Technical Report #74, Rutgers University Center for Cognitive Science

Individuation of Visual Objects over Time

Jacob Feldman and Patrice D. Tremoulet
Dept. of Psychology, Center for Cognitive Science, Rutgers University

How does an observer decide that a particular object viewed at one time is actually the *same* object as one viewed at a different time? We explored this question using an experimental task in which an observer views two objects as they simultaneously approach an occluder, disappear behind the occluder, and re-emerge from behind the occluder, having switched paths. In this situation the observer either sees both objects continue straight behind the occluder (called "streaming") or sees them collide with each other and switch directions ("bouncing"). This task has been studied in the literature on motion perception, where interest has centered on manipulating spatiotemporal aspects of the motion paths (e.g. velocity, acceleration). Here we instead focus on *featural* properties (size, luminance, and shape) of the objects. We studied the way degrees and types of featural dissimilarity between the two objects influence the percept of bouncing vs. streaming. When there is no featural difference, the preference for straight motion paths dominates, and streaming is usually seen. But when featural differences increase, the preponderance of bounce responses increases. That is, subjects prefer the motion trajectory in which each continuously existing individual object trajectory contains minimal featural change. Under this model, the data reveal in detail exactly what magnitudes of each type of featural change subjects implicitly regard as reasonably consistent with a continuously existing object. This suggests a neat mathematical definition of "individual object:" an object is a path through feature-trajectory space that minimizes feature change, or, more succinctly, an object is a *geodesic in Mahalanobis feature space*.

Objects

An important component of our perception of a stable and unified world is the subjective impression of *coherent objects* having continuous existence over time. Yet the full psychological meaning of the term "object" in this context remains elusive. What causes an object at one time to be regarded as the "same object" as another at a previous time, and what does "same" mean in this connection? This problem has sometimes been referred to as *temporal grouping* (Gepshtein & Kubovy, 2000) (as contrasted with *spatial grouping*, in which elements within a given visual image are aggregated together). In this paper we will use the term *object individuation*, to emphasize the mental construction of individual phenomenal objects having continuous existence.[1]

Pioneering research in the study of the object concept has come from the developmental literature (Baillargeon, 1994; Spelke, 1990). Infants understand objects to be bounded and coherent three-dimensional entities (Spelke, Breinlinger, Macomber, & Jacobson, 1992), and as young as four months of age believe that objects continue to exist when they disappear behind occluders (Baillargeon, 1987). Thus over time infants develop something like the adult's conception of objects, including expectations of boundedness and coherence, continuity of existence over time, and stability of featural attributes. Yet the exact meaning of many of these terms in the adult's conception is still somewhat unclear; the relevant questions in adults have scarcely been studied. Adults presumably have "object constancy" in the sense in which the term is usually used; but exactly what does this mean? What is held subjectively constant over the course of an object's existence? In this paper we study the problem of object identification in adult observers, and attempt to shed light on true psychological meaning of the term "object" and the computations underlying it.

We focus on the notion of featural stability, and on how expectations about the stability of objects' features over time influences observers' object interpretations in ambiguous situations. Clearly, one expects objects generally to retain their properties over the course of time. Yet it is obvious that an object can change its properties to some degree and yet re-

This research was supported by NSF SBR-9875175 to J.F. and NIH MH 19975-03 to P.D.T. We are grateful to two anonymous reviewers for helpful comments on an earlier version of this manuscript, to Randy Gallistel and Whitman Richards for helpful discussions, and to Elan Barenholtz, Daniel Drucker, Lena Fantuzzi, John Fulton, James Matthes, Joseph Rosenblatt, Michael Stickloon, and Jennifer Sutton for assistance in data collection and analysis.

Please direct correspondence to Jacob Feldman, Dept. of Psychology, Center for Cognitive Science, Rutgers University – New Brunswick, 152 Frelinghuysen Rd., Piscataway, New Jersey, 08854, or by e-mail at jacob@ruccs.rutgers.edu.

[1] The term "temporal grouping" is admittedly elegant, but we feel that the analogy with spatial grouping is not perfectly apt. In spatial grouping elements are aggregated together while continuing to maintain separate existence, whereas in the problem at hand distinct elements are interpreted as being in fact the *same* individual, and thus unified.

main, phenomenally, the "same" object (Fig. 1). *No* change in features may be the most likely case (a); but clearly some featural changes, such as the change in retinal shape associated with rotation in depth, are quite plausible (b); while other changes, such as non-rigid changes in shape, are less plausible (c); and still other changes highly implausible (d). Later in this paper, we will seek to capture this nexus of vague expectations about the evolution of an object's properties as a concrete probability distribution defined over a feature space, which we call the *object evolution function*. In the theory we develop below, this probability distribution will then serve as the centerpiece of the observer's decisions about object individuation in an ambiguous situation, such as our experimental paradigm. Our main conclusion will be that observers perceive as continuously existing objects those paths that entail the minimum of feature change over time—or, more precisely, the *least unlikely* feature change given the observer's subjective probabilistic expectations as captured in the evolution function.

The developmental literature has at times explicitly counterposed *spatiotemporal* properties, such as continuity of location over time, with *featural* properties, usually visual features such as shape and color. Infants as young as four months of age can individuate objects based on spatiotemporal properties such as continuity (Spelke, Kestenbaum, Simons, & Wein, 1995). But even at the age of nine months, infants do not reliably individuate based on featural properties (Tremoulet, Leslie, & Hall, 2000), but develop this ability by 12 months (Xu & Carey, 1996), though this issue remains controversial. Experiments with adults have also suggested that location is primary while properties are secondary (Johnston & Pashler, 1990; Nissen, 1985). Our experimental paradigm is designed so that all candidate object paths are continuous and hence spatiotemporally possible. This allows us to manipulate featural differences and investigate their influence on object interpretations.

The bouncing/streaming paradigm

The paradigm we will use in the experiments below is a variant of one introduced by Michotte (1946/1963, exp. 24), and later, independently, by Julesz (in about 1959; see Julesz, 1995, p.50). More recently it was reintroduced (apparently without knowledge of these earlier uses) by Bertenthal, Banton, and Bradbury (1993) and Sekuler and Sekuler (1999) as a tool to study motion perception.

In a typical display, two objects approach each other from the left and right edges of the screen, "collide" in the middle, and then two objects emerge from the collision moving in opposite directions. The question for the subject is: after the collision, which object is which? In the simplest case of two identical objects moving at constant velocity, the most common percept is that the objects appear to pass through one another ("streaming"), but under certain circumstances the objects appear to strike each other and abruptly reverse motion direction ("bouncing"). The preference for streaming is thought to reflect a preference for straight, constant-velocity motion paths, an important bias of the motion interpretation system (Ramachandran & Anstis, 1983), and hence this task has most often been used to investigate basic motion mechanisms. Such studies usually use featurally identical objects while manipulating spatiotemporal aspects such as speed and acceleration (Sekuler & Sekuler, 1999), attentional demands (Watanabe & Shimojo, 1998), or exogenous cues such as sound (Sekuler, Sekuler, & Lau, 1997).

In our slightly modified version of this task (Fig. 2), the two objects appear from the upper left and right corners of the screen, moving down and towards a central occluder; simultaneously disappear behind the occluder; and then reemerge on the two original paths, but having switched properties. Crucially, one can regard the two objects as having constant properties, but exchanging paths (in which case one sees bouncing); or as having constant (straight) paths, but swapping properties (streaming).

Subjectively, the percept of "bouncing" or "streaming" in this task is very vivid: one either has an immediate percept of two objects crossing without touching or, alternatively, of an abrupt collision, with a concomitant sense of which object is which after they emerge from behind the occluder.

Our displays differ from those of Bertenthal et al. (1993), Sekuler and Sekuler (1999) and others in two respects. First, our two paths cross transversally at the occluder, while others' are strictly horizontal. We felt that the "accidental" collinear alignment between perfectly horizontal paths might bias observers to see the two paths as causally related in some way. Moreover Michotte (1946/1963), using a horizontal display, had observed in a small number of subjects an anomalous depth-rotation interpretation which we wished to avoid. Second, in our displays, an occluder covers the point of intersection, while in others' displays there is no occluder. The occluder was necessary to ensure that displays with certain featural differences (especially in shape and size) were completely ambiguous between bouncing and streaming (i.e. that both interpretations were consistent with the display).

Related phenomena and literature

Before presenting our experiments we briefly review some relevant phenomena already studied in the literature.

Apparent motion

An analogous and related area of research is apparent or phi motion (see Anstis, 1980). In an apparent motion display (Fig. 3a), one visual item is briefly flashed at one time, and then another item is flashed at a different location at a slightly later time; the usual result is a perception of motion between the two locations. Some authors (Burt & Sperling, 1981; Navon, 1976) have found that apparent motion is influenced primarily by spatiotemporal properties (e.g. the magnitudes of the spatial and temporal gaps between the two items), and that featural properties of the items play little role. However others (Shechter, Hochstein, & Hillman, 1988; Prazdny, 1986) have found a measurable benefit of featural similarity between the items. Many authors have suggested a distinction between short-range and long-range apparent motion

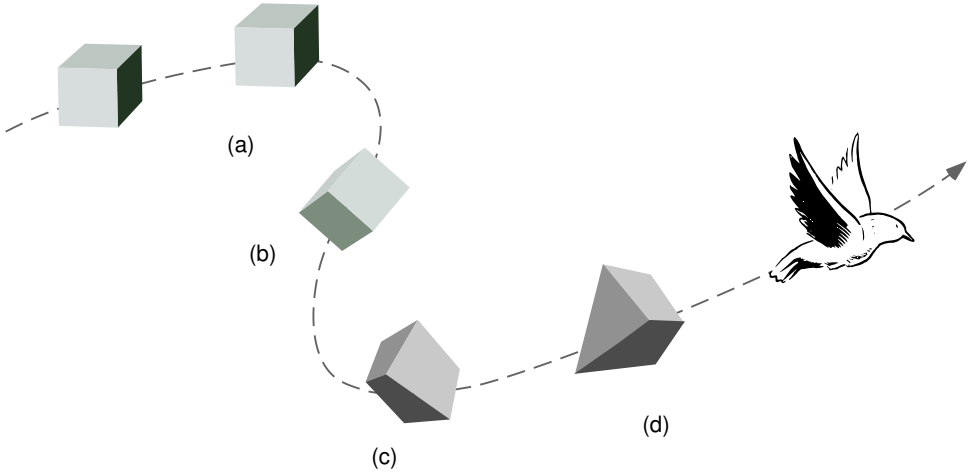

Figure 1. As an object evolves over time, zero feature change (a) is the most likely case, but certain featural changes are highly plausible (b), while others are less plausible (c), and others are highly implausible (d).

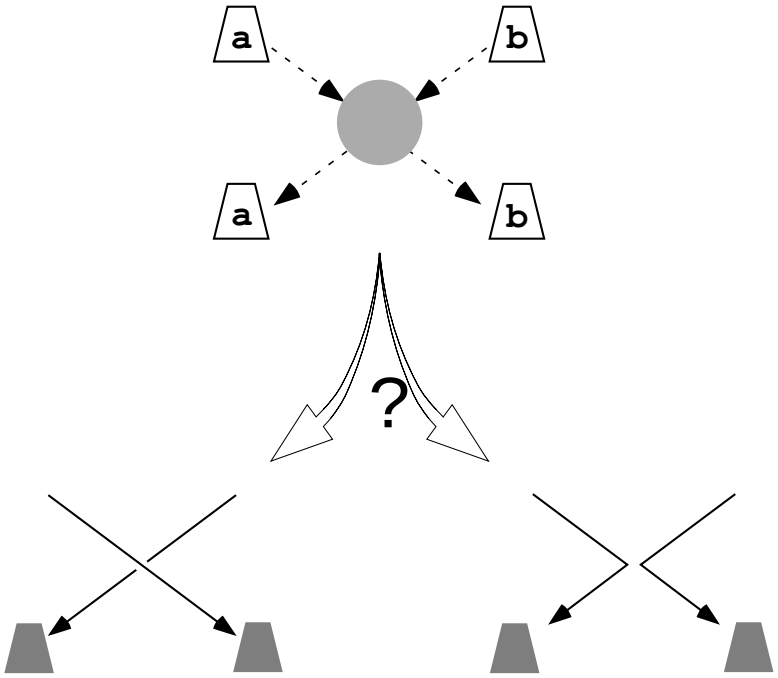

Figure 2. The bouncing/streaming task.

