
- Battelli L, et al. (2001) Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron* 32(6):985–995.
- Claeys KG, Lindsey DT, De Schutter E, Orban GA (2003) A higher order motion region in human inferior parietal lobule: Evidence from fMRI. *Neuron* 40(3):631–642.
- Berniker M, Körding K (2008) Estimating the sources of motor errors for adaptation and generalization. *Nat Neurosci* 11(12):1454–1461.
- Knill DC, Kersten D (1991) Apparent surface curvature affects lightness perception. *Nature* 351(6323):228–230.