Journal of Experimental Psychology: General
2006, Vol. 135, No. 4, 513–541

0096-3445/06/$12.00 DOI: 10.1037/0096-3445.135.4.513

A Theory of Dynamic Occluded and Illusory Object Perception

Evan M. Palmer and Philip J. Kellman
University of California, Los Angeles

Thomas F. Shipley
Temple University

Humans see whole objects from input fragmented in space and time, yet spatiotemporal object perception is poorly understood. The authors propose the theory of spatiotemporal relatability (STR), which describes the visual information and processes that allow visible fragments revealed at different times and places, due to motion and occlusion, to be assembled into unitary perceived objects. They present a formalization of STR that specifies spatial and temporal relations for object formation. Predictions from the theory regarding conditions that lead to unit formation were tested and confirmed in experiments with dynamic and static, occluded and illusory objects. Moreover, the results support the identity hypothesis of a common process for amodal and modal contour interpolation and provide new evidence regarding the relative efficiency of static and dynamic object formation. STR postulates a mental representation, the *dynamic visual icon,* that briefly maintains shapes and updates positions of occluded fragments to connect them with visible regions. The theory offers a unified account of interpolation processes for static, dynamic, occluded, and illusory objects.

Keywords: dynamic occlusion, motion, object perception, contour interpolation, illusory contours

A distinguishing characteristic of organisms possessing sophisticated visual systems is that they move. Guiding locomotion and conveying information about events are among the most fundamental functions of vision (Gibson, 1966, 1979). Accordingly, one would expect the visual systems of humans and other mobile animals to extract information from the constantly changing visual stimulation produced by object and observer motion. Indeed, such processes are evident in ordinary visual activity. For example, a moving person, vehicle, or animal may be perceived through foliage despite the fact that the shape information reaching the eyes is fragmented in both space and time. Optical changes given by self-motion are even more pervasive. We move our eyes, turn our heads, walk, run, or drive. These events place important demands on perception but also enable active perceptual exploration. This point was made eloquently by Gibson (1979), who wrote,

> One sees the environment not just with the eyes but with the eyes in the head on the shoulders of a body that gets about. We look at details with the eyes, but we also look around with the mobile head, and we go-and-look with the mobile body. (Gibson, 1979, p. 222)

Motions of objects and observers thus produce both challenges and opportunities for perception. One fundamental challenge is that they produce constantly changing patterns of occlusion. As objects in ordinary environments are often partially occluded by other objects, perceiving depends not only on utilizing fragmented information but on coping with continuously changing visible and invisible regions of objects and surfaces. Ordinary seeing suggests that perceivers routinely recover and represent the unity and shape of objects despite partial occlusion and motion. The processes involved, however, are poorly understood.

Considerable progress has been made in understanding the geometric relations and processes that allow visible areas of objects to be connected under occlusion in stationary scenes (e.g., Fantoni & Gerbino, 2003; Field, Hayes, & Hess, 1993; Geisler, Perry, Super, & Gallogly, 2001; Kellman & Shipley, 1991; for a review, see Shipley & Kellman, 2001), and neural-style models have been proposed suggesting how interpolation may be computed in neural circuitry (e.g., Heitger, von der Heydt, Peterhans, Rosenthaler, & Kubler, 1998; for a review, see Neumann & Mingolla, 2001). Although perception of partly occluded objects has been studied with moving stimuli (e.g., Kellman & Spelke, 1983), such research has usually examined the perceptual integration of visual areas that are simultaneously visible. In the real world, however, this is often not the case.

We call any object that moves behind an occluding surface and is seen through apertures *dynamically occluded.*[1] Dynamic occlusion arises from relative motions between objects, other objects that occlude them, and observers. In many cases, visible areas belonging to a single object are received over time. Furthermore, some regions of dynamically occluded objects may never project to the eyes, necessitating interpolation processes that connect visible regions to produce accurate percepts of unity and shape. How does the visual system recover the connectivity and shape of dynamically occluded objects from spatially and temporally frag-

Evan M. Palmer and Philip J. Kellman, Department of Psychology, University of California, Los Angeles; Thomas F. Shipley, Department of Psychology, Temple University.

We gratefully acknowledge support of this research from National Eye Institute Grant EY13518 and National Science Foundation Grant SBR9496112 to Philip J. Kellman and Thomas F. Shipley and a National Science Foundation Graduate Fellowship to Evan M. Palmer.

Correspondence concerning this article should be addressed to Evan M. Palmer, who is now at the Visual Attention Laboratory, Brigham and Women's Hospital and Harvard Medical School, 64 Sidney Street, Suite 170, Cambridge, MA 02139. E-mail: palmer@search.bwh.harvard.edu

---

[1] For the purposes of this discussion, the terms *kinetic* and *dynamic* are regarded as synonymous.

mented patterns of stimulation? This question is important in its own right and is also an example of a basic and pervasive problem for the visual system: how to assemble accurate representations of reality from partial information.

Consider an example. Suppose one views a street scene from a window and through foliage, as in the left column of Figure 1. As a fire truck drives by, gaps in the trees allow only bits and pieces of the truck's shape to reach the eyes at any instant. One might see a fragment of the truck's front in the left visual field, as in Frame A; part of a door in the center of the visual field, as in Frame B, and then on the right, as in Frame C; and, finally, part of a wheel in the right visual field, as in Frame D. Furthermore, these fragments are seen through vertically and horizontally misaligned apertures, and a large horizontal region of the fire truck is never projected to the eyes due to occlusion by the window frame. Nonetheless, these fragments are perceived not as disconnected parts but as a single object—a fire truck—driving down the street.

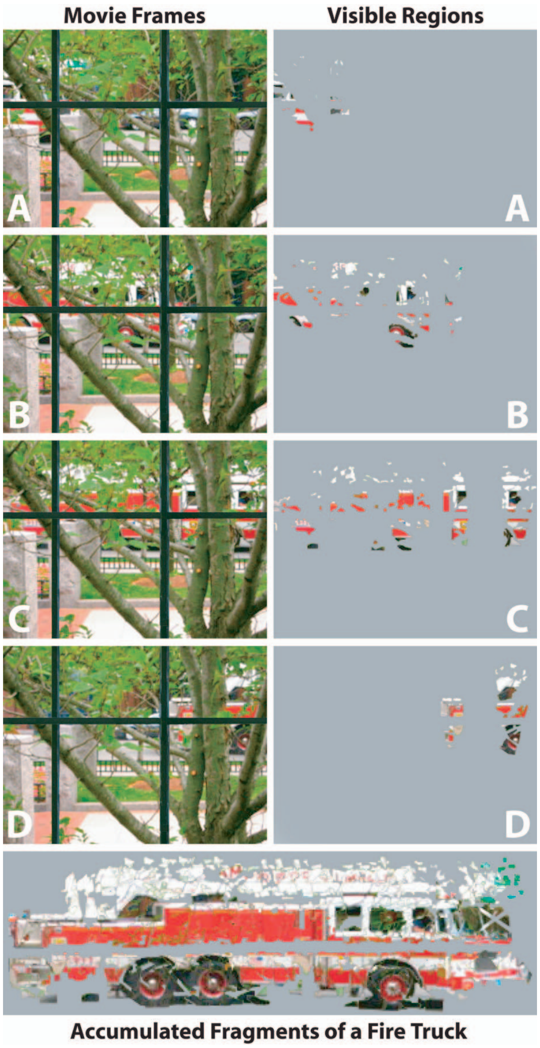

Figure 1. Several momentary views of a heavily occluded scene containing a moving fire truck. Each view of the scene (left column) reveals fragmentary bits of objects, such as the fire truck. Mere accumulation of visible parts over time would not allow recovery of objects. The right column is a graphical depiction of the theory of spatiotemporal relatability. A dynamic visual icon representation generated through processes of persistence and position updating could preserve visible areas and update their positions over time after occlusion, consistent with extracted velocity information. Bottom: Accumulation of positionally updated visible regions allows interpolation processes characteristic of static scenes to occur. These spatiotemporal integration and interpolation processes allow perception of coherent objects, such as the fire truck.

Some evidence suggests that human object and scene perception under these conditions is quite good. Classic studies have shown that observers can perceive the shape of an object when it passes behind a narrow slit (Helmholtz, 1867/1962; Parks, 1965; Plateau, 1836; Zöllner, 1862; see below). Vishton, Kellman, and Shipley (1991) showed that brief viewing of a static scene through a moving occluder with small apertures was sufficient to allow accurate answers to queries about the presence or absence of specific objects even when only a very small percentage (e.g., 10%) of the scene was projected (across all frames taken together). In a classic paper, Hochberg (1968) argued for a form of post-retinal visual storage that could account for findings such as the fact that observers could distinguish between possible and impossible plus-shaped figures after seeing successive frames of a single corner at a time through a small aperture. It seems that the visual system naturally integrates fragmented shape information over time into coherent and accurate representations of objects.

Mateeff, Popov, and Hohnsbein (1993) reported that participants were able to accurately perceive objects that moved behind an occluding surface containing many small, separated pinholes. Perception of a continuously moving figure through a black surface with tiny apertures was better than perception of the same figure when it was presented at random locations behind the occluding surface. This result suggests that continuous motion engages specific visual processes that support object formation in dynamically occluded displays.