(Lu & Sperling, 2001 for a recent review), with only the latter being influenced by later visual processing involving overt featural properties (see also our discussion below of motion energy models). It is thus certainly possible that object individuation is related to long-term apparent motion; the two processes are at the very least closely analogous. However the items in our paradigm disappear briefly behind an occluder at the critical moment in each trial, so it seems unlikely that the two tasks are identical. In any case, as will be seen later, the assumption that our task does not primarily involve early motion mechanisms is bolstered by the finding that performance in the bouncing/streaming task is not well accounted for by standard early motion models (see section below on motion energy models).

Multiple-object tracking

Another relevant literature is that on multiple-object tracking (MOT) (Pylyshyn & Storm, 1988; see Fig. 3b). In this task, subjects are asked to track a small set of moving objects amid a field of (visually identical) distractor objects. Most subjects can track about four such objects among a field of eight. Normally in this task all the items are featurally identical, so tracking is based on continuous monitoring of spatial trajectories rather than featural information. To our knowledge the influence of featural information on tracking in MOT has not been studied. Subjects in an MOT task can track objects behind occluders (Scholl & Pylyshyn, 1999), and in a similar task can track using continuity in abstract

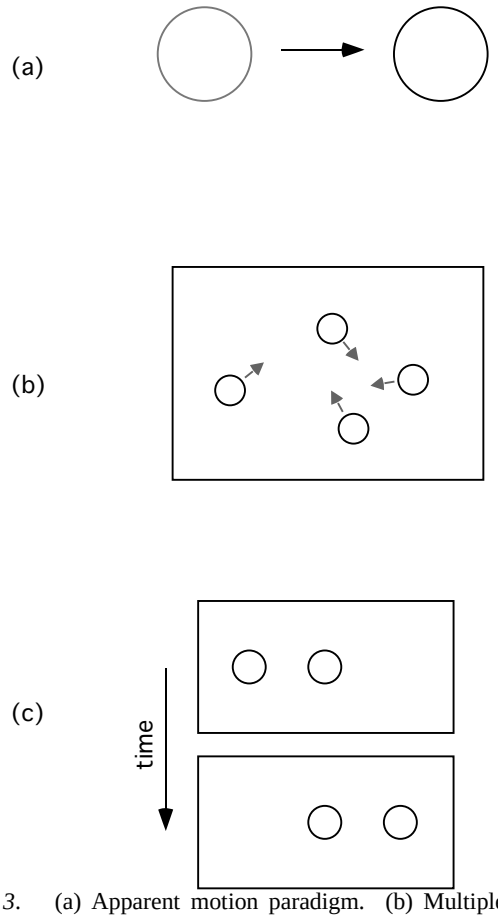

Figure 3. (a) Apparent motion paradigm. (b) Multiple-object tracking paradigm. (c) Ternus illusion. One either sees two objects rigidly translating, or the left-hand object "leap-frogging" over the static center object.

feature-space (Blaser, Pylyshyn, & Holcombe, 2000), a notion closely related to the abstract feature-change space we will develop below. Hence the notion of "individual object" tapped by our bouncing/streaming task is probably the same as that tapped by MOT. The emphasis in MOT studies though is on the division of attention among the various items to be tracked, and how this attentional load is affected by the number of items. Our displays do not vary the number of items; rather the emphasis is on how observers solve the ambiguous correspondence between the items before the occluder and those after the occluder, an ambiguity not normally present in the MOT task.

Finally, we also mention two other phenomena that relate to object individuation. A study by Gepshtein and Kubovy (2000) considered spatiotemporal grouping using a temporal variant of the method of dot lattices introduced by Kubovy (1994) to study spatial grouping. This study drew several interesting conclusions about temporal grouping, in particular concerning the effect of spatial and temporal factors, but did not consider featural differences. Second, the well-known Ternus illusion (see Fig. 3c) features an ambiguity of object identity over motion. This illusion is known to depend heavily on spatiotemporal factors (e.g. the interframe interval; see Yantis, 1995), but extant studies have all used featurally identical elements, so the effect of featural differences is, again, unknown. Hence notwithstanding a great deal of speculation about the role of featural continuity in determining object identity, the actual influence of featural information in adults' judgments remains largely unstudied.

Experiments

When describing our displays, for clarity of exposition we will refer to the left-hand item as **a** and the right-hand item as **b** (see Fig. 2); hence the symbols **a** and **b** each refer consistently to an entity with constant properties. (This terminology is for convenience; in the actual displays left and right sides were counterbalanced.) We parameterize the displays with respect to the featural difference ΔF between **a** and **b**; e.g. $\Delta F = 0$ means $\mathbf{a} = \mathbf{b}$. Later we will present a theoretical model in which we predict the probability of a bounce response as a function of the featural difference ΔF.

Summing up, when confronted with any of our displays, the observer has a choice between a percept of streaming—but with an attendant abrupt feature change of magnitude ΔF as each object passes behind the occluder; or a percept of bouncing, with no attendant featural change. Hence this task directly measures the observer's judgment of the likelihood of a featural change of ΔF within the lifeline of a single, coherent object. The subjective likelihood of a given feature change, in turn, reflects the nexus of subjective expectations about plausible feature change, which we hope will illuminate the subjects' underlying mental model of "objects."

As noted above, when $\Delta F = 0$, subjects usually report streaming, consistent with (or analogous to) the general preference for straight motion paths and "inertia" in apparent motion (Bertenthal et al., 1993; Ramachandran & Anstis, 1983). Hence clearly featural differences are not the only factor influencing interpretation of our displays. However this bias is a constant throughout all our conditions, while featural differences ΔF are manipulated. In the formalism presented below, we will assume the bias for streaming is a single scalar weight (in effect, the subjective prior probability attached to the streaming hypothesis) that does not interact with any of the manipulated effects.

In the following experiments, we manipulated the featural difference ΔF between items **a** and **b** (see Fig. 2), and measured its influence on "bouncing" vs. "streaming" responses. As features, we focused on luminance (Exp. 1), size (Exp. 2), and shape (Exp. 3). We were also especially interested in the manner in which multiple cues are combined (a topic of substantial recent interest among perceptual theorists; see discussion below), so we also separately ran conditions manipulating all three pairs of features: ; luminance $\times$ size (Exp. 4), luminance $\times$ shape (Exp. 5), and size $\times$ shape (Exp. 6).

Notation. Each of our objects can be thought of as a point in luminance-size-shape space (Fig. 4), which we denote by F:

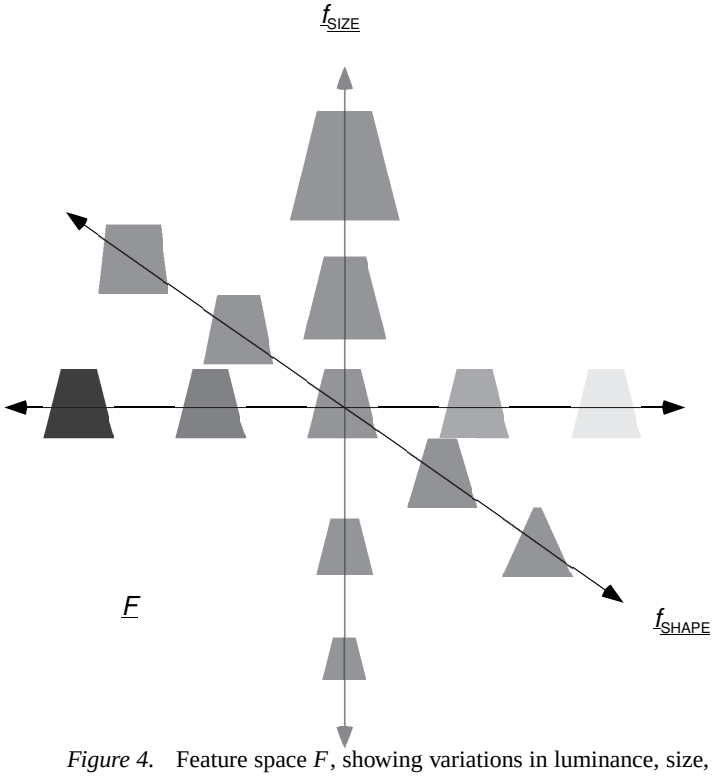

Figure 4. Feature space F, showing variations in luminance, size, and shape.

$$F = \langle f_{\text{LUM}}, f_{\text{SIZE}}, f_{\text{SHAPE}} \rangle . \tag{1}$$

We denote by $F_{\mathbf{a}}$ and $F_{\mathbf{b}}$ the feature vectors for objects **a** and **b** respectively. We are interested in the difference ΔF between them, given by

$$\Delta F = F_{\mathbf{a}} - F_{\mathbf{b}}, \tag{2}$$

which is a vector of three individual feature differences,

$$\Delta F = \langle \Delta f_{\text{LUM}}, \Delta f_{\text{SIZE}}, \Delta f_{\text{SHAPE}} \rangle . \tag{3}$$

The main manipulation in the experiments is always ΔF, with each experiment focusing on a single component or pair of components.

By Fechner's law, we actually expect psychological differences to reflect ratios rather than differences of the raw values. So in order to make F a simple vector space, we use logarithmic features, e.g.

$$f_{\text{LUM}} = \log(\text{raw luminance}), \tag{4}$$

and similarly for f_{SIZE} and f_{SHAPE}. This means that vector differences in this space reflect ratios in the original raw values, e.g.

$$\Delta f_{\text{LUM}} = \log\left(\frac{\text{raw luminance of } \mathbf{a}}{\text{raw luminance of } \mathbf{b}}\right), \tag{5}$$

and similarly for size difference Δf_{SIZE} and shape difference Δf_{SHAPE}. This means that "no difference" is always signified by $\Delta f = 0$ (because $0 = \log(1)$). In the experiments, we attempt to choose values of each Δf giving a wide range of differences, and always including a same-feature case ($f_{\text{LUM}} = f_{\text{SIZE}} = f_{\text{SHAPE}} = 0$, i.e. $F_{\mathbf{a}} = F_{\mathbf{b}}$). Then we choose pairs of values of f that center their difference Δf symmetrically around an intermediate value of the parameter.

Parameters

Luminance. Each object was a uniform gray region (on a white background) with reflectance drawn from the range between 0 (black) and 1 (brightest white). As explained, Δf_{LUM} thus represents the log ratio of the raw luminances (percent white) of **a** and **b**.

Size. Objects were uniformly scaled to create size differences. Δf_{SIZE} thus represents the log ratio of the linear span of **a** to that of **b**.

Shape. As a shape parameter, we created a one-parameter continuum of shapes running from square to triangle (see shape axis in Fig. 4), with intermediate values producing a spectrum of of trapezoids. As with the other parameters, we then take the log ratio, so Δf_{SHAPE} represents the log of the ratio of **a**'s position along this spectrum (i.e., percent triangle) to that of **b**.

In each experiment, the feature(s) not manipulated were fixed at an intermediate value (50% white, medium size [about 1° of visual angle at 45cm of viewing distance], or 50% triangle). Hence all shapes in the luminance condition (Exp. 1), size condition (Exp. 2), and luminance × size condition (Exp. 4) were trapezoids with an intermediate value of the shape parameter.

In the two-parameter experiments (Exps. 4–6), we used the same values of each of the parameters as had been tested in the single-parameter experiments, so subjects' treatment of parameters in combination could be compared as directly as possible to the same parameters taken singly.

Method

Methods for all six experiments were identical except for the choices of feature change vectors ΔF. For clarity of presentation we give the general method first, then details of each of the experiments' parameters in sequence, before giving results.

Subjects. Exps. 1–6 used 16, 16, 14, 16, 15, and 17 subjects respectively. Subjects were undergraduate students participating for course credit and were naive to the purposes of the experiment.

General method. The subject was seated in front of the computer screen at a viewing distance of approximately 45cm. The subject would then fixate on the central occluder, a textured circular patch subtending about 2° of visual angle (the exact size is calculated to be sufficient to fully occlude the largest object in any trial). After pressing the spacebar, the two objects would appear from the upper left and

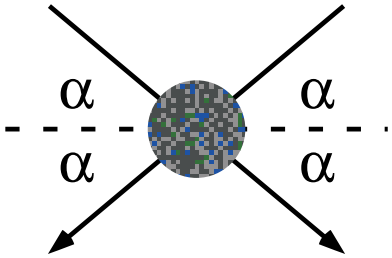

Figure 5. Illustration of the incidence angle α in the displays. The figure also illustrates the texture pattern on the occluder.

right, traveling towards the occluder. The objects would then both simultaneously disappear behind the occluder, becoming progressively occluded until they were both completely invisible. Then two objects would progressively emerge from behind the occluder traveling along the original motion paths, as explained above, except with object **a** now continuing **b**'s motion path and vice-versa. The subject was then asked to indicate whether he or she had perceived a "bounce" or "stream" via response keys on the computer keyboard.

Several other variables in addition to ΔF were also manipulated, fully crossed with the main manipulation of ΔF. In all experiments, the speed of the moving objects was either 6, 12, or 18 degrees of visual angle per second. The incidence angle (Fig. 5) was either 25 or 45 degrees.

On each trial, subjects were forced to respond "bounce" or "cross" (stream). The main dependent variable was the proportion of "bounce" responses (one minus the proportion of "cross" responses), which later we model as function of the featural difference ΔF.

Exp. 1 (luminance). Exp. 1 used the following raw luminance pairs (**a**:**b**, raw percent white): 50:50, 52.5:47.5, 55:45, 60:40, and 70:30 (i.e. spreads of 0, 5, 10, 20, and 40 percentage points centered around 50). These values correspond to ratios of $1.00, 1.1, 1.22, 1.5$ and 2.33, or log ratios Δf_{LUM} of $0, 0.1, 0.2, 0.41$ and 0.85. The actual appearance of these values is illustrated along the abscissa in Fig. 6.

Exp. 2 (size). We used size ratios (**a**:**b**, linear span) of 1:1, 1.05:1, 1.1:1, 1.15:1 and 1.21:1 (actually equal intervals in log units before rounding; the real values are 1.1 raised to the power of 0, 0.5, 1, 1.5, and 2 respectively). These values correspond to log ratios of about f_{SIZE} of $0, 0.05, 0.1, 0.14$ and 0.19. These ratios are illustrated along the abscissa in Fig. 7.

Exp. 3 (shape). We used shape ratios (**a**:**b**, percent triangle) of 50:50, 55:45, 60:40, 70:30, and 90:10. These values correspond to ratios of $1.00, 1.22, 1.5, 2.33$ and 9, or log ratios Δf_{SHAPE} of $0, 0.2, 0.41, 0.85$ and 2.2. The corresponding shapes are illustrated along the abscissa in Fig. 8.

Exp. 4 (luminance × size). Exp. 4 used the same five levels of luminance change f_{LUM} as in Exp. 1, and the same five levels of size change f_{SIZE} as in Exp. 2, fully crossed, for a total of 25 feature-change combinations (i.e., values of ΔF).

Exp. 5 (luminance × shape). Exp. 5 used the same five levels of luminance change f_{LUM} as in Exp. 1, and the same five levels of shape change f_{SHAPE} as in Exp. 3, fully crossed, for a total of 25 feature-change combinations (i.e., values of ΔF).

Exp. 6 (size × shape). Exp. 6 used the same five levels of size change f_{SIZE} as in Exp. 2, and the same five levels of shape change f_{SHAPE} as in Exp. 3, fully crossed, for a total of 25 feature-change combinations (i.e., values of ΔF).

Results

Figs. 6–11 show results for Exps.1–6 respectively. Each plot shows the proportion bounce responses as a function of featural change ΔF. Each plot also shows a theoretical model, which is explained in detail below.

We followed a two-tiered analysis strategy. First, we entered the data from each experiment into an analysis of variance (ANOVA), in order to establish the significance of each of the manipulations. Generally, these analyses show significant effects of all of the featural manipulations, as well as their interactions in the two-feature experiments (and, with a few exceptions, no effects of the nuisance variables speed and incidence angle). The plots suggest complex but highly systematic nonlinear interactions between the featural variables. Hence in the second phase of our analysis, we attempt to model these nonlinearities with a detailed quantitative model. The model is a simple Bayesian observer, which predicts the bounce/stream classification as a function of the featural variables. This model give a good account of the exact nonlinear shape of the decision surfaces shown in the plots.

Analyses of variance

Exp. 1 (luminance). The effect of luminance change was highly significant ($F(4,60) = 12.521, p < .0001$). The effect of speed was also significant ($F(2,30) = 5.697, p = .008$), with bounce responses generally increasing with faster speeds. No other effects or interactions were significant ($p > .1$ in all cases).

Exp. 2 (size). The effect of size change was highly significant ($F(4,60) = 35.995, p < .0001$). The interaction between size and speed was also significant ($F(8,120) = 3.353, p = .002$), with bounce responses rising more quickly with size change at low speeds than at high speeds. No other effects or interactions were significant ($p > .05$ in all cases).

Exp. 3 (shape). The effect of shape change was highly significant ($F(4,52) = 24.434, p < .0001$). The interaction of speed and incidence angle was also significant ($F(2,26) = 3.637, p = .044$), with more of a (non-monotonic) influence of speed at 45° incidence angle than at 25°. No other effects or interactions were significant ($p > .05$ in all cases).

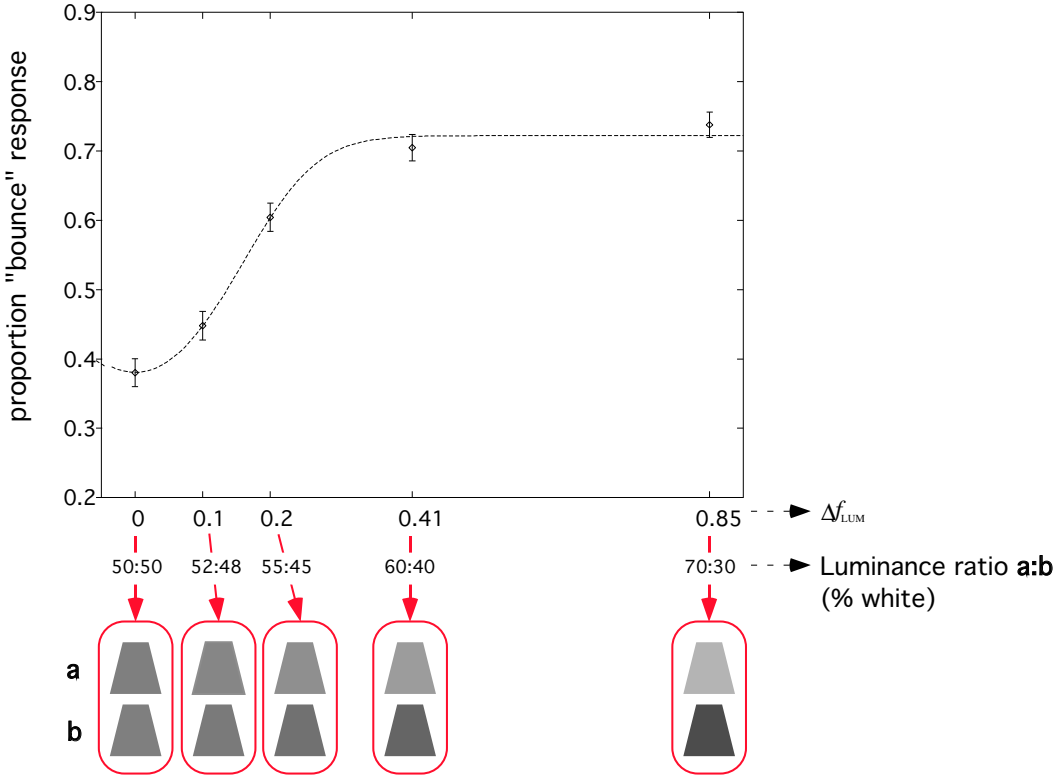

Figure 6. Results from Exp. 1 (luminance), showing the proportion "bounce" responses as a function of luminance difference f_{LUM}. The dotted line shows the Bayesian model (see text). Error bars show standard error.

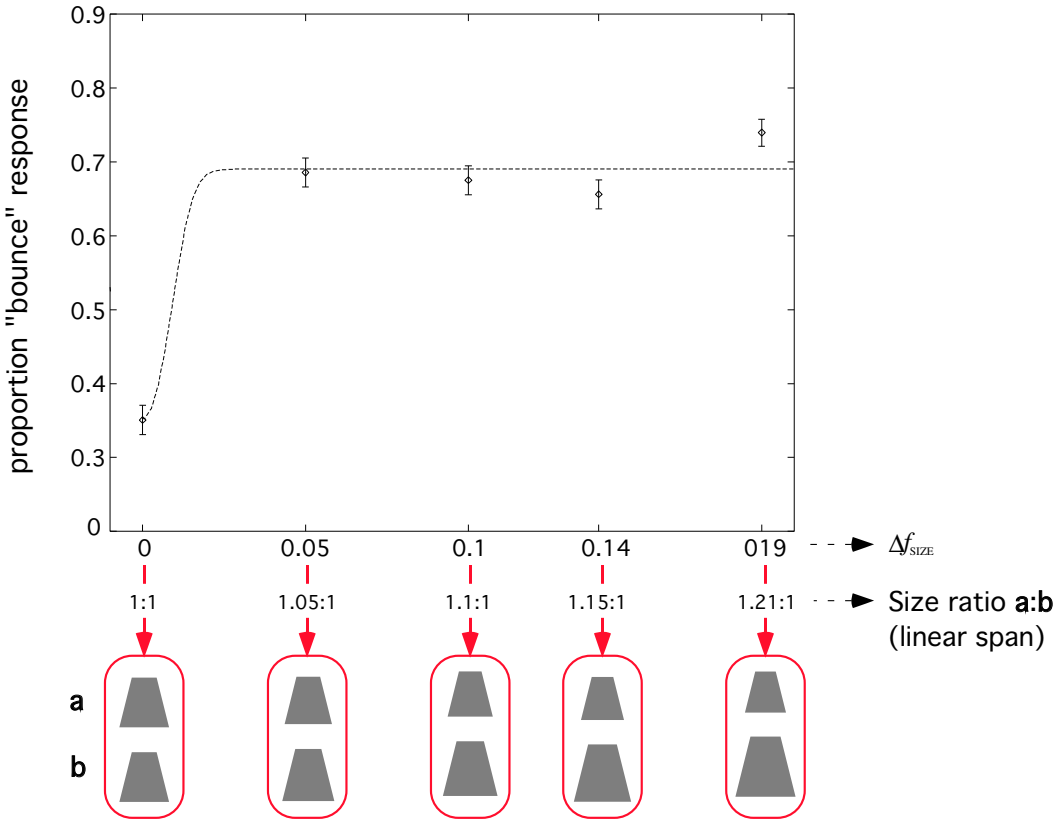

Figure 7. Results from Exp. 2 (size). The dotted line shows the Bayesian model (see text). Error bars show standard error.