What relations in space and time allow perception of complete objects from fragmentary information? Little, if any, research has addressed this question for dynamic interpolation situations in which some parts of an occluded object are never projected to the eyes. In such a situation, integration of information is important, but interpolation is also required to connect visible fragments that have appeared at different times and places. Given the pervasive interactions of observer motion, object motion, and occlusion, no understanding of object formation can be considered complete without grappling with these issues.

The theory and experiments reported here focus on how the visual system assembles spatially discontinuous shape information given sequentially in time to perceive dynamically occluded objects. Specifically, we sought (a) a geometric account of the spatial and temporal relations between visible regions that allow their interpolation across gaps to form perceptual units and (b) an account of the perceptual processes that support the use of these stimulus relations. The result is a new theory of contour interpolation and visual unit formation that handles spatially and temporally fragmented input resulting from occlusion and object or observer motion. In the perspective that emerges, static stimuli appear as a special case of dynamic stimuli with zero temporal separation and velocity. Additionally, this theory (and the data of Experiment 3) extends the *identity hypothesis* (Kellman, Yin, & Shipley, 1998) to the dynamic domain, proposing that contour completion for occluded objects and contour completion for illusory objects share a common interpolation mechanism. The theory of contour interpolation proposed here addresses static, dynamic, occluded, and illusory stimuli in a unified manner.

As becomes evident below, a fruitful approach to dynamic object formation is to build upon extant geometric accounts of visual unit formation for static occluded objects. To provide the context for a theory of spatiotemporal object formation, we first describe some aspects of unit formation in static scenes.

Visual Unit Formation in Static Scenes

In static scenes, contour interpolation across occluded regions depends on particular spatial relationships between visible contours. Michotte, Thinès, and Crabbé (1964) suggested that a number of Gestalt principles of organization, such as good continuation, good form, and closure, could be applied to the problem of partial occlusion. Investigations by Kanizsa (1979) suggested that local contour continuity dominates other influences in visual segmentation and grouping. More global variables, such as good form, symmetry, familiarity, or simplicity of the outcome, prove weak or ineffective when pitted against local continuity.

Kellman and Shipley (1991) proposed a geometric account of the stimulus variables in static, two-dimensional (2D) scenes that support contour interpolation. Interpolation is triggered by contour junctions (more formally, *tangent discontinuities*) in the optical projection, such as the points at which an object's edge reaches an occlusion boundary. Pairs of contours leading into tangent discontinuities are connected by the visual system if they fulfill the geometric conditions of *relatability*. We briefly explain each of these notions.

Contour interpolation begins and ends at tangent discontinuities (Rubin, 2001; Shipley & Kellman, 1990), which are points where a contour has no unique slope (e.g., a contour junction or corner). Whereas a zero-order discontinuity would be a spatial gap in a contour, a first-order or tangent discontinuity is a point at which the direction of the contour changes abruptly. Besides first-order discontinuities, some have suggested that second-order discontinuities (as where a straight segment joins a constant curvature segment, with the slopes matching at the join point) might also play a role in triggering interpolation (Albert, 2001; Albert & Hoffman, 2000; Albert & Tse, 2000; Shipley & Kellman, 1990; for recent discussion, see Kellman, Garrigan, & Shipley, 2005). The importance of tangent discontinuities in visual processes coping with occlusion stems from an ecological invariant: It has been proven that when one object occludes another, the intersection of the two objects' contours generically forms a tangent discontinuity in the visual projection (see Kellman & Shipley, 1991, Appendix A). In illusory contour displays (which appear to share a common edge interpolation process with occluded displays; see Experiment 3, below), the presence or absence of tangent discontinuities can be manipulated by rounding the corners of inducing elements. Experimental evidence indicates that this manipulation reduces or eliminates contour interpolation (e.g., Albert & Hoffman, 2000; Kellman, Garrigan, Shipley, Yin, & Machado, 2005; Shipley & Kellman, 1990).

Among contours leading into tangent discontinuities, interpolation operates only for relatable contours. As with the principle of good continuation, proposed to explain segmentation of contiguous arrays (Wertheimer, 1923), relatability has at its heart a smoothness constraint. (For a recent discussion of good continuation and its relation to relatability, see Kellman, Garrigan, Kalar, & Shipley, 2003.) Relatability is defined formally (Kellman, Garrigan, & Shipley, 2005; Kellman & Shipley, 1991), but it can be described intuitively as follows. Two contours are relatable if they can be connected by a smooth curve that bends monotonically through no more than about 90°. A number of empirical studies of the contour interpolation process have suggested that the geometric notion of relatability describes the conditions under which observers perceive fragmented regions of objects as unified into a single perceptual unit for both occluded and illusory objects (e.g., Field et al., 1993; Kellman, Garrigan, Shipley, et al., 2005; Kellman & Shipley, 1991; Kellman et al., 1998; Saidpour, Braunstein, & Hoffman, 1994).

Analyses of natural scenes by Geisler et al. (2001) have suggested that relationships between partially occluded edges arising from the same object tend to be consistent with the conditions of relatability. Contour grouping mechanisms may incorporate a geometry of unit formation that matches in important respects the statistics of natural scenes as projected to the eyes.

In the foregoing review of static completion processes, we have emphasized contour interpolation, as it is most relevant to the following theory and results on spatiotemporal interpolation. Kellman and Shipley (1991) proposed that, in addition to contour processes, a process of surface interpolation under occlusion connects visible areas, a claim that has received experimental support (Yin, Kellman, & Shipley, 1997, 2000). It has also been proposed that inferred volumes may play a role in object formation (Tse, 1999). A discussion of the complementary roles of contour and surface processes, as well as the possibility of volume completion in static, three-dimensional objects, can be found in Kellman, Garrigan, and Shipley (2005).

Extending Contour Interpolation Over Time

In the ecology of human perception, static scenes constitute a limiting case. Given the combination of occlusion and motion of objects and observers, perception of objects despite fragmentary information requires a spatiotemporal interpolation process.

What must a spatiotemporal interpolation process accomplish? The problem is illustrated in Figure 2. The figure shows how an object might appear as it passes behind an occluder if shape information were merely accumulated over time (i.e., if fragments visible at different times were retained in the positions at which they appeared). Juxtaposed in a single frame, these fragments appear as a bizarre hodgepodge. The problem of spatiotemporal unit formation is how to recover the actual shape of a moving object from fragmentary information of this kind.

The accumulated views in Figure 2 represent a simplified picture of temporal summation. In real occlusion situations, surface regions of the object would sweep past a given aperture, filling all locations visible through it over time. Yet even the few views shown in the simplified figure present quite a poor assemblage of shape information; filling an entire aperture with object color would be even less informative. These considerations motivate the question, What relations between momentarily viewed object fragments can be used to recover the connectivity and shapes of dynamically occluded objects in the scene?

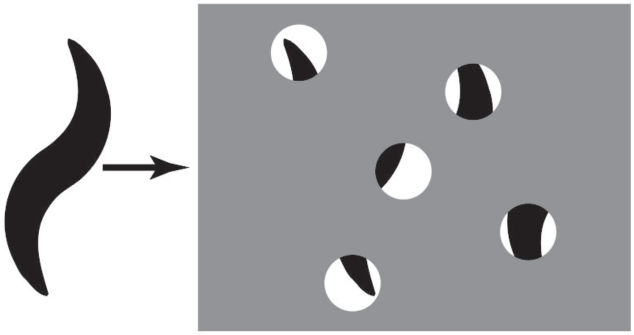

Figure 2. Successive views of a dynamically occluded object. (Only some views are shown.) Without a representation of object motion, the temporally summed inputs to the spatial relatability computation would consist of all positions of the object behind the occluding surface that are visible through the apertures.

Although the question seems dauntingly open ended, it turns out there are ways to extend what is known about static contour interpolation to answer it. In spatiotemporal object perception, the visual system must integrate information not only across space but over time. Suppose one assumes that the relevant spatial relations between fragments apply to both static and dynamic cases but that, in the latter, motion is somehow taken into account. This is the basic intuition behind what we call *spatiotemporal relatability* (STR).

The ability of the human visual system to perceptually integrate edges that are revealed over time via object motion was discussed by Kellman and Shipley (1991, 1992), but they neither proposed nor tested any specific theoretical account. A variety of experimental efforts have examined the role of motion relationships in determining perceptual grouping, including the effect of common motion in object formation (e.g., Johnson et al., 2003; Kellman, Gleitman, & Spelke, 1987; Kellman & Spelke, 1983), the role of motion information in perceiving illusory and occluded stimuli (Anderson & Barth, 1999; Bruno & Bertamini, 1990; Kellman & Cohen, 1984; Lorenceau & Shiffrar, 1992, 1999; Shipley & Cunningham, 2001), and the perception of boundaries and form in the absence of oriented edge inputs (Bruno, 2001; Cunningham, Shipley, & Kellman, 1998a, 1998b; Shipley & Kellman, 1992, 1994, 1997).

Though contour relatability has been mentioned in these contexts, it is important to note that, as proposed by Kellman and Shipley (1991), relatability does not have the ability to predict contour interpolation in dynamically occluded objects. The shortcomings of spatial relatability with regard to these situations are twofold: (a) The theory does not describe the ability to accumulate edge information over time, and (b) the theory does not incorporate information about the motion of edges behind an occluding surface. For the theory of relatability to apply to dynamically occluded objects, it must be extended.