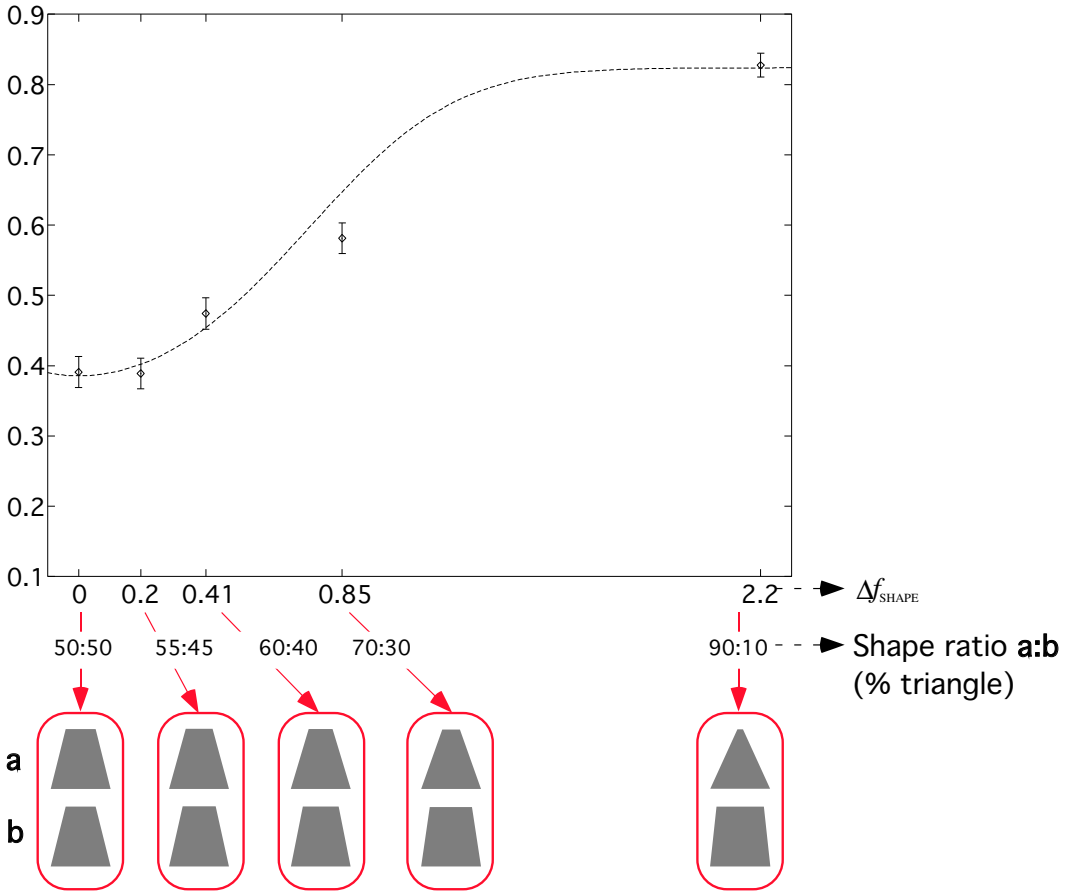

Figure 8. Results from Exp. 3 (shape). The dotted line shows the Bayesian model (see text). Error bars show standard error.

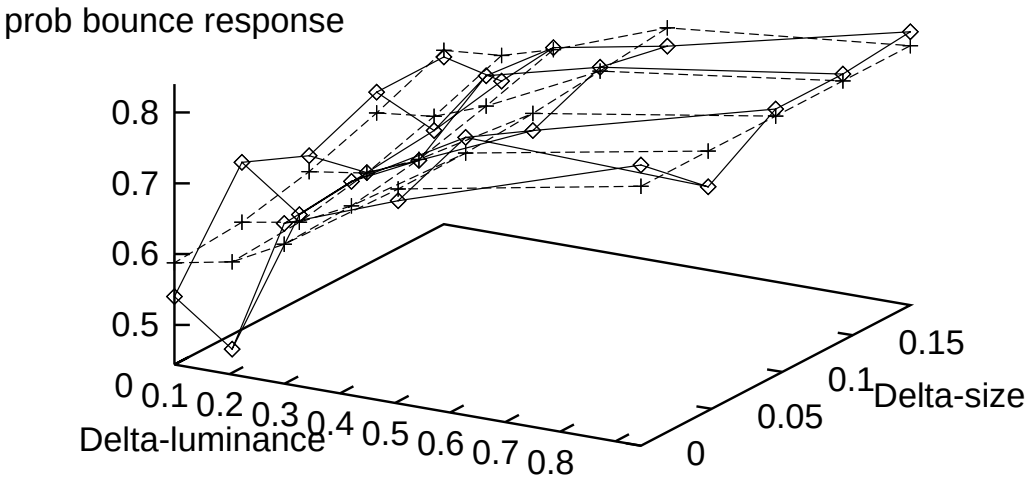

Figure 9. Results from Exp. 4 (luminance $\times$ size). The dotted surface shows the fitted Bayesian model (see text).

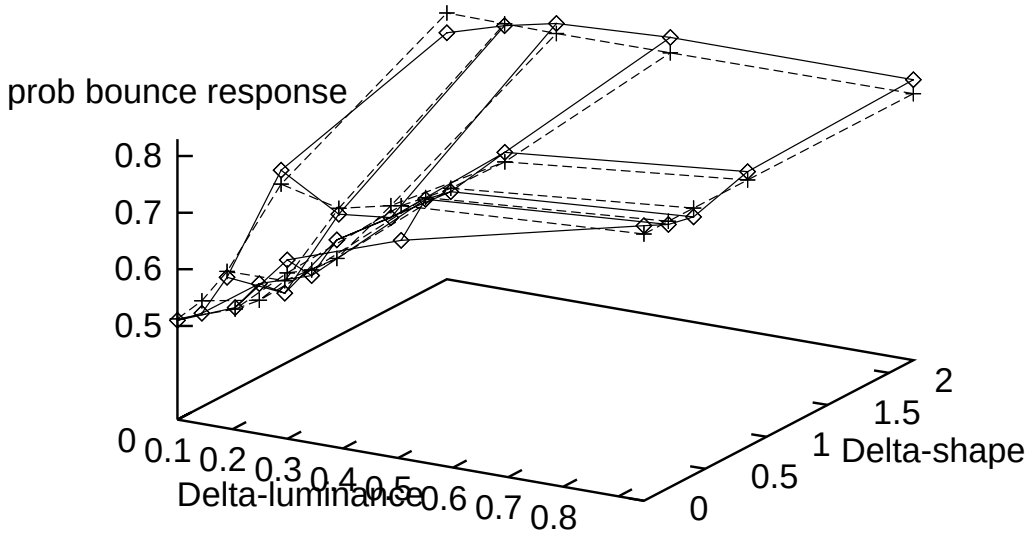

Figure 10. Results from Exp. 5 (luminance × shape). The dotted surface shows the fitted Bayesian model.

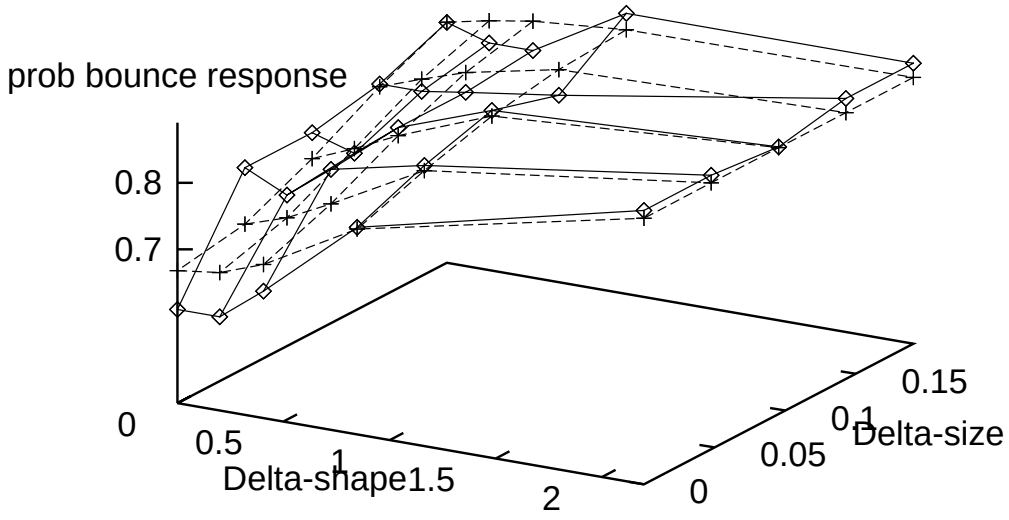

Figure 11. Results from Exp. 6 (size × shape). The dotted surface shows the fitted Bayesian model.

Exp. 4 (luminance × size). The main effects of luminance change and size change were again significant (luminance, $F(4,60) = 11.796, p < .0001$; size, $F(4,60) = 4.367, p = .004$). The luminance by size interaction was also significant ($F(16,240) = 1.887, p = .022$). No other effects or interactions were significant ($p > .05$ in all cases).

Exp. 5 (luminance × shape). The main effects of luminance change and shape change were again significant (luminance, $F(4,52) = 15.916, p < .0001$; shape $F(4,52) = 11.121, p < .0001$). The effect of speed was also significant ($F(2,26) = 6.611, p = .005$), with more bounce responses at higher speeds. The luminance by shape interaction was also significant ($F(16,208) = 3.054, p = .0001$). No other effects or interactions were significant ($p > .05$ in all cases).

Exp. 6 (size × shape). The main effect of size and shape were again significant (size, $F(4,64) = 11.010, p < .0001$; shape, ($F(4,64) = 13.669, p < .0001$). The interaction of size and shape was also significant ($F(16,256) = 1.942, p = .017$). The three-way interaction of shape, speed, and incidence angle was also significant ($F(8,128) = 2.179, p = .033$), with bounce responses rising rapidly at low speeds and 25° incidence angle but more slowly at high speeds and 45° incidence angle. No other effects or interactions were significant ($p > .05$ in all cases).

Summary. The main conclusion from the ANOVAs is that, as expected, bounce responses generally increased with greater featural change. The more **a** differed from **b**, the more often subjects report seeing the bouncing percept; that is, the *less* likely they were to see the given featural change as consistent with a single, coherent object. This result is not in itself surprising, but it confirms the basic idea that the subjectively continuous existence of an object is mentally associated with small changes in its features.

This conclusion requires one caveat. We have assumed so far that the visual system first reduces each visual item to a featural representation, and then determines correspondence over time based on the features. Gepshtein and Kubovy (2000) have shown however that the process of perceptually interpreting each individual time-slice can be influenced by the inferred correspondence with subsequent frames, suggesting in their terms an interactive rather than sequential model. They drew this conclusion based on displays (using their spatiotemporal dot lattice paradigm) in which the relative grouping strengths of within-frame and between-frame correspondences were deliberately manipulated. In our displays, the interpretation of each individual frame is not ambiguous in this way. Hence we assume that the spatiotemporal interactivity discovered by Gepshtein and Kubovy (2000) will play only a minimal role.

The challenge next is to model the bounce/stream classification data more precisely, in order to understand exactly what mental assumptions and mechanisms they reflect. We take up this challenge in the next section by postulating a Bayesian observer endowed with a simple subjective model of "objects."

A Bayesian observer model

When confronted with one of our displays, or indeed with any real stream of images, an observer is faced with an uncertain decision. If an object in the current image perfectly matches the features of exactly one in the previous image, then the individuation might be unambiguous. But far more often no match is perfect, because of noise in the image, and also more substantively because objects' features really can change, due to pose change, non-rigidity, and other varieties of common transformations.

Such a decision can be modeled effectively in a Bayesian framework, often applied recently to perceptual inference (see Bülthoff & Yuille, 1991; Feldman, 2001; Knill & Richards, 1996; Landy, Maloney, Johnston, & Young, 1995 for examples). In this section we formulate a Bayesian model of observers in our task. As with any observer model, the critical issue is what to assume about the observer's state of knowledge and beliefs. In our model, we assume only that the observer has certain subjective expectations about the probability of feature changes, which are encoded in a probability distribution function. The observer can then "turn the Bayesian crank," and place an interpretation on a particular display by, in essence, plugging this distribution into Bayes' rule. This yields a decision function that, we then show, serves very well as a model of the data from our six experiments.