Spatiotemporal Relatability: A Theoretical Framework for the Perception of Dynamically Occluded Objects

Two hypotheses allow the spatial geometry of relatability to be broadened into the concept of STR for handling dynamically occluded objects. Figure 3 may help to illustrate them. We propose first a *persistence hypothesis*: Briefly viewed object fragments remain perceptually available for a short time after occlusion so that they can be integrated with later-appearing fragments. In Figure 3A, an occluding surface with two apertures moves in front of an object, revealing one part of the object at time t_1 and another part at time t_2. If shape information from the object fragment seen at t_1 persists in a perceptual buffer until the second part appears at t_2, then the standard spatial relatability computation can be performed on both the visible and occluded regions to connect them into a single perceptual unit.

Persistence alone may be helpful in the case of a moving occluder but is not sufficiently general. If the occluding surface remains stationary and the object moves behind it (as depicted in Figure 3B), the same two parts of the object are again revealed, but at different locations within the visual field. Perceptually connecting two such parts in the positions at which they were visible would result in a bizarre outcome, as Figure 2 illustrated. For the fragment seen at t_1 to be appropriately related to the one seen at t_2, the current position of the t_1 fragment must be used when the t_2 fragment appears. Accordingly, we propose the *position updating hypothesis*: The visual system maintains some representation of the velocity of the occluded region of the object, in addition to its shape and edge orientations, and uses this velocity signal to represent the changing position of the occluded region as it moves behind the occluding surface. Subsequently, when the second part of the object appears at t_2, it can be integrated with the updated position of the first part from t_1 according to the standard spatial relatability computation.

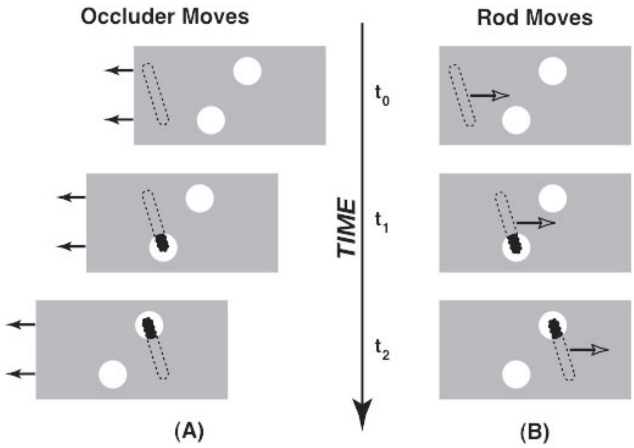

Figure 3. Components of spatiotemporal relatability. A: Persistence—The moving occluder reveals relatable parts of the rod sequentially in time (t_1 and t_2). Perceptual integration of object fragments requires that some representation of the initially visible part persist over time. B: Persistence and position updating—Parts of the moving rod become visible through apertures sequentially in time. Perceptual integration of the parts requires not only persistence but also position updating based on velocity information. From "Geometric and Neural Models of Object Perception," by P. J. Kellman, S. E. Guttman, and T. D. Wickens, in *From Fragments to Objects: Segmentation and Grouping in Vision* (p. 238), edited by T. F. Shipley and P. J. Kellman, Amsterdam: Elsevier. Copyright 2001. Reprinted with permission.

It should be noted that although persistence and position updating are logically distinct hypotheses, it is likely that they work in a coordinated fashion. Even cases of zero position change in which object parts are revealed over time (as when occluders move) may involve position tracking; in such a case, position change has the value zero. In the General Discussion, we propose a unified representation—the dynamic visual icon—that incorporates both persistence and position updating. Below, we review evidence for the persistence and position updating hypotheses, as well as for their role in STR.

Persistence

The persistence hypothesis—that visual interpolation can connect currently and previously visible regions—is novel with respect to current models of contour interpolation. The general notion that visual information remains perceptually available after its offset, however, is neither new nor controversial. The classic experiments of Sperling (1960) were among the earliest to document the continuing visibility of displays after their physical termination. The perceptual representation responsible for Sperling's effect was termed the *visual icon* by Neisser (1967). In a landmark review and synthesis of research on the visual icon, Coltheart (1980) proposed that information from an extinguished display remains active in the visual system in several forms, including neural persistence (e.g., Duysens, Orban, & Cremieux, 1985; Kratz & May, 1990), visible persistence (e.g., Di Lollo, 1980; Dixon & Di Lollo, 1994; Efron, 1970; Haber & Standing, 1969), and iconic memory (e.g., Di Lollo & Dixon, 1988; Irwin & Yeomans, 1986).

Many experimenters have investigated the contents of the visual icon using Sperling's (1960) poststimulus sampling technique. Among the properties that have been established to exist in iconic memory following stimulus termination are color (Banks & Barber, 1977; Clark, 1969), size and orientation (Von Wright, 1968), shape (Turvey & Kravetz, 1970), and stimulus motion direction (Demkiw & Michaels, 1976; Shioiri & Cavanagh, 1992).

Of particular relevance to this discussion are the findings that both motion and shape information continue to be available to the perceiver after stimulus termination. As we show below, the perceptual persistence of such information beyond stimulus termination may enable visual processing that overcomes spatial and temporal gaps in dynamically occluded object perception.

Position Updating

The position updating hypothesis expresses the idea that the spatial positions of occluded but persisting object fragments are updated over time as an object moves behind an occluding surface. These occluded, persisting, positionally updated fragments can then be combined with currently visible regions to allow computation of object unity and shape (as in the bottom panel of Figure 1). Position updating can be accomplished through smooth pursuit eye movements, or it may be computed using velocity information extracted from shape fragments while they are visible through apertures. Below, we review several lines of evidence suggesting that the positions of moving objects are actively updated for a short time after they become occluded.

Hochberg (1968) carried out studies in which observers saw a plus-shaped outline figure rotate behind a circular aperture in a series of discrete frames. The figure was depicted as three dimensional, in either a possible or an impossible form. Even though only fragmentary shape information was available through the aperture at any given time, participants were able to integrate the visible regions into an accurate representation of shape and correctly judge whether the line drawings were possible or impossible. Hochberg proposed that the successive views of the figures were integrated in a *schematic map* that stored shape information and updated it according to the object's motion observed through the aperture.

Classic experiments on the *tunnel effect* (Michotte, 1950) involved stimulus sequences in which one small dot moved behind an occluding surface (the tunnel) and then, after an appropriately short delay, a second small dot emerged on the other side of the tunnel at the same speed. Under these conditions, participants reported perceiving the two moving dots as the same object and made very specific reports about the apparent trajectory and path shape of the dot behind the tunnel (Burke, 1952). These results are consistent with the idea that the dots on either side of the tunnel were perceptually represented as a single entity whose position was perceived behind the tunnel even though it was not physically visible.

More evidence consistent with the position updating hypothesis comes from Shioiri, Cavanagh, Miyamoto, and Yaguchi (2000), who investigated participants' ability to judge the position of an apparent motion stimulus during the interstimulus interval between presentations of the dot stimuli. They found that participants were quite accurate at judging the interpolated position of the dot even when queried at a place and time when no stimulus was actually on the screen. Their results suggest that the positions of objects are perceptually represented even when they are not physically specified.

Scholl and Pylyshyn (1999) used a multiple object tracking paradigm and found that physical transformations of tracked elements that provided occlusion cues of accretion and deletion (Gibson, Kaplan, Reynolds, & Wheeler, 1969) enabled participants to track the stimuli during a period when they were physically absent. However, if the tracked elements either disappeared or shrank at an occlusion boundary, then participants were unable to reliably track them behind the occluder. This difference indicates that particular stimulus transformations, such as accretion–deletion information, are important for the continued representation of hidden elements (Gibson et al., 1969; Michotte et al., 1964). Thus, when given information for occlusion, the visual system appears to maintain a representation of an object after it becomes hidden. Note also that the high level of tracking performance in the accretion–deletion condition could not have been accomplished solely with pursuit eye movements because it is impossible to track simultaneously the independent motions of four disks with only two eyes.

Studies of *anorthoscopic perception* furnish evidence for both the visible persistence and the position updating hypotheses. (The curious term anorthoscopic, used by Plateau, 1836, means abnormally viewed.) Perception researchers long ago reported that observers are able to perceive the entire shape of an object when it moves behind a narrow slit, even though only a small fraction of the object is visible at any given instant

(Helmholtz, 1867/1962; Plateau, 1836; Zöllner, 1862). Several findings from the anorthoscopic perception literature imply the involvement of persistence and position updating. First, the perceived figure can be far larger than the slit through which it is observed (e.g., Anstis & Atkinson, 1967; Parks, 1965; Zöllner, 1862), suggesting that shape information persists and is accumulated over time. Second, the anorthoscopic image appears outside the slit, farther along in the direction of its motion (Sohmiya & Sohmiya, 1994), consistent with the possibility that the positions of persisting shape parts are updated over time according to their previously extracted velocity information. Third, the entire anorthoscopic image is perceived as a complete figure despite the fact that only parts of the image are visible at any given instant (Parks, 1965; Rock, Halper, DiVita, & Wheeler, 1987; Sohmiya & Sohmiya, 1994), suggesting that visual unit formation operates on both currently visible and recently occluded portions of an object's shape.

It is important to distinguish the anorthoscopic perception studies mentioned above from the present experiments testing spatiotemporal interpolation. In anorthoscopic displays, the entire figure is projected to the eyes over time as it moves behind a single aperture. In spatiotemporal interpolation and in the dynamic occlusion displays we used in particular, the figure is much larger vertically than any of the apertures, and the apertures are offset and misaligned. These conditions involve integrating information across several narrow, misaligned apertures. Crucially, large regions of the dynamically occluded objects are never projected to the eyes at all; in contrast to anorthoscopic perception, this task requires interpolation across unspecified regions in the image.

What representations and processes might accomplish position updating? It seems intuitively clear that smooth pursuit eye movements are often involved when perceiving dynamically occluded objects. To preview, evidence suggests that eye movements are not necessary for position updating. However, there are other questions of what representations might be involved and in what reference frame—retinotopic (eye-centered) or distal (environment-centered)—position and shape information might be encoded.

Shipley and Cunningham (2001) argued that two types of representations are used in dynamic occlusion: (a) a retinotopic representation (frequently and inaccurately referred to as retinal painting), where spatial relations are passively received during pursuit eye movements and encoded with reference to the retina, and (b) a distal representation that is abstracted away from the retinal reference frame and in which spatial relations must be updated mentally to represent current spatial positions. They cited evidence for both representations and delineated the conditions that favor one over the other (see also Morgan, Findlay, & Watt, 1982). If STR were implemented using only a retinotopic representation, position updating could be accomplished solely by pursuit eye movements. Using the distal representation, on the other hand, visual unit formation could be accomplished via persistence, relatability, and position updating not dependent on eye movements.

Consider again the example of perceiving a fire truck through foliage (see Figure 1). It seems natural that one's eyes might pursue moving visible fragments of the fire truck, tracking its current position and allowing accumulation over time of its visible regions. In this example, persistence would allow occluded and unoccluded regions to be combined, and position updating via eye movements would keep them properly aligned within a retinotopic reference frame. Several experiments have suggested that sensory memory for position (i.e., persistence) is maintained in a retinotopic reference frame during smooth pursuit eye movements of this type.

Kerzel and Ziegler (2005) showed that apparent motion is computed between successive moving frames during a smooth pursuit eye movement, indicating that sensory memory integrates information between the two images. They also found that performance for stationary targets with fixation was equivalent to moving targets with pursuit eye movements in a visual short-term memory task. Their findings are consistent with a retinotopic reference frame for smooth pursuit eye movements (see also Sun & Irwin, 1987, for a similar conclusion).

Despite being sufficient for position updating, pursuit eye movements are not necessary to perceive shapes moving through an aperture (Fendrich, Rieger, & Heinze, 2005; Haber & Nathanson, 1968; Parks, 1965; Rock & Halper, 1969; Shipley & Cunningham, 2001). Research in support of a distal representation for position updating (see Shipley & Cunningham, 2001, for a review) has shown that it is possible to perceive the form of an occluded object when the eyes move in such a way as to make any retinotopic representation useless. For example, Haber and Nathanson (1968) used a stationary object that was seen through a moving slit and showed that participants could perceive the object even when they tracked the slit with their eyes. In that case, the shape information from the object was always projected onto the same part of the retina, making it impossible for a retinotopic representation to be solely responsible for the position updating necessary to perceive the form correctly (Shipley & Cunningham, 2001).

Recently, Fendrich et al. (2005) investigated observers' ability to perceive the shapes of anorthoscopically presented figures either with or without retinal stabilization. When an anorthoscopic display is stabilized, it is impossible for the observer to pursue the object moving behind the slit and paint its extended shape onto the retina. Across two experiments, Fendrich and colleagues found no differences between stabilized and unstabilized presentations in terms of the time required for an initial report of the shape percept, the duration of shape perception episodes, or the frequency of the episodes. They concluded that their data provided no evidence that the retinotopic representation contributes anything to the formation of anorthoscopic percepts and attributed anorthoscopic figure perception to post-retinal processes, a conclusion consistent with some earlier analyses (e.g., Parks, 1965).

Evidence for position updating that is not dependent on eye movements has also come from research more directly involving spatiotemporal interpolation—cases where partial edge information is received sequentially in time and used to support connections across gaps. In the earliest studies of kinetic illusory contours, displays involved object parts that rotated around a stationary center point (Bruno & Bertamini, 1990; Kellman & Cohen, 1984). In such rotational displays, position updating cannot be accomplished by eye movements because rotation of the eyeballs around the line of sight is severely limited in humans.

Thus, though pursuit eye movements can be responsible for position updating, they should not be considered the only source—mental representations of velocity and position are also implicated. The simple act of walking past a hedge-lined fence and seeing the yard behind demonstrates that not all position updating in dynamic occlusion is handled by pursuit eye movements. Because the optical displacements of the objects in the yard occur at different rates relative to the observer (due to the geometry of motion parallax), it would be impossible to move one's eyes in a compensatory fashion for all depth planes simultaneously. Indeed, complete objects are perceived at multiple depth planes despite only small fragments of the objects being visible at any given time and despite those fragments moving at different speeds relative to the observer.

Regardless of whether the input to STR comes from eye movements or velocity information extracted from shape fragments while visible, the important concept is that visible and occluded regions of moving objects are aligned within the same reference frame so that the relatability computation can be performed. This notion is similar to Feldman's (1985) proposal of a *stable features frame* in which operations are performed on static representations of the visual scene, with motion removed through pursuit eye movements or parameterization.

Why might the visual system maintain and update representations of object fragments that become occluded? One reason might be the importance of tracking moving objects through cluttered environments (e.g., Scholl & Pylyshyn, 1999). Another is our present focus: the perception of objects from moving shape information that arrives piecemeal over time.

Spatiotemporal Relatability

Taken together, the persistence and position updating hypotheses allow for the continued representation of moving image fragments for a short time after they pass behind an occluding surface. This brings us to the final hypothesis in STR: We propose that when new portions of a dynamically occluded object are revealed, currently visible regions and previously stored, positionally updated regions both enter the same spatial relatability computation as occurs in static scenes, leading to visual unit formation. The extension of spatial relatability into the spatiotemporal domain via persistence and position updating is parsimonious because the application of relatability to dynamic situations may be the same as in static situations, for which there is considerable evidence.

In Figure 1, for example, the visible regions of the fire truck are depicted in the right column in the positions at which they were seen. The persistence hypothesis proposes that these fragments survive as perceptual representations after occlusion, whereas the position updating hypothesis suggests that their positions move along the same trajectory (left to right) as the visible regions. If this were the case, then the visible fragments of the fire truck could be accumulated over time, as in the bottom panel. Because parts of the fire truck never project to the eyes, interpolation processes are required to connect visible regions across gaps. Thus, the three processes of persistence, position updating, and relatability—acting in concert on currently and previously viewed fragments—accomplish dynamic object formation.

Experiment 1A: Perception of Shape Relations in Dynamically Occluded Objects

Real-world situations, such as perceiving a moving fire truck through foliage, inspired us to create an experimental situation in which participants viewed an object that translated laterally behind the multi-aperture occluder depicted in Figures 4 and 5. Apertures in the occluder were offset horizontally and vertically so that no two apertures were directly above or below each other. Fragments of the translating objects were visible only for very short periods of time through narrow (10-arcmin), spatially separated, and misaligned apertures. To give some intuitive idea of the size of these narrow apertures, their width was equal to the thickness of a quarter (1.75 mm) viewed at about arm's length (60 cm). Following each dynamic occlusion display, participants made a two-choice discrimination judgment about which of two displays they had seen moving behind the occluder (depicted schematically in Figure 4, below each occluder). For each discrimination, the two displays were versions of the same three fragments, differing only in the horizontal alignment of one of their three pieces.

The objects and occluder were designed so that objects could not be perceptually completed from information in any static frame of the animation sequence (see Figures 5A–5D). The task required spatiotemporal integration in that shape information had to be accumulated over time due to the small size of the slits relative to the objects. The task required spatiotemporal interpolation because

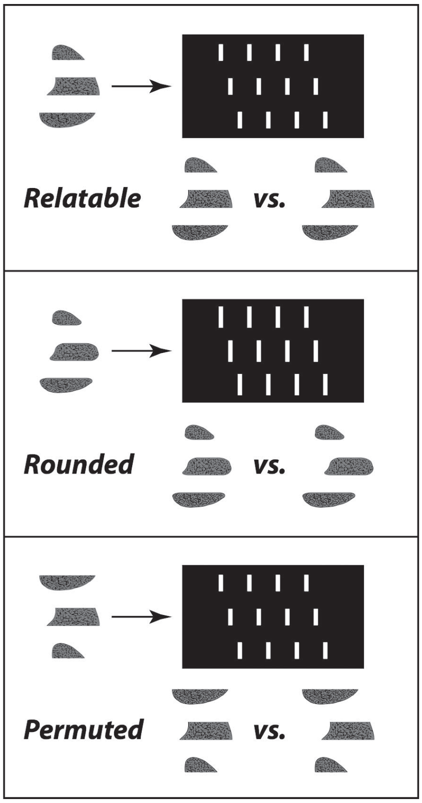

Figure 4. A schematic depiction of the two-choice decision task. Participants watched a shape move behind an occluding surface and then attempted to choose which of two piece configurations they had seen.