Assumptions

In Bayesian theory, the observer's subjective belief in a particular hypothesis H given data D is associated with the *posterior probability* $p(H|D)$, which can be computed via Bayes' rule,

$$p(H|D) = \frac{p(D|H)p_H}{\sum_i p(D|H_i)p_i)}. \tag{6}$$

Here the numerator is the product of *likelihood* of hypothesis H, $p(D|H)$ (which gives the probability of observing D if H were in fact true) and its *prior probability* p_H (which says how likely H was before this particular trial was observed). The denominator sums this product over all possible hypotheses, including both H and all other alternative hypotheses H_i. Hence the whole expression says how plausible H is relative to the set of competing hypotheses.

In our situation, we are interested in the posterior probability of the bounce interpretation given the display as parameterized by the observed featural difference ΔF, denoted $p(\text{BOUNCE}|\Delta F)$. Via Bayes' rule, this is given by

$$p(\text{BOUNCE}|\Delta F) = \frac{p(\Delta F|\text{BOUNCE})p_{\text{BOUNCE}}}{p(\Delta F|\text{BOUNCE})p_{\text{BOUNCE}} + p(\Delta F|\text{STREAM})p_{\text{STREAM}}} \tag{7}$$

where p_{BOUNCE} and p_{STREAM} are the priors on bouncing and streaming respectively, and $p(\Delta F|\text{BOUNCE})$ and $p(\Delta F|\text{STREAM})$ are respectively the likelihoods of a given

feature change under the bouncing and streaming interpretations. We make the following simple assumptions concerning these parameters. First, we assume that all displays are either bouncing or streaming, so $p_{\text{BOUNCE}} = 1 - p_{\text{STREAM}}$ and $p(\text{BOUNCE}|\Delta F) = 1 - p(\text{STREAM}|\Delta F)$. More substantively, we assume that objects may take on any feature value F with equal probability,[2] or, more precisely, that all values of ΔF are equally likely when it represents the featural difference between two *distinct* objects (as opposed to two different incarnations of the *same* object, as under the streaming interpretation). In mathematical terms, this means that $p(\Delta F|\text{BOUNCE}) = 1$ always: any feature change is perfectly consistent with a bounce interpretation.

With these assumptions, we can now rewrite the posterior on the bounce interpretation as

$$p(\text{BOUNCE}|\Delta F) = 1 - \frac{p(\Delta F|\text{STREAM})p_{\text{STREAM}}}{p(\Delta F|\text{STREAM})p_{\text{STREAM}} + 1 - p_{\text{STREAM}}}. \tag{8}$$

This equation has only two variables on the right-hand side: the prior p_{STREAM}, which is a simple scalar, and the likelihood term $p(\Delta F|\text{STREAM})$, which is a function mapping feature change vectors to probabilities. In our analysis, we treat the prior p_{STREAM} as a free parameter to be estimated from the data. The main focus then is on the likelihood term $p(\Delta F|\text{STREAM})$, which represents the likelihood of a particular feature change ΔF under the streaming interpretation. The next section asks what this crucial function might be expected to look like.

The object evolution function

The function $p(\Delta F|\text{STREAM})$ expresses the subject's probabilistic expectations about how features may change from before an object disappears behind the occluder until after it reappears, given that it is actually *the same individual entity.* More generally, we assume, this function expresses how likely it is for a given featural change to occur within the lifeline of a single individual object from time t to time $t + \Delta t$: that is, the probability that an object will "evolve" by ΔF during an interval Δt (Fig. 12). Hence we will refer to this function as the *object evolution probability density function*, or more briefly, as the *evolution function*, and denote it as $\Psi(\Delta F)$,

$$\Psi(\Delta F) = p(\Delta F|\text{STREAM}). \tag{9}$$

This function $\Psi(\Delta F)$ lies at the heart of the subject's beliefs about how objects tend to behave. What can we say about its form?

We begin by asking what *mean value* the evolution function ought to have, formally expressed by the mathematical expectation $E[\Psi(\Delta F)]$. That is, given that we observe an object F at time t, by how much do we typically expect it to change after an interval Δt?

Our basic assumption is that, all else being equal, object properties tend to be stable. That is, at each time slice, an object is most likely to have the *same* features as at the previous time slice. This is a very basic and non-trivial assumption about feature change, which we refer to as the *feature stability assumption:*

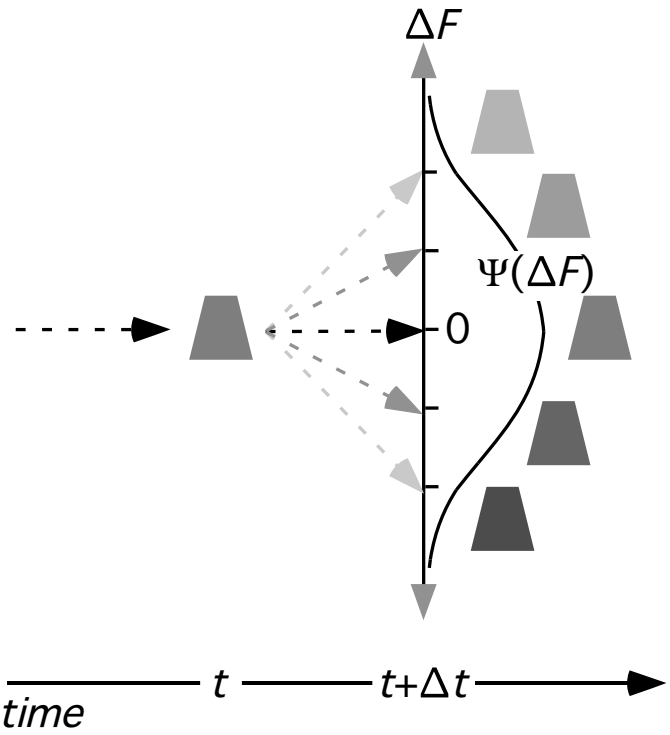

Figure 12. Illustration of the evolution function $\Psi(\Delta F)$. The function gives the expected distribution of change ΔF in the object's over the time interval Δt.

[Feature Stability assumption]

$$E[\Psi(\Delta F)] = 0. \tag{10}$$

This means that $\Psi(\Delta F)$ will be centered at $\Delta F = 0$.

Another way of thinking of the same idea is to think of the object in terms of its feature vector over time $F(t)$, that is, over the "evolution" of the object. If an object has feature vector $F(t_0)$ at a certain time t_0, then at time $t_0 + \Delta t$ we expect it to have feature vector:

$$E[F(t_0 + \Delta t)] = F(t_0) + E[\Psi(\Delta F)]. \tag{11}$$

Plugging in the feature stability assumption $E[\Psi(\Delta F)] = 0$, this immediately yields

$$E[F(t_0 + \Delta t)] = F(t_0). \tag{12}$$

In words, at each time slice, all else being equal, an object is most likely to have the *same* features as at the previous time slice.

[2] If the space F is unbounded this creates the possibility of what are called *improper priors* in the Bayesian literature. However standard methods for dealing with this situation have been developed (see Box & Tiao, 1973 for introduction).

Having established the mean of the evolution function, we next have to worry about its functional form. In what follows, we assume that it is normal (Gaussian) in form—a very common assumption in the Bayesian literature, for a variety of good and bad reasons.[3] In our case this means that in general we assume $\Psi(\Delta F)$ is Gaussian over ΔF, centered at $\Delta F = 0$ and with covariance matrix Σ, notated

$$\Psi(\Delta F) = N(\Delta F; 0, \Sigma). \tag{13}$$

More specifically, in our experiments ΔF is always a vector of either one featural dimension (Exps. 1–3) or two (Exps. 4–6). Hence in Exps. 1–3 the evolution function is a simple univariate normal,

$$\Psi(\Delta f_i) = N(\Delta f_i; 0, \sigma_i), \tag{14}$$

where Δf_i is either $\Delta f_{\text{LUM}}, \Delta f_{\text{SIZE}}$, or Δf_{SHAPE}, and σ_i is an associated standard deviation. Similarly in Exps. 4–6 Ψ is a bivariate normal

$$\Psi(\Delta f_i, \Delta f_j) = N(\Delta f_i, \Delta f_j; \langle 0,0 \rangle, \sigma_i, \sigma_j, r), \tag{15}$$

where Δf_i and Δf_j are the two relevant feature-change parameters, σ_i and σ_j are their associated standard deviations, and r is the correlation between them.

To produce our final Bayesian model, we plug these assumptions about Ψ back into Eq. 8, and place a scaling coefficient h in front of the entire expression:

$$\begin{aligned} p(\text{BOUNCE}|\Delta F) &= h\left[1 - \frac{p_{\text{STREAM}}\Psi(\Delta F)}{p_{\text{STREAM}}\Psi(\Delta F) + 1 - p_{\text{STREAM}}}\right] \\ &= h\left[1 - \frac{p_{\text{STREAM}}N(\Delta F; \Sigma)}{p_{\text{STREAM}}N(\Delta F; \Sigma) + 1 - p_{\text{STREAM}}}\right] \end{aligned} \tag{16}$$

This will now serve as a model of the data from Exps. 1–6, with the free parameters fitted to the data. In the single-feature experiments (Exps. 1–3), the model has three free parameters: the leading scaling term h, the streaming prior p_{STREAM}, and the single feature standard deviation σ. In the two-feature experiments (Exps. 4–6), the model has five free parameters: the leading scaling term h, the streaming prior p_{STREAM}, the two featural standard deviations σ_i and σ_j, and the correlation coefficient r between them. Fitting the data in the single-parameter experiments, which have only five data points (i.e., mean bounce responses for each of five levels of feature change) is relatively easy; the main question is whether the model gives qualitatively the right behavior. The more serious challenge to the Bayesian model is in the two-parameter experiments, where we will use the five-parameter model to fit 25 data points (a 5×5 grid of feature change levels); here the question is whether the same basic Bayesian model will fit the more complex dataset.

Sources of the evolution function

The evolution function $\Psi(\Delta F)$ represents the observer's expectations about how an object is prone to change over time. We have so far assumed that its mean will be at zero (no change the most likely) and that its form will be generally Gaussian. However this leaves open several questions about the nature and sources of these distributions. Where do the observers' subjective expectations about object evolution come from? Our data do not speak directly to this question, so our discussion is necessarily speculative.

Some changes to observed object properties are intrinsic, in the sense that the properties are tied to the object themselves, and other extrinsic, in that they depend on viewing conditions. For example despite perceptual invariances, objects may appear different colors or luminances at different moments despite constant material properties. Such extrinsic changes add uncertainty to the data our observer is using to track identity. Thus from a formal point of view they would be folded into the evolution function. Another important kind of extrinsic property change is change in shape due to viewpoint change. However as pointed out by Ullman (1977), such changes are potentially almost unbounded; one can construct objects whose appearances from orthogonal viewing directions are arbitrarily dissimilar. Of course the potentially large changes in shape introduced by viewpoint changes are one of the central problems studied in the object recognition literature (Tarr & Pinker, 1989).

As for intrinsic property changes, many objects in the natural world can alter their intrinsic shape, color, or size, though generally not on the sub-second time-scale of our experiments. It is intriguing in this regard to consider the chameleon, blowfish, and hognose snake, three animals that alter respectively their color, size, and shape in response to threat. Such adapations seem designed to conceal the animal's identity precisely by fooling predators' perceptual systems via their assumption that such changes are generally unlikely. More mundanely, shape changes in articulated and non-rigid objects such as animal bodies are commonplace. One would not want to think that your cat had been replaced by a *different* cat simply because it moved its tail.

More generally, one might imagine that observers' expectations about how intrinsic properties might change over time would relate to their beliefs about the material properties of the surfaces in question. Many computational models of surface perception, in which smooth surfaces are reconstructed from isolated depth values (e.g. see Blake & Zisserman, 1987), rely on assumptions about the physical flexibility of the underlying surfaces, which would naturally relate to their likelihood of changing shape over time. Thus a surface judged to be made of wood would be expected to have a tighter shape evolution function than one judged to be made of rubber. Of course in our impoverished displays, the subjects had little data on which to base estimates of material properties, so they might been led to employ some kind of

[3] Good reasons include that the Gaussian is the maximum entropy function with a given fixed mean and variance (see Bernardo & Smith, 1994), and that the Gaussian is the limiting sum of a large number of independent distributions (the Central Limit theorem). Bad reasons include that it is mathematically simple and convenient to work with.

neutral default distribution.