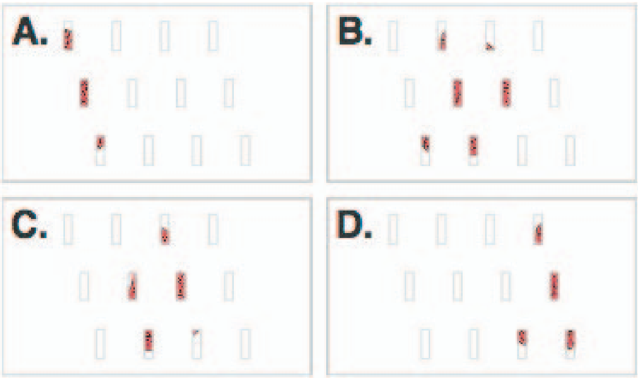

Figure 5. A depiction of the paucity of shape information available to participants during the animation sequence (figure and occluder drawn to scale). In the actual experimental displays, the occluder was black, and the object was red.

visible fragments given sequentially in time were separated by two large horizontal gaps. Substantial regions of the objects never projected to the eyes (44%, on average), and the areas that did project were spatially discontinuous.

The objective performance task used to study object formation in this experiment was designed with two key assumptions in mind. The first assumption was that encoding a dynamically occluded object as a single perceptual unit would allow an observer to detect a change in spatial relations better than would encoding the object as several perceptual units. In conditions that produced perception of a single, unified object, participants should have shown higher sensitivity to spatial misalignments of the pieces than in conditions that led to perception of three separated fragments. A number of prior studies indicated that visual unit formation aids in processing of relations within a single object (e.g., Behrmann, Zemel, & Mozer, 1998; Ringach & Shapley, 1996). If STR captured conditions for dynamic visual unit formation, we expected that observers would show better performance for fragments that were spatiotemporally relatable.

The second assumption was closely related to the first. We expected that in our two-choice discrimination paradigm, a decision between a choice perceived as a single unit versus another choice perceived as multiple units would be easier than a discrimination between choices with the same number of perceived units. For instance, it should have been comparatively easy to identify the target shape if it appeared to be one object moving behind the occluder when the choices in the discrimination phase consisted of an image that looked like one object and an image that looked like three pieces. In that case, there would be a salient qualitative difference between the two percepts leading to better discrimination performance. On the other hand, if participants discriminated between two displays that both looked like a single object or two displays that both looked like three disconnected fragments, then performance in the task should have been less good, with judgments between single objects being less difficult than judgments between multiple pieces.

The logic of this experiment was that if the geometry of STR was not a factor in perceiving dynamic occlusion displays, then disrupting it should have had no effect on participants' performance in the dynamic occlusion task. If participants did not perceptually connect object fragments within and between the apertures, then they should perform no better in the STR case than with control displays in which the same fragments were presented but in relationships that did not satisfy STR. Data that disconfirmed this null hypothesis would constitute support for STR as a determinant of spatiotemporal object perception.

Method

Participants. Twenty undergraduates (10 female and 10 male students, mean age 19.8 years) from the University of California, Los Angeles, participated in the experiment in partial fulfillment of requirements for an introductory psychology class. All participants reported normal or corrected-to-normal vision and were naïve as to the purposes of the experiment.

Apparatus. A program written in MacProbe (Hunt, 1994) controlled the timing and presentation of stimuli and gathering of participant responses using an Apple Macintosh G3 computer and an Apple Colorvision 17-in. (43.8-cm) display. Stimuli were presented on the display at 1,024 × 768 pixel resolution, with each pixel subtending one minute of visual angle. A chinrest was used to stabilize participants' heads at a distance of 114.5 cm from the monitor. Responses were gathered with an Apple Macintosh extended keyboard.

Stimuli. Four bloblike objects were used as prototypes (see Figure 6), generating 18 stimuli each (3 display types × 6 misalignment levels, as depicted in Figure 7). The objects subtended about 2.73° vertically and ranged from 1.46° to 2.15° horizontally. The four objects were red, with speckled black texture. The texture was added to the shapes to overcome the aperture problem (Marr & Ullman, 1981), that is, to help disambiguate their motion direction within the apertures.

The objects in this study were either relatable, rounded, or permuted depending on their display condition (see Figure 7). The three display conditions were designed to evaluate particular aspects of the spatiotemporal interpolation process. Relatable figures had edge relationships that fulfilled the conditions for STR, especially in the 0 misalignment condition (see Figure 6 and Figure 7, top left). If STR provides an accurate account of dynamic object formation, one would predict this condition to produce the best performance. Note that large amounts of misalignment also disrupted STR; for convenience, we use the label *relatable* to refer to the display set in the top row of Figure 7, although the rightmost two panels show misalignments larger than the small misalignment tolerance known to exist for relatability. For parallel edges in static objects, strength of interpolation has been shown to be strongly disrupted by 15–20 arcmin of

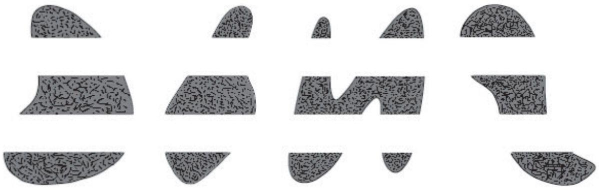

Figure 6. The four displays used in the relatable condition of Experiment 1A. Spatial relatability (Kellman & Shipley, 1991) leads the three fragments composing each object to be perceptually unified despite the absence of a visible occluder. The two gaps separating the three visible pieces of each shape represent areas that were never visible through the occluder. The two-choice decision task used displays as shown here (having fragments and gaps).

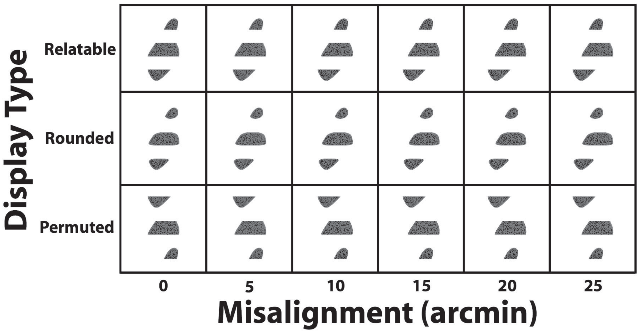

Figure 7. The 18 configurations of one display used in the study. Misalignment magnitudes are drawn to scale, relative to object dimensions. Visual angles used in the experiments may be approximated by viewing the objects in this display from 21 times their vertical extent.

misalignment (Kellman, Yin, & Shipley, 1998; Shipley & Kellman, 1992).[2]

The permuted condition (see Figure 7, bottom row) used the same image fragments as in the relatable condition except that the top and bottom ones were reversed. This manipulation maintained the same local image information as the relatable condition but disrupted STR[10] because the top and bottom pieces no longer had occluded edges that were relatable with the middle piece. If object perception suffered in this condition, it would provide evidence that perceptual processes that create visual units are sensitive to the positional relations of dynamically occluded object fragments rather than to just the fragments themselves.

Finally, the rounded condition (see Figure 7, middle row) was developed to evaluate the importance of tangent discontinuities in the perception of dynamically occluded objects. If STR engages contour interpolation processes, then we would expect that tangent discontinuities are important in initiating those processes, as in the static domain. To our knowledge, this notion has not previously been tested with dynamically occluded objects.

Figures in the rounded condition were created by rounding the edges of relatable image fragments near their occlusion points. This manipulation maintained the global geometry of the objects. Each corner of a figure that touched the occluder had pixels removed until it could fit within a 10-arcmin-diameter circle. On average, this eliminated 138 pixels from each piece or 414 pixels per image, a change of roughly 6.6% from the original pixel count. Previous research provided variable evidence about whether the rounding of tangent discontinuities eliminates (Kellman, Garrigan, & Shipley, 2005; Shipley & Kellman, 1990, Experiment 1) or weakens interpolation (Albert & Hoffman, 2000; Guttman & Kellman, 2004; Shipley & Kellman, 1990, Experiment 2). Specifically, displays whose elements have perfectly circular boundaries over large extents produce no evidence of interpolation, whereas more local rounding (of the sort we did here) produces reduced interpolation effects. The ability of locally rounded elements to produce weak interpolation derives either from the presence of second-derivative discontinuities (where locally circular boundary parts join other parts of a boundary) or from triggering of tangent discontinuity detectors in low-spatial-frequency channels (Albert & Hoffman, 2000; Guttman & Kellman, 2004; Shipley & Kellman, 1990). To preserve most of the overall shape of visible areas in this experiment, we used local rounding. Thus, we expected weakening, but not necessarily elimination, of interpolation effects.

If performance in both the permuted and rounded conditions was reliably worse than in the relatable condition, it would provide support for the notion that STR specifies conditions for dynamic unit formation. Again, referring to the null hypothesis, if unconnected fragments were all that were perceived, task performance would be expected to be similar across all three display conditions.

The computer-generated occluder[3] contained 12 narrow windows (subtending 10 × 36 arcmin each) that were staggered horizontally and vertically (see Figures 4 and 5). The apertures were arranged in three rows of four windows with two occluded regions between the three rows. The horizontal staggering and vertical separation ensured that shape information available through the apertures at any given moment in time was sparse at best (see Figure 5). Two horizontal regions, comprising about 44% of the objects' total extents, were never visible, making any object formation occurring in this paradigm necessarily dependent on interpolation.

Whenever the three separated fragments of a rounded condition stimulus were displayed, the windows in the occluding black surface were enlarged vertically by 7 arcmin in each direction (14 arcmin total) so that no part of the rounded stimuli touched the top or bottom border of any of the windows. That is, we arranged to have some white space between the object fragments and the occluding surface in the rounded condition. This eliminated all tangent discontinuities (both T and L junctions) coincident with occluding edges at the vertical gaps in the occluder while at the same time preserving the qualitative geometric relationships between pieces. The distances between the centers of the pieces and also the unrounded parts of the top and bottom edges of the pieces were the same as in the relatable condition.

Design. On each trial, participants watched a display of three object fragments move horizontally behind a multi-aperture occluding surface and then attempted to identify the object from among two choices: a target image (the display that was used in the animation sequence) and a distractor image (that differed only in the horizontal alignment of one of its pieces relative to the target image). The arrangement of the objects' pieces either did or did not satisfy the conditions for STR. STR was disrupted either through changes in how the pieces were displayed or through changes in misalignment, respectively (see Figure 7).

The misalignment variable had six levels: horizontal displacements of 0, 5, 10, 15, 20, or 25 minutes of visual angle (see Figure 7). On each trial, one of the four objects was chosen at random, and the display type and misalignment combination used were randomly selected from a counterbalanced set. Discriminations tested included 0/5, 0/10, 0/15, 0/20, 0/25, 5/10, 10/15, 15/20, and 20/25. Additionally, the direction (left or right) of misalignment was chosen randomly, and the particular piece (top, middle, or bottom) to be shifted was chosen quasi-randomly. The top and bottom pieces were chosen with a probability of 25% each, whereas the middle piece was picked with a probability of 50%. The probabilities were based on the number of edge relationships that could be disrupted by moving each piece. By moving either the top or bottom piece, only edge relationships between that piece and the middle piece were disrupted, whereas moving the middle piece disrupted edge relationships between both the top and bottom pieces relative to the middle piece. By quasi-randomly choos-

---

[2] The four different shapes, because of different relative orientations of partner edges, had somewhat different susceptibilities to misalignment before visual unit formation was disrupted. Some edge pairs were disrupted by 15 arcmin, and some were not disrupted until about 20 arcmin. These differences follow from the definitions of the limits of relatability and tolerance for small misalignments beyond those limits. Specifically, the limit of relatability is reached when the linear extension of one edge intersects the tip of a partner edge. For parallel vertical edges, any horizontal misalignment immediately goes beyond the limits, whereas this is not the case for nonparallel edges. Thus, misalignments induced by horizontal shifts in the experiments did not correspond precisely to misalignment beyond relatability limits in every case.

[3] A separate experiment, not reported here, used a physical occluder attached to the computer monitor, which provided a real depth difference from the screen objects. Results were statistically indistinguishable from those reported here.

ing the piece to be moved in this manner, we hoped to discourage response strategies in which participants always attended to a particular piece or pair of edges when performing the task. On each trial, participants made one of the nine possible discrimination judgments from the misalignment conditions listed above. There were 18 possible trial types because the target could be either of the two misalignments per pair.

The 54 experimental levels (3 display types × 18 misalignment combinations) were administered twice for each shape for a total of 432 experimental trials per participant.

Procedure. At the start of the experiment, a participant was seated at the computer, and his or her head was stabilized in a chinrest. On each trial, a target animation sequence was presented consisting of one of the four shapes translating laterally at a rate of 8.8°/s behind the occluder from left to right and then back to the left again. Following the animation sequence, a target image and a distractor image were displayed at the bottom of the screen. The left or right position of the target and distractor was chosen randomly on each trial. A practice block of 10 trials was performed at the beginning of the study to acquaint participants with the task. (These data were not included in the analyses.) During the experimental phase, two minute-long breaks were given after Trials 144 and 288. Participants initiated each trial with a keypress. The computer prepared the animation for display and then sounded a beep immediately before the image appeared, alerting the participants that the motion display was beginning.

The task was perceptually demanding because the pieces of the target and distractor objects had the same shape, with the only difference between the two alternatives being the relative alignment between the three fragments (see Figure 4). Apertures were arranged so that accurate shape perception was not possible from static views of the objects at any position (see Figures 6A–6D); however, all participants reported being able to perceive a shape (or shapes) moving behind the occluding surface. (Participants' subjective reports were gathered informally during the instruction phase.) Observers were not given fixation instructions, and eye movements were not monitored in this experiment. Informal observation suggested that participants made pursuit eye movements, tracking the objects as they moved across the screen.

So as not to bias participants' responses, the two horizontal regions that were occluded in the dynamic display were not shown to the participant during the two-choice discrimination phase; that is, discriminations were made between the same fragments of the image that would have been visible if the observer could accumulate information over time about the projected object fragments and their spatial relations (as in Figure 4). Whole objects were never shown in this experiment.

Dependent measures and data analyses. Data from the four shapes (see Figure 6) were combined. To assess participants' ability to detect a misalignment difference between the pieces, we submitted accuracy scores from the 18 stimulus combinations per shape condition to a signal-detection analysis. This analysis removed response biases that might occur in this paradigm (e.g., always choosing the more aligned version of the two choices). Twenty-seven sensitivity (d') scores (nine each for relatable, rounded, and permuted) were calculated for each participant.

We used an alpha level of .05 on all tests of statistical significance. For analyses of variance (ANOVAs), we computed the partial eta squared (η_p^2) statistic to indicate effect size. To indicate t-test effect size, we calculated Cohen's d.

Results

The primary results of Experiment 1A are shown in Figure 8, which plots participants' sensitivity (d') for discriminating a 0-arcmin-misaligned figure from other values of misalignment. Across all misalignment comparisons, performance was best for relatable displays, somewhat worse for the rounded displays, and poorest for permuted displays.

These observations were confirmed by the analyses. We submitted the d' scores to a 3 × 9 (Display Type × Misalignment) repeated measures ANOVA. There were reliable main effects of display type, $F(2, 38) = 48.33$, $p < .001$, $\eta_p^2 = .72$, and misalignment, $F(8, 152) = 53.86$, $p < .001$, $\eta_p^2 = .74$, as well as a reliable Display Type × Misalignment interaction, $F(16, 304) = 3.05$, $p < .001$, $\eta_p^2 = .14$.

Planned comparisons between the three display conditions revealed that sensitivity in the relatable condition was higher than in both the rounded condition, $t(19) = 3.49$, $p = .0024$, $d = 1.11$, and the permuted condition, $t(19) = 12.04$, $p < .001$, $d = 3.81$. Sensitivity in the rounded condition was reliably higher than in the permuted condition, $t(19) = 5.50$, $p < .001$, $d = 1.74$. All 20 observers performed better in both the relatable and rounded conditions than in the permuted condition, and 16 of the 20 performed better in the relatable condition than in the rounded condition. Rounding the corners of the shapes lowered discrimination performance relative to the relatable condition, though not to the level of the permuted condition.

The Display Type × Misalignment interaction appears to have been driven by greater differences among conditions for misalignment comparisons beyond 0/5 and also by more similar patterns of improvement across increasing misalignments for the relatable and rounded conditions than for the permuted condition (see Figure 8). Participants were not able to reliably discriminate small misalignment differences in the permuted condition until they became larger than 10 arcmin.

The main effects and interactions are consistent with an interpretation of the data in terms of visual unit formation by STR. We explored these comparisons further with regard to displays that may have been perceived as one versus several units. For the data depicted in Figure 8, our hypotheses suggested that the relatable condition (and, to a lesser extent, the rounded condition) consisted of comparisons between one visual unit versus one visual unit or one unit versus multiple units. These comparisons should have been easier than the comparison of multiple units versus multiple units for all values of misalignment in the permuted condition.

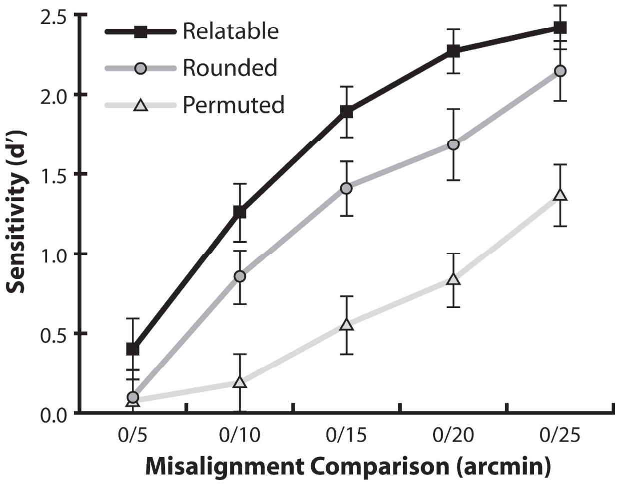

Figure 8. Main results for Experiment 1A (dynamic occlusion): Sensitivity (d') by display condition as a function of misalignment difference between target and distractor images. Error bars indicate ±1 standard error of the mean.

Several studies have shown that for misaligned parallel edges, contour interpolation in static displays degrades continuously and, by 15 to 20 arcmin of misalignment, is substantially disrupted (Kellman et al., 1998; Shipley & Kellman, 1992). Using 15 arcmin as a rough estimate of the point at which displays are sufficiently misaligned to look like multiple units, we divided the relatable and rounded conditions into three groups based on the hypothesized discrimination task demands (see Figure 9). In the relatable groups, comparisons between small amounts of misalignment (0/5, 0/10, and 5/10) were considered discriminations between one unit versus one unit (1vs1). Comparisons between large amounts of misalignment (15/20 and 20/25) were considered discriminations between three units versus three units (3vs3).[4] The remaining misalignment comparisons (0/15, 0/20, 0/25, and 10/15) were between displays that appeared to be a single unit and displays that appeared to be three units (1vs3). Given that we did not expect the total elimination of interpolation effects from rounding and that the data suggest weak interpolation effects were present, we predicted rounded displays would share in the 1vs3 advantage but would do so more weakly than in relatable displays with clear tangent discontinuities.