Finally we note that the observer's subjective expectations may not themselves be firmly fixed, but may vary depending on context and mental set. Specifically it is certainly possible that our subjects' distributions were "tuned" by their experiences viewing our stimuli; over the course of trials they may have gradually developed a sense of the range of feature changes at play in the experiments. This possibility means that we cannot draw any very firm conclusion about the meaning of the specific values of σ (standard deviation) observed in our data. Rather it is the general form of the decision procedure that is of interest.

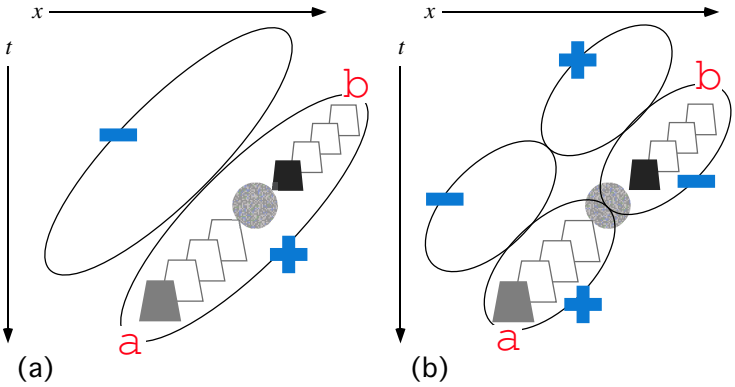

Figure 13. Two ways of applying a motion energy model to our displays. (a) Standard arrangement, with a pair of mutually inhibitory receptive fields (one positive and one negative), each oriented in space-time x and t (or x, y and t in the full model). In the full model, such fields would normally be coupled in quadrature pairs of opposite contrast polarity. (b) An alternative arrangement, yielding the difference or change in motion energy as the items cross behind the occluder.

Fits of the Bayesian model

We fitted the data (probability of a bounce response as a function of featural difference ΔF) to the Bayesian model (Eq. 17) using Levenburg-Marquardt (a common nonlinear model estimation technique). Estimated parameters and goodness-of-fit (R^2) for each experiment are given in Tables 1 (Exps. 1–3) and 2 (Exps. 4–6). The fitted models are plotted alongside the data in Figs. 6–11.

The Bayesian model fits the data very closely in all six experiments, as demonstrated by the high R^2 values, and even more vividly by the extremely close matches visible in the figures. In the single-parameter experiments, as mentioned above, because the dataset has few degrees of freedom compared to the model, the very good fit ($R^2 > .95$ in all cases) is not in itself very probative; it shows only that the model has qualitatively the correct form. But in the two-parameter experiments, where the number of data-points (25 per experiment) greatly exceeds the number of degrees of freedom in the model (5), the good fit ($R^2 > 0.79$ in all cases) is far more demonstrative (and is significant in each case: for Exps. 4, 5, and 6, $F(5,19) = 14.50, 94.65$ and 14.96 respectively, $p < .00001$ in each case). Informally, the fact that all the fitted parameters take on reasonable and meaningful values (e.g., prior probabilities between 0 and 1, correlation coefficient between -1 and 1, etc., none of which conditions are forced by the fitting procedure) suggests very strongly that the model is qualitatively correct in form. In practice, when the model is qualitatively defective in even a small way, some parameters will tend to diverge (go to infinity or minus infinity), which never happened here.

In summary, the Bayesian model gives a very accurate prediction of the subject's responses. Subjects weigh the evidence they observe from each of the feature changes they observe—and combine these cues to form an impression of which object is which in the display—in a manner very close to that prescribed by Bayes.

Comparison with motion energy models

A natural competitor for the Bayesian model in explaining our subjects' responses are spatiotemporal energy models of motion perception, such as that of Adelson and Bergen (1985). Such models, which have been very successful tools in understanding early motion perception, are based on the idea of receptive fields that are oriented in space-time (rather than just in space; see Fig. 13). In order to apply such a model to our displays, we need to make a few assumptions: (i) that the positive lobe of one such receptive field exactly covers the straight ("streaming") path of our moving shapes, symmetrically around the occluder (and presumably no single receptive field exactly covers the bent "bouncing" path); (ii) that we can ignore the presence of the occluder (which in reality would diminish the motion energy but not change the direction of any of the model's predictions); and (iii) that a filter is available at each of the velocities (space-time orientations) used in the experiment. With these assumptions, such a model would indeed explain the general preference for streaming percepts, because there would always be more motion energy along the streaming path than along the bouncing path.

However, the motion energy model cannot explain the way responses varied with featural difference ΔF. The total energy is computed as the total stimulation within the excitatory lobe (minus that in the inhibitory lobe, which we are assuming is empty and thus zero). In our displays this means that the energy is proportional to the luminance of each item (which depends on f_{LUM}), integrated over the total area of the item (which depends on f_{SIZE} and f_{SHAPE}), integrated over all of the items that fall within the receptive field. The area of a trapezoid with shape parameter f_{SHAPE} is $(1 - f_{\text{SHAPE}}/2)s^2$, where s is the width at the bottom edge (a constant). Hence the total motion energy due to each item is proportional to

$$E_{\text{item}} = f_{\text{LUM}} f_{\text{SIZE}} (1 - \frac{f_{\text{SHAPE}}}{2}) s^2. \qquad (17)$$

This is linear in all three parameters, increasing with f_{LUM} and f_{SIZE} and decreasing with f_{SHAPE}. The total motion energy from a given display sums this energy over all the items within the positive lobe of the receptive field (Fig. 13a), which by assumption includes an equal number of **a** items

		Fit of model	*Estimated parameters*		
Exp.	*Manipulation*	R^2	h	p_{STREAM}	σ
1	f_{LUM}	.9948	.722 (.011)	.205 (.013)	.115 (.007)
2	f_{SIZE}	.9600	.690 (.025)	.016 (.018)	.007 (.007)
3	f_{SHAPE}	.9900	.824 (.026)	.589 (.029)	.503 (.042)

Table 1
Summary of fits of data from Exps. 1–3 to the Bayesian model, for each experiment showing goodness of fit (R^2) and estimated values of parameters h, p_{STREAM} and σ (with asymptotic standard errors).

		Fit of model	*Estimated parameters*				
Exp.	*Features ($i \times j$)*	R^2	h	p_{STREAM}	r	σ_i	σ_j
4	luminance × size	.7923	.811 (.020)	.275 (.028)	.055 (.388)	.235 (.043)	.154 (.033)
5	luminance × shape	.9614	.806 (.008)	.364 (.014)	.459 (.123)	.189 (.014)	.705 (.065)
6	size × shape	.7974	.869 (.015)	.231 (.024)	-.617 (.652)	.125 (.094)	.784 (.636)

Table 2
Summary of fits of data from Exps. 4–6 to the Bayesian model, for each experiment showing goodness of fit (R^2) and estimated values of parameters $h, p_{\text{STREAM}}, r, \sigma_i$ and σ_j (with asymptotic standard errors).

and **b** items. By the design of the experiment, whatever parameters **a** has, **b** has values that are equally extreme but in the opposite direction. Hence every manipulation of any feature change parameter induces a linear change in the motion energy due to **a** and an *equal and opposite* linear change in the motion energy due to **b**, with zero net effect on total motion energy. Hence the motion energy from any one filter is approximately[4] constant over our entire experiment.

However it is possible to rig the spatiotemporal receptive fields in a slightly more complex way in order to give a better account of our data (Fig. 13b). This arrangement includes positive and negative lobes covering respectively the **a** and **b** parts of the path (with similar lobes nearby in opposite phase), yielding a "difference of motion energy," or motion energy differential, as the item crosses behind the occluder. This quantity seems more apt for our displays, in that it reflects how much the motion energy along the streaming path *changes* as item **a** changes to **b**. Presumably the streaming response is maximally consistent with zero motion energy differential along the streaming path (i.e. simple coherent object motion). Hence we would expect bounce responses to increase with the motion energy differential.

However, this motion energy differential model makes several predictions that are qualitatively at odds with the data. Note that motion energy increases with item size f_{SIZE} but *decreases* with the shape parameter f_{SHAPE} (Eq. 17). This means that an increase in one parameter coupled with a simultaneous decrease in the other parameter leaves energy constant, with zero change in the differential; the two feature changes "cancel each other out" from a motion energy perspective. The same applies to any two feature parameters that have opposite effects on motion energy, such as luminance and shape. Note that this is an inevitable result of the way motion energy is computed; by design it is blind to feature values per se, but simply integrates stimulation in its spatiotemporally oriented window (cf. Chubb & Sperling, 1991).

Specifically, this means that the differential model predicts a deep "valley" in the luminance × shape and size × shape data with zero mean bounce responses, despite arbitrarily large total ΔF, as the respective Δf's cancel each other out. There is no such valley in the data, and thus no evidence of this cancellation characteristic of motion energy. To test the fit of model more systematically, we regressed the mean bounce responses onto the calculated motion energy differential (using a quadratic model as in Adelson & Bergen, 1985) in the two cases where this trade-off exists, Exp. 5 and 6. The fit in Exp. 5 (luminance × shape) was $F(2,22) = 6.69, p = .0054, R^2 = .3783$; good but far weaker than the fit of the Bayesian model (again $F = 14.95, p < .000001, R^2 = .7974$). Similarly the motion energy fit in Exp. 6 was good ($F(2,22) = 6.75, p = .0052, R^2 = .3804$) but much poorer than the Bayesian model (again $F = 94.64, p < .000001, R^2 = .9614$). Thus we can reasonably conclude that the Bayesian model gives a better account of human judgments, and that our task does not primarily reflect simple motion energy.

It is worth noting that motion energy is insensitive to *any* pure shape change which does not change area or luminance, (unlike our shape parameter that does change area). This is inherent in the fact that motion energy does not encode shape features directly, but only insofar as they affect the luminance integral within the receptive field. (Indeed, this is the entire point of motion energy models—to get away from overt featural representations, and this seems to fit early motion computations well.) So for example any motion energy model

[4] In the case of size change, the net effect is only approximately zero because size changes are equal and opposite in log space, while energy depends on actual linear area (not log area). However in the case of luminance change, where luminance themselves are proportions, changes yield zero net motion energy change, incorrectly predicting constant mean bounce responses. For example a luminance pair of **a** = 55%, **b** = 45% gives sums to 100% (standard) luminance over the entire receptive field, exactly the same as a = b = 50%. The same applies to change in the shape parameter. Hence the simple motion energy model predicts no effect of luminance or shape changes by themselves, which is obviously at odds with the results of Exps. 1 and 3 respectively.

would predict equal bounce responses when **a** and **b** were both circles as when when **a** was a rabbit and **b** an (equal-size, equal-luminance) hat. This prediction seems implausible in our displays, though admittedly this extreme condition was not tested.

The role of spatiotemporal information

So far we have explicitly ignored spatiotemporal information such as the position and velocity of candidate objects. As mentioned, such information is definitely important to perceived object identity, and in fact probably dominates over featural information when the two are counterposed (Johnston & Pashler, 1990; Nissen, 1985). How might this type of information be integrated into our framework?

In our view, spatiotemporal information can be integrated into the framework we have developed above in a very seamless way by observing that spatiotemporal factors do not influence observers' object assignments, as it were, directly, but rather only via observers' *expectations* about them. That is, a spatiotemporal feature such as the object's position can be regarded as just another type of feature, in no way qualitatively distinct from other types of properties, except that the observer has particularly strong subjective expectations about its value. As in the theory so far, such subjective expectations express themselves via the evolution function Ψ. For example a strong expectation that objects ought to be stationary would be represented by a very tight (low-variance) distribution $\Psi(\Delta\mathbf{x})$ (with $\mathbf{x}$ representing spatial position). Similarly, a strong expectation that objects move in straight paths would be expressed as a very tight distribution $\Psi(\Delta\mathbf{v})$ (with $\mathbf{v}$ representing velocity). These expectations can be integrated into the evolution function Ψ simply by considering its domain to be the full feature-change space ΔF viewed as including spatiotemporal feature change as well as featural factors.