In the permuted condition, the three pieces of the display were never relatable and thus always constituted a 3vs3 comparison. In Figure 9, the permuted condition is divided into the same misalignment comparison conditions as the relatable condition for the sake of comparison.

To test the hypothesis that comparisons that crossed the 15-arcmin misalignment boundary had a stronger effect on the relatable (and rounded) condition than on the permuted condition, we computed the advantage of the relatable (and rounded) condition over the permuted condition for each of the visual unit comparison groups. This was done by subtracting the permuted score for that group from the relatable (or rounded) score for the same group. We opted for this comparison rather than simply assessing the sensitivity of the 1vs3, 1vs1, and 3vs3 groups directly because the 1vs3 group had larger misalignment differences than either the 1vs1 group or the 3vs3 group. Thus, in the comparisons we performed between groups, the misalignment values were always equated. We predicted that if the visual unit hypothesis account of the data was correct, then the advantage of the relatable condition over the permuted condition should be larger in the 1vs3 group than in the 3vs3 and 1vs1 groups.

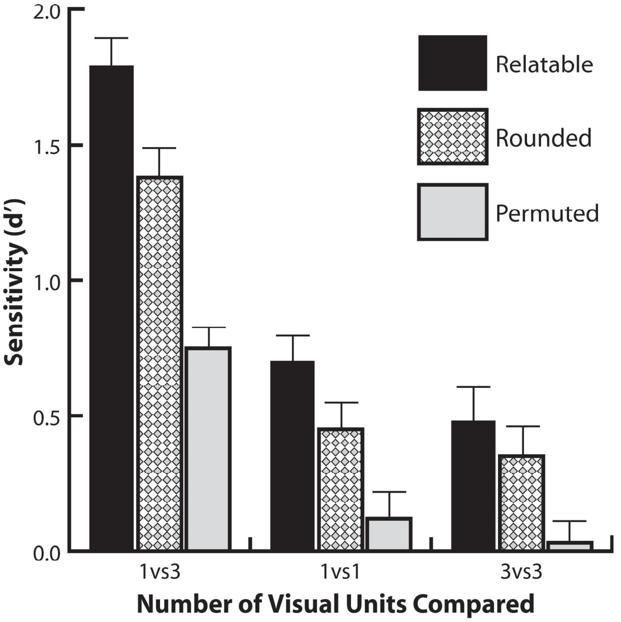

Figure 9. Sensitivity (d') to misalignment as a function of the number of visual units in the display and the type of comparison being performed in Experiment 1A (dynamic occlusion). Error bars indicate ±1 standard error of the mean.

Using *t* tests on the magnitude of the sensitivity improvement (using a Bonferroni correction for Type I error, $p_{crit} = .0056$) for the relatable condition over the permuted condition across the three visual unit comparison categories, we found that the relatable advantage was bigger in the 1vs3 comparison group than in the 1vs1, $t(19) = 4.40$, $p < .001$, $d = 1.39$, and the 3vs3 comparison groups, $t(19) = 3.2$, $p = .0046$, $d = 1.01$. The rounded condition also showed a larger advantage in performance over the permuted for the 1vs3 condition than for the 3vs3 condition, $t(19) = 3.17$, $p = .0050$, $d = 1.00$. No other comparisons, including comparisons between the relatable and rounded conditions, reached significance (all *p*s ≥ .09).

Discussion

Participants in Experiment 1A showed a sizable sensitivity advantage for relatable relative to permuted displays. This effect occurred despite the fact that these two types of displays had identical local image fragments. The two display types differed in that the permuted displays interchanged the top and bottom pieces of the image relative to relatable displays, a manipulation that eliminated relatability between adjacent visible fragments. Why did this manipulation produce a large difference in sensitivity, shown by every participant? The data are consistent with and predicted by the theory of STR. Specifically, STR operates to accumulate object fragments over time and update their positions; the geometry of spatial relatability is then applied to the relations of stored and currently visible object parts to form connections. As in numerous paradigms used to study stationary displays, we predicted that the operation of spatiotemporal interpolation for relatable displays would lead to perception of the display fragments as a unitary object and confer an advantage in task performance. Poor performance in the permuted condition likely occurred because the displays were perceived as three separate pieces, not one unit, consistent with the importance of visual unit formation in the task. Spatiotemporal relations between successively visible fragments appeared to be important for perceiving dynamically occluded shapes.

A convergent measure of unit formation in this paradigm was furnished by the condition in which tangent discontinuities were rounded. Unlike the permuted displays, the displays in this condition preserved the overall spatial relations of the visible object parts. The sensitivity of the paradigm to object formation is evi-

---

[4] The notion that highly misaligned relatable displays are divided into three units is strictly true only for the trials on which the center piece is moved. In other cases, two of the three object parts remain relatable.

denced by the disruptions in performance produced by two different factors expected to affect object formation. Specifically, results with rounded displays, which matched the positional relations of fragments in the relatable condition, indicate that superior performance for relatable displays was not an accidental benefit of fortuitous spatial positions of the fragments. The present results support the importance of tangent discontinuities in dynamic, spatiotemporal interpolation, parallel to their importance in static, spatial interpolation.

The modification of tangent discontinuities in the rounded condition did not decrease performance to the level of the permuted condition. This result, consistent with earlier ones in static interpolation (Albert, 2001; Albert & Hoffman, 2000; Shipley & Kellman, 1990), suggests that our rounding manipulation substantially weakened, but did not eliminate, interpolation effects. In stationary displays, two hypotheses have been proposed to explain why local rounding of tangent discontinuities does not completely eliminate interpolation effects. One is that some rounded displays still contain second-order discontinuities, which may be capable of weakly triggering interpolation (Albert, 2001; Albert & Hoffman, 2000; Shipley & Kellman, 1990). A second hypothesis is that tangent discontinuities may be processed at multiple spatial scales (Albert, 2001; Wurtz & Lourens, 2000). It is likely that rounding the corners of the objects over a relatively small contour region removed tangent discontinuities at high-spatial-frequency, but not low-spatial-frequency, channels. Activation of detectors for tangent discontinuities at some, but not all, spatial scales may be sufficient to partially initiate contour interpolation processes. We considered these issues further in Experiment 2.

More fine-grained analysis of the results suggested that for small misalignments in the relatable and rounded conditions, but not in the permuted condition, successively visible fragments were assembled into single objects and thus were relatively easily discriminated from fragments not predicted to form unitary objects. This interpretation is consistent with the hypothesis that shape comparisons between two displays where at least one appears to be a single unit (the relatable condition) are better than shape comparisons between two displays where both appear to be several units (the permuted condition). However, we were not able to find reliable support for the notion that 1vs1 comparisons were easier than 3vs3 comparisons.

The results of Experiment 1A are consistent with the visible persistence and position updating hypotheses of STR. The arrangement of apertures within the occluding surface precluded the possibility of participants interpolating the bounding regions of the shape fragments across the occluded region at any instant. In the relatable condition, partner edges above and below the occluded regions were visible only in succession. Therefore, if the three horizontally translating pieces of the target object were combined into a single unified shape (as observers universally reported perceiving in the relatable, 0-arcmin misalignment condition), then this integration process depended on accumulation of information over time, consistent with the visible persistence hypothesis. Furthermore, the fact that participants were highly sensitive to horizontal misalignments of the three pieces indicates that they achieved a representation of the relative position of the fragments during intervals when pieces were traveling behind the occluder and were not physically visible, consistent with the position updating hypothesis.

The relevance of object formation to the patterns of performance obtained is supported by the fact that both particular geometric relations in space and time (expressed as STR) and the presence of tangent discontinuities were important. The reductions in performance caused both by the disruption of STR and by rounding tangent discontinuities are straightforwardly predicted by the notion that the experimental task is facilitated by contour interpolation and subsequent object formation. These two manipulations are qualitatively very different, making it difficult to construct an alternative hypothesis that predicts the effects of both on performance without implicating perceptual unit formation processes.

Looking at the permuted condition alone, for example, one might consider other factors that could contribute to lower performance compared with the relatable condition. The three pieces in the relatable displays all bordered the two horizontal strips between the apertures in the occluder, whereas only the middle piece in the permuted condition bordered them. This arrangement made the three constituent pieces of relatable objects slightly closer together than the three constituent pieces of permuted objects (3.6 arcmin, on average). It would be conceivable, although we believe unlikely, that this small difference in separations between the two conditions could account for the large performance differences observed. Such a notion, however, could not explain the reliably worse performance in the rounded condition, as spatial separations there were identical to those in the relatable condition.

Another difference between permuted and relatable displays was that shapes in the relatable condition extended directly up to the edges of the horizontal strip between the apertures. This arrangement could allow for surface spreading, a process that can connect retinally separated visual areas even without relatable contours (Yin et al., 1997, 2000). Surface spreading, rather than contour interpolation, might have contributed to visual unit formation. Such a possibility would still require the processes of visible persistence and position updating, but it would raise the interesting possibility that these processes facilitate surface completion. We examined the possible contribution of surface spreading in Experiment 3.

Experiment 1B: Perception of Shape Relations in Statically Occluded Objects

The results of Experiment 1A support the theory that STR is used by the visual system to perceive objects from information over time. However, not much prior work has been done on dynamic unit formation, and the inferences in Experiment 1A depend upon earlier findings in static unit formation, such as the role of static relatability (Kellman, Garrigan, & Shipley, 2005; Kellman & Shipley, 1991), tangent discontinuities (Albert & Hoffman, 2000; Shipley & Kellman, 1990), and the enhancement of discrimination performance by object formation (Guttman & Kellman, 2004; Kellman, Garrigan, Shipley, et al., 2005; Ringach & Shapley, 1996). Although the relations to earlier work are straightforward, neither the unfamiliar objects nor the discrimination paradigm we used to test dynamic object formation were identical to those of previous studies. To be certain of the comparability of our results to known effects in static interpolation, we conducted an experiment employing a direct static analogue of the displays and conditions used in our dynamic paradigm.

We sought to verify in Experiment 1B that well-known unit formation effects—effects of relatability, misalignment, and rounding of tangent discontinuities—would occur in a direct static analogue of the paradigm used in Experiment 1A. In this static version, however, all visible areas of partly occluded displays were simultaneously visible in an unchanging position (see Figure 10).

Perhaps the most direct evidence of effective persistence and updating in STR would have been results showing that, despite temporal fragmentation and spatial displacement, STR allows perceivers to assemble objects much the same as occurs when fragments and their relations are given simultaneously and in correct spatial register. One way to summarize our theory is to say that STR converts the chaos of moving fragments, seen through apertures in different places and times, to the object formation task confronted by the visual system in connecting stationary visible areas across spatial gaps.

From this perspective, we predicted that a static version of Experiment 1A would show effects similar to those found with dynamic occlusion displays. Conversely, if the superiority of relatable over permuted displays, the effects of misalignment, or the effects of rounding of tangent discontinuities failed to appear in a static, simultaneous version of the paradigm, it might suggest that our paradigm, unlike some others used previously with static displays, is not sensitive to unit formation. If so, the effects in the dynamic case might be attributable to other causes. In short, if the effects in dynamic occlusion displays occurred for reasons other than STR and object formation, clear differences between static interpolation performance and the results of the dynamic conditions of Experiment 1A would be expected.

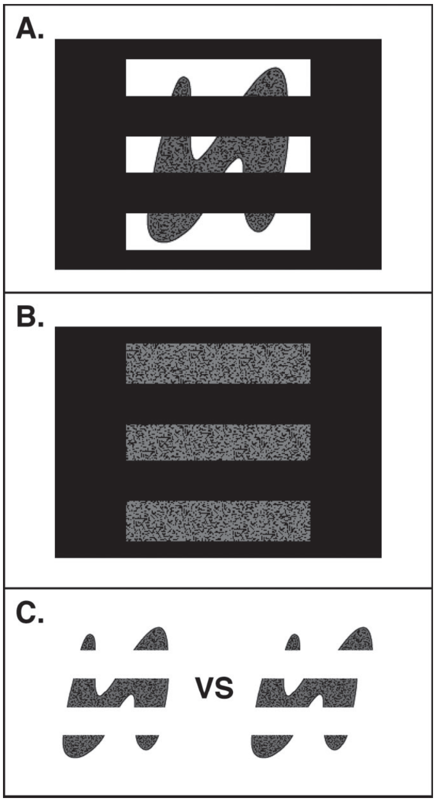

Figure 10. Format of Experiment 1B. A: The occluded objects in this experiment were stationary and displayed behind a modified occluding surface with three large horizontal apertures. B: After 493 ms, the object was masked by a red surface with the same texture as the objects. C: At the end of each trial, participants determined whether they had seen the object on the left or the right.

We altered the presentation conditions from Experiment 1A so that all the various object fragments that appeared during the dynamic sequences were now shown as correctly spatially arrayed and simultaneously visible. The static, simultaneous views of object parts were displayed for 493 ms before being masked. After this presentation, participants made a forced choice between the display that had been presented and another display in which one of the visible areas was shifted (see Figure 10). In all other respects, the procedure and stimuli of this experiment were identical to those of Experiment 1A.

Method

Participants. Ten volunteers (4 men and 6 women, mean age 25.3 years) from the Brigham and Women's Hospital Visual Attention Laboratory and the surrounding community participated in the study without compensation. All participants reported normal or corrected-to-normal vision and were naïve to the purposes of the experiment.

Apparatus. A program written in MacProbe (Hunt, 1994) controlled the timing and presentation of stimuli. Stimuli were presented and responses gathered by a 450-MHz Apple Macintosh G4 computer with a 21-in. (53.3-cm) RasterOps SuperScan Mc801 display at 1,024 × 768 resolution. Participants were seated 140 cm from the monitor, a distance at which one pixel on the monitor subtended one minute of visual angle.

Stimuli and procedure. Objects, conditions, and procedure were the same as in Experiment 1A, with the following exceptions. The target object was stationary and displayed through an occluder that contained three horizontal apertures. The apertures allowed projection of the same portions of the occluded object that would have been visible in a dynamic occlusion display if shape information were accumulated over time in accordance with the persistence and position updating hypotheses (see Figure 10A). The target shape configuration was displayed for 493 ms and then masked (see Figure 10B). Following the presentation of the figure, participants completed the two-choice discrimination phase (see Figure 10C).

Dependent measures and data analyses. Measures and analyses were the same as in Experiment 1A.

Results

Participants' sensitivity (d') to increasing misalignment differences from zero are depicted in Figure 11. Sensitivity in all conditions increased with misalignment differences. As in Experiment 1A, performance in the relatable condition was better than performance in the rounded condition, which was in turn better than performance in the permuted condition.

These trends were confirmed by a 3 × 9 (Display Type × Misalignment) repeated measures ANOVA. The analyses revealed main effects of display type (relatable, rounded, or permuted), $F(2, 18) = 24.82$, $p < .001$, $\eta_p^2 = .73$, and misalignment, $F(8, 72) = 25.58$, $p < .001$, $\eta_p^2 = .74$, as well as a significant Display Type × Misalignment interaction, $F(16, 144) = 2.12$, $p < .05$, $\eta_p^2 = .19$. Planned comparisons between display conditions revealed that participants' sensitivity to misalignment was higher in the relatable condition than in the rounded condition, $t(9) = 4.02$, $p = .003$, $d = 1.80$, which in turn was higher than in the permuted condition,

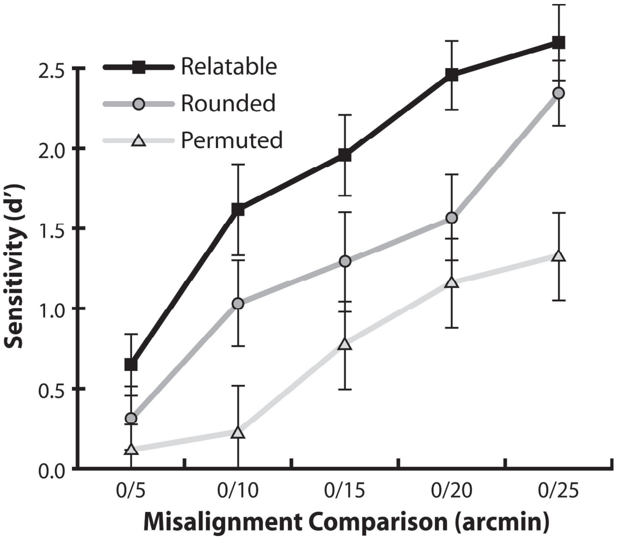

Figure 11. Main results for Experiment 1B (static occlusion): Sensitivity (d') by display condition as a function of misalignment difference between target and distractor images. Error bars indicate ±1 standard error of the mean.

$t(9) = 5.21$, $p < .001$, $d = 2.33$. All 10 observers performed better in the relatable condition than in the rounded or permuted conditions, and 9 of 10 performed better in the rounded condition than in the permuted condition. As in Experiment 1A, participants were most sensitive to differences in misalignment for these static displays when they fulfilled the conditions for relatability.

The data from Experiments 1A and 1B were submitted to a 2 × 3 × 9 (Experiment × Display Type × Misalignment) repeated measures ANOVA, with the first factor as a between-subjects variable. The analysis detected no significant difference in overall performance between the two experiments, $F(1, 28) = 0.31$, $p = .58$, $\eta_p^2 = .01$, *ns,* nor did the experiment variable interact with any other factors (all *p*s > .10). The main effects of display type, $F(2, 56) = 66.51$, $p < .001$, $\eta_p^2 = .70$, and misalignment, $F(8, 224) = 71.93$, $p < .001$, $\eta_p^2 = .72$, as well as the Display Type × Misalignment interaction, $F(16, 448) = 4.11$, $p < .001$, $\eta_p^2 = .13$, were all significant. Additionally, planned comparisons between the three display types showed no reliable differences between the static and dynamic versions of the experiment, every $t(9) \leq 1.56$, all *p*s ≥ .15.

As in Experiment 1A, we calculated the sensitivity advantage for the relatable and rounded conditions over the permuted conditions for the 1vs3, 1vs1, and 3vs3 data groups (see Figure 12). To reiterate, the visual unit formation hypothesis predicts that the relatable (and possibly the rounded) conditions would show a greater advantage over the permuted condition for the 1vs3 group than for the 3vs3 group. Planned comparisons of the relatable