The tendency for spatiotemporal information to dominate over featural information then simply corresponds to the tendency for spatiotemporal changes to have relatively tight subjective distributions. In standard Bayesian theory, the influence of a cue turns out to depend inversely on its variance (see Box & Tiao, 1973) Thus the Bayesian observer in our model, having tight distributions around its expected spatiotemporal predictions, would consequently tend to weigh spatiotemporal factors correspondingly heavily in its object individuations. No special mechanisms or dominance rules are required.

A similar situation exists in the literature on haptic vs. visual cues, where classical studies had suggested that visual cues dominated over haptic cues in cases of conflict. A recent study (Ernst & Banks, 2002) has shown instead that subjects' behavior is consistent with a uniform Bayesian model integrating both visual and haptic cues, while the superiority of visual cues is accounted for by the relative tightness of their noise distributions (i.e., their greater reliability).

Object individuation: a more extended view

Summing up, we have established so far that the subjects observing our displays make in effect a Bayesian decision about what the most likely interpretation is: which type of event (bouncing or crossing), and thus which assignment of individual identities to the two objects, best explains the observed featural differences. We now attempt to show how this decision procedure entails what is in effect a particular "theory of objects" on the part of the observer.

To this end, in this section we recast the mathematics of the Bayesian decision procedure established in the previous section in a more complete and naturalistic setting. In particular, we assume that the bounce/stream decision in our experimental task is a proxy for the more ubiquitous decision that must be made at each point in a real image stream, where a correspondence must be subjectively established between objects in one "frame" of the stream and the next (we consider the case of a continuous image stream below). That is, we assume that subject's expectations about feature change as an object passes behind the occluder in our displays correspond closely to their expectations about feature change whenever an object in one image evolves into a subjectively co-individual object in the next image. Thus the evolution function $\Psi(\Delta F)$ refers not only to subjective expectations in the experimental displays but also, more generally, to the evolution of objects over time in a natural setting.

We emphasize that the our proposals in this section, perhaps despite appearances, actually represent only a rather modest extension of the Bayesian model discussed above. The "objects as geodesics" hypothesis presented below is a direct mathematical consequence of the properties of the Bayesian observer, except generalized to continuous time (instead of a single discrete decision as the object encounters the occluder) and to an arbitrary continuous feature space. The relevance of Mahalanobis distance (discussed below), similarly, is a mathematical entailment of the subjective dependence on the likelihood of feature change as evidenced in the experimental data. Hence unless the success of the Bayesian model in some way critically depended on the details of the experimental situation and the featural variables used, then the theory below is only a modest extrapolation of the data at hand.

Extending the Bayesian model

We begin by postulating an arbitrary feature space F, no longer limited to the three features in our experiments, but now encompassing all potential observable properties of objects in the visual field. Assume that at time t the observer sees a single object with feature vector F_0. At time $t+\Delta t$, the observer is confronted with some set $F_1, F_2, \ldots$ of possible candidate objects, each of which might be the same object as F_0 but with somewhat altered features. Of course, these objects may all be different locations and distances from the original location of F_0; we take up this issue below. For the moment assume that they are all equally plausible spatiotemporally (e.g., all equidistant from the location of F_0) so we

can restrict our discussion to the effects of featural cues.

The Bayesian model discussed above says that in this situation, the observer will perceive as the continuation of F_0 that object whose featural difference $\Delta F_i = F_i - F_0$ has the highest likelihood $p(\Delta F_i|\text{STREAM})$ (i.e., what in the experiments we would have called the "likelihood under the streaming interpretation"). Note that this does not exactly mean the *smallest* featural change per se, but rather the *least unlikely* featural change given the expected distribution of feature change. This distribution is none other than what in the previous section we called the object evolution function $\Psi(\Delta F)$. Thus the observer faced with the choice of $F_1, F_2, \ldots$ simply ought to—and by our data, will—choose the one that maximizes $\Psi(F_i - F_0)$.

Fig. 14 illustrates the situation by placing all the objects under discussion in the context of the evolution function $\Psi(\Delta F)$, illustrated schematically as a Gaussian via a contour plot. The original object F_0 is at dead center ($\Delta F = 0$). Candidates for the role of continuation of F_0 sit at various positions in feature-change space. In the example shown, F_1 is closer to F_0 in Euclidean distance: it has the minimum feature change if a step any direction in ΔF space is taken as equally important. But F_2 is closer in a *probabilistic* sense, as can be checked by examining the isoprobability contours closely: F_2 is less than two bands from F_0, while F_1 is two whole bands away. Hence F_2 has fallen less far "down the hill" from F_0, and is thus more likely under the evolution function; it represents a less-unexpected magnitude of feature change from F_0, and is thus the subjective winner as the evolution of F_0.

This fairly intuitive notion of probabilistic distance is termed the *Mahalanobis distance* in the mathematical literature (see Duda, Hart, & Stork, 2001 for an introduction). Intuitively, Mahalanobis distance is Euclidean distance scaled by probability in the underlying distribution.[5] Thus rephrasing, we can say that the Bayesian observer in our set-up simply chooses the candidate object at time $t + \Delta t$ which is at minimum Mahalanobis distance from the original object at time t. This expression of the rule emphasizes that the observer is indeed finding a "minimally-distant" extension of the original object, but doing so under a distance metric which is itself informed (and indeed determined) by subject expectations about the probability of feature change.

Extending this to a sequence of discrete times is simple. Now instead of one step, whose Mahalanobis distance we would like to minimize, we have a sequence of steps, each of which the observer would like to make as small as possible in the Mahalanobis sense. The resulting chain of steps (imagine a sequence of minimum-distance jumps from rock to rock as one crosses a river) constitutes the observer's judgment of the most likely continuous existence of the object through the world under observation. More pointedly, one can think of this chain of choices as in effect *constituting* the "object" itself: that is, a subjectively continuous stream of existence over the sequence of frames.

Another natural generalization is to consider a continuous rather than discrete progression of images over time, i.e. taking $\Delta t \to 0$. In this case the featural change ΔF becomes infinitesimal, and as a result the likelihoods of the various candidate objects will differ from each other only infinitesimally: all will have moved only infinitesimally "down the hill" of $\Psi(\Delta F)$ (see below for a more careful explanation). In this case, as in the discrete case, the choice of where the given objects' identity ought to go is still perfectly well-defined; it depends on the directions in which the candidates lie and the structure of Ψ.

Objects as geodesics

Extrapolating this to a full-fledge continuous image stream yields a particularly succinct way of stating the proposed "object concept." In the discrete version of our theory, an object is viewed as a sequence of Bayesian choices among candidate identities, such that the winning chain minimizes Mahalanobis distance through the feature space at each step. In the continuous version, an object is a continuous path through feature space such that each *infinitesimal step* minimizes Mahalanobis distance. In mathematical terminology, a minimum-length path on a curved surface is called a *geodesic* (a generalization of the notion of "straight line" appropriate for curved spaces; see below for a more technical explanation). Hence in our proposal, a subjective mental object is a *geodesic through Mahalanobis feature space* (see Fig. 15 for a schematic illustration).

A physical analogy may be helpful. Just as in relativistic gravity, where physical objects move along geodesics in a spacetime that is warped by massive objects, in our framework *psychological* objects move along geodesics in feature-space-time that is warped by subjective probability distributions. In the vivid phrase sometimes used, in physics an object is a "space-time worm;" by our theory, a *psychological* object is a minimal-length worm through a subjectively warped feature space.

A related proposal was made by Carlton and Shepard (1990), developing an earlier suggestion of Shepard (1957). They suggested that the motion path mentally interpolated between two viewed objects tends to be geodesic in psychological space (cf. Tenenbaum, de Silva, & Langford, 2000). Their proposal was primarily aimed at understanding apparent motion without feature change, but with an obvious extension to featural similarity spaces similar to that developed here.

This idea can be fleshed out formally a bit more, as follows. A continuous image stream can be thought of as func-

[5] More technically, Mahalanobis distance simply replaces the Euclidean norm, which in our notation would be expressed as

$$(\Delta F)^t (\Delta F) \tag{18}$$

with the transformed norm

$$(\Delta F)^t \Sigma^{-1} (\Delta F), \tag{19}$$

in which Σ is the covariance matrix of the subjective probability distribution $\Psi(\Delta F)$, and the superscript t indicates the matrix transpose. Thus the Mahalanobis norm is simply the Euclidean norm scaled by the (co-)variance of the underlying distribution in the given direction.

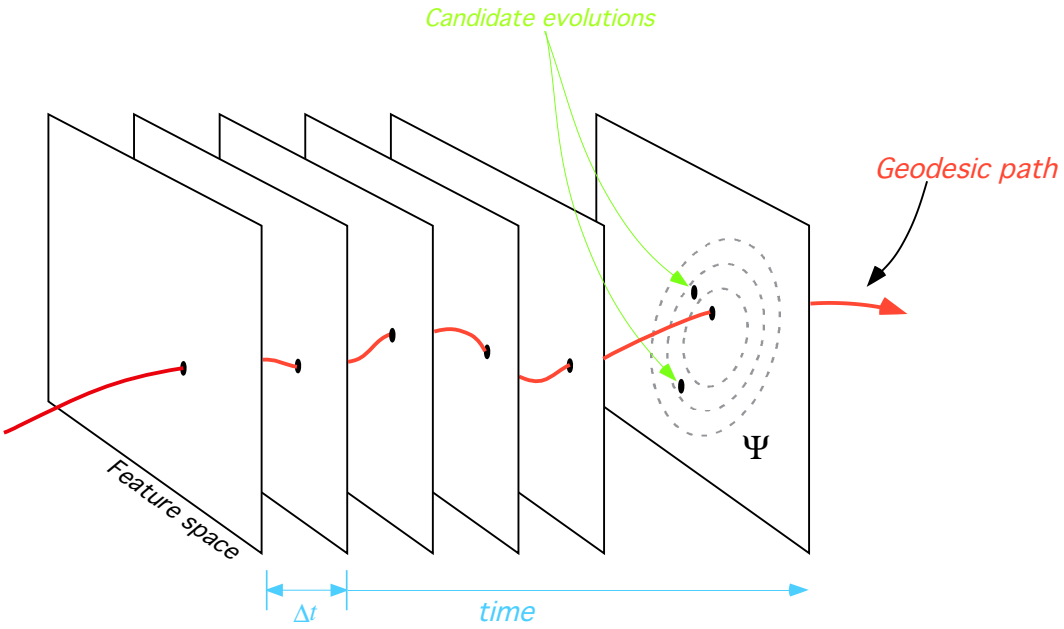

Figure 15. Schematic illustration of a geodesic through Mahalanobis feature space. At each (possibly) infinitesimal step, the observer chooses the path of minimum Mahalanobis distance from the features of the object at the previous step.

tion of space and time, $I(x,y,t)$. Running this stream through the observer's feature representation yields a featural representation $F(x,y,t)$, in which each point in space-time $\langle x,y,t\rangle$ maps to a feature vector F, which we understand to include any object features that the observer cares to represent.

Now consider an individual object located at a particular point in space-time $\langle x_0,y_0,t_0\rangle$, with feature vector $F(x_0,y_0,t_0)$. To subjectively continue this object's existence to point x_1,y_1 at time $t_1 = t_0 + \Delta t$ entails space-time motion of the object's identity, which we denote $\Delta x = \langle x_1,y_1,t_1\rangle - \langle x_0,y_0,t_0\rangle$, and an associated feature change ΔF, which as before denotes $F(x_1,y_1,t_1) - F(x_0,y_0,t_0)$. As discussed, this feature change ΔF has an associated subjective probability $\Psi(\Delta F)$, which determines its plausibility, and thus under the Bayesian model the probability with which the observer will subjectively continue the object in the direction Δx.

In the case of continuous time, we simply take the limit as $\Delta t \to 0$. The space-time motion Δx becomes a vector, denoted $\vec{x}$, meaning the instantaneous direction in which the object's identity is moving at time t_0. (In the experiments, this was forced to be either in the streaming direction or in the bouncing direction; here we consider all possible directions.) In place of the discrete featural difference ΔF, we now take the limit as $\Delta t \to 0$

$$\lim_{\Delta t \to 0} \frac{\Delta F}{\Delta x}, \tag{20}$$

which is simply the partial derivative

$$\frac{\partial F}{\partial \vec{x}}, \tag{21}$$

meaning the instantaneous change in feature vector as one moves in the direction $\vec{x}$.

The hypothesis, then, is that at each point in space-time subjective object identity moves in the direction $\vec{x}$ that minimizes the subjective probability of feature change

$$\Psi\left[\frac{\partial F}{\partial \vec{x}}\right]. \tag{22}$$

We can now state the objects-as-geodesics hypothesis more formally as follows. Each point $\langle x,y,t\rangle$ in space-time maps to a feature vector $F(x,y,t)$. Each motion $\vec{x}$ through this space-time entails a particular feature change, with associated subjective probability given by Ψ. Now, impose upon space-time the Mahalanobis metric under the probability distribution Ψ. Objects, as conceived by the Bayesian observer, are geodesics through this space:

> [Objects as geodesics]
>
> *An individual object is a geodesic through space-time under the Mahalanobis metric given the subjective probability function* $\Psi(\partial F/\partial \vec{x})$.

Again, it should be understood that this proposal is really a direct consequence, or more accurately a restatement, of the properties of the Bayesian object observer as proposed above, simply extrapolated to a continuous stream of inferences each of which is analogous to the single decision made by our subjects in the bouncing/streaming task. The geodesic characterization of objects is simply a way of capturing the idea that psychological individual objects represent subjectively maximally-probable paths through the space of possible feature changes.

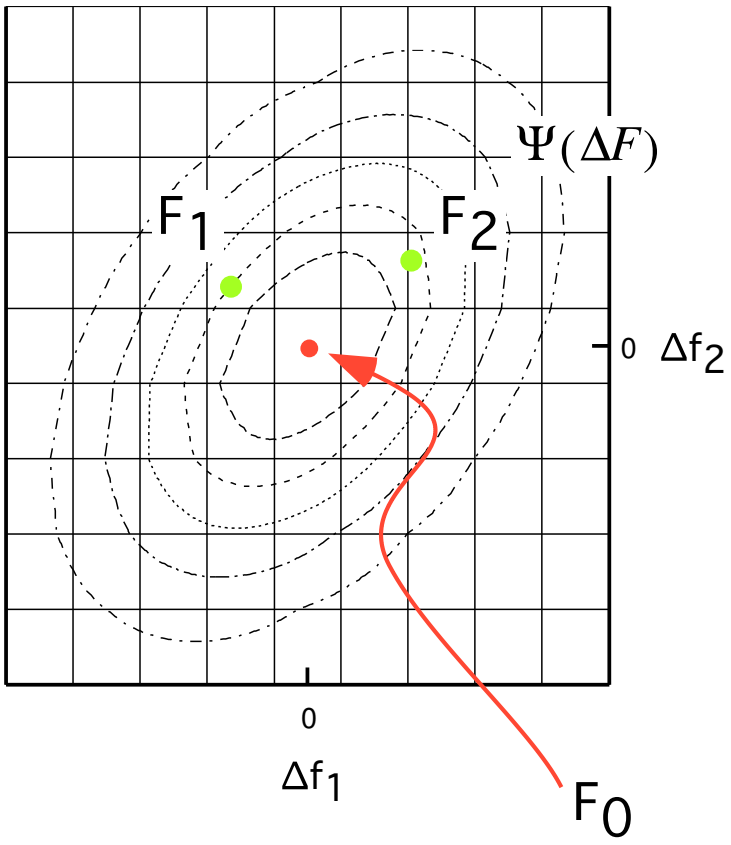

Figure 14. An illustration of object F_0 and its possible extensions $F_1, F_2, \ldots$ as situated in the evolution distribution function $\Psi(\Delta F)$ (schematically indicated as a contour plot). Given object F_0 at time t, which of $F_1, F_2, \ldots$ is perceived as the evolution of the same object? The Bayesian answer is: the one that has the highest likelihood according to the evolution function—the one that has fallen the least "down the hill," in this case F_2. As illustrated in the example, this is not necessarily the one with the smallest total feature change, but rather the one with the *least unlikely* feature change. F_1 is closer in Euclidean distance, but F_2 is closer probabilistically, as can be checked by examining the isoprobability contours in the figure. This sense of "probabilistic proximity" is called *Mahalanobis distance*.)

The object hypothesis

We conclude this theoretical discussion with one additional remark. In any sequence of images, no matter how structured or unstructured, *some* path will be of minimal length. Our object definition so far simply says the observer chooses the shortest from available alternatives. If the world is very random, the best available hypothesis may still involve a large amount of feature change at each time step—truly a "blooming, buzzing confusion," in William James's famous phrase. However, implicit in this entire scheme (and more particularly, in the feature stability assumption, which led to the assumption that the evolution function is centered at zero) is the presumption that *some* paths will in fact be much shorter (i.e. entail much less featural change) than one would expect in a totally random world. Indeed, if the world in fact did contain some objects with relatively stable properties, than some of these geodesics will be extremely short. The hypothesis that such paths *do in fact exist* is thus a version of what is sometimes called the "object hypothesis" (see Feldman, 1999; Gregory, 1970; Reynolds, 1985), and is a particularization of W. Richards' (1988) "Principle of Natural Modes." The underlying idea is that the world we inhabit does in fact contain stable entities:

> [Object hypothesis]
>
> *In the natural world, some geodesics in Mahalanobis feature space will be short.*

This assumption is not necessary for our scheme to be well-defined: again, even if the world were random (and thus did not obey the object hypothesis), *some* path would be shortest. But something like this assumption is necessary in order for the Bayesian decision scheme to be a sensible one. Without it, the Bayesian observer would be choosing among hypotheses *none* of which actually corresponds to a stable object as hoped. That the object hypothesis is implicitly believed by human observers is testified by our data, which demonstrate that subjects do have evolution functions centered near zero, meaning that they do expect evolving individual objects to have stable properties. Without such a hypothesis in their mental arsenal, our subjects' pattern of responses makes little sense.

Conclusion

In summary, our experiments suggest that human observers form a correspondence between items in successive time-slices—and thus create a representation of individual objects bearing continuous existence—by determining the most plausible featural correspondence given subjective expectations about objects are likely to change over time. These expectations, encoded as a subjective probability distribution (our "evolution function"), are then combined in a simple way, via Bayes' rule, to establish object individuation. Again, our model is expressed in terms of featural properties, because these were the only ones that were informative in our displays; spatiotemporal properties would probably have dominated were they useful to the observer, but in our studies they were completely ambiguous. However, as discussed above, spatiotemporal properties could be incorporated into the Bayesian observer model, and thus into our "object concept," without substantially altering it.

In everyday conception, the individuation of physical objects is often described as if there were an objective, physical fact of the matter: one object is at a certain time is regarded as *in fact* the same as another at a previous time, in virtue of continuous intervening existence. Considering the bouncing/streaming task, however, it becomes apparent that the very notion of "continuous intervening existence" has a subjective element. One must *decide* whether existence has in fact intervened continuously; and in doing so, all one has to work with are observable properties. Our proposal is that

this situation extends to object individuation generally, not just when the choices are made artificially ambiguous as in our laboratory. All object individuation, in the end, is based on observables, and no further ground truth is available.

Our data suggest that mental representation of object individuation depends in fact on subjective apprehension of how things change (the object evolution function), and in particular, on the subjective expectation that real physical objects tend not to change too much (the feature stability assumption). These assumptions lead via Bayes' rule to a very specific quantitative prediction of how choices will be made when individuation is rendered ambiguous, as in our experimental task, which are borne out by the data. The conclusion is that human observers follow a very reasonable strategy when individuating objects, one based on making the best guess possible given the data available.

References

Adelson, E., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A.*, *2*(2), 284–299.

Anstis, S. (1980). The perception of apparent movement. *Phil. Trans. R. Soc. Lond. B*, *290*, 153–168.

Baillargeon, R. (1987). Object permanence in 3 1/2 and 4 1/2 month-old infants. *Developmental Psychology*, *23*(5), 655-664.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science*, *3*(5), 133–140.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* Chichester: John Wiley & Sons.

Bertenthal, B. I., Banton, T., & Bradbury, A. (1993). Directional bias in the perception of translating patterns. *Perception*, *22*, 193–207.

Blake, A., & Zisserman, A. (1987). *Visual reconstruction.* New Cambridge: M.I.T. Press.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, *408*, 196–199.

Box, G. E. P., & Tiao, C., George. (1973). *Bayesian inference in statistical analysis.* Reading, Massachussetts: Addison-Wesley.

Bülthoff, H. H., & Yuille, A. L. (1991). Bayesian models for seeing shapes and depth. *Comments on Theoretical Biology*, *2*(4), 283–314.

Burt, P., & Sperling, G. (1981). Time, distance and feature trade-offs in visual apparent motion. *Psychological Review*, *88*(2), 171–195.

Carlton, E., & Shepard, R. N. (1990). Psychologically simple motions as geodesic paths: I. Asymmetric objects. *Journal of Mathematical Psychology*, *34*, 127–188.

Chubb, C., & Sperling, G. (1991). Texture quilts: basic tools for studying motion-from-texture. *Journal of Mathematical Psychology*, *35*, 411–442.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification.* New York: John Wiley and Sons, Inc.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Feldman, J. (1999). The role of objects in perceptual grouping. *Acta Psychologica*, *102*, 137–163.

Feldman, J. (2001). Bayesian contour integration. *Perception & Psychophysics*, *63*(7), 1171–1182.

Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in space-time. *Proceedings of the National Academy of Science*, *97*, 8186–8191.

Gregory, R. L. (1970). *The intelligent eye.* New York: McGraw-Hill.

Johnston, J. C., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 843–856.

Julesz, B. (1995). *Dialogues on perception.* Cambridge: M.I.T. Press.

Knill, D., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference.* Cambridge: Cambridge University Press.

Kubovy, M. (1994). The perceptual organization of dot lattices. *Psychonomic Bulletin and Review*, *1*(2), 182–190.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*(3), 389–412.

Lu, Z.-L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: review and update. *Journal of the Optical Society of America A*, *18*(9), 2331–2370.

Michotte, A. (1946/1963). *The perception of causality.* New York: Basic Books.

Navon, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology*, *2*, 130–138.

Nissen, M. J. (1985). Accessing features and objects: is location special? In M. Posner & O. Marin (Eds.), *Attention and performance* (Vol. XI, pp. 205–220). Hillsdale, NJ: Erlbaum.

Prazdny, K. (1986). What variables control (long-range) apparent motion? *Perception*, *37*, 37–40.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 1–19.

Ramachandran, V. S., & Anstis, S. M. (1983). Extrapolation of motion path in human visual perception. *Vision Research*, *23*, 83–85.

Reynolds, R. I. (1985). The role of object-hypotheses in the organization of fragmented figures. *Perception*, *14*, 49–52.

Richards, W. A. (1988). The approach. In W. A. Richards (Ed.), *Natural computation* (pp. 3–13). Cambridge: M.I.T. Press.

Scholl, B., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290.

Sekuler, A. B., & Sekuler, R. (1999). Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? *Perception*, *28*, 415–432.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.

Shechter, S., Hochstein, S., & Hillman, P. (1988). Shape similarity and distance disparity as apparent motion correspondence cues. *Vision Research*, *28*(9), 1013–1021.

Shepard, R. N. (1957). Stimuls and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605–632.

Spelke, E. S., Kestenbaum, R., Simons, D., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, *13*(2), 113–142.

Tarr, M., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*, 233–282.

Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*, 2319–2323.

Tremoulet, P. D., Leslie, A. M., & Hall, G. (2000). Infant attention to the shape and color of objects: Individuation and identification. *Cognitive Development*, *15*, 499–522.

Ullman, S. (1977). Transformability and object identity. *Perception & Psychophysics*, *22*, 414–415.

Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, *27*, 1041–1054.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, *30*, 111–153.

Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science*, *6*(3), 182–186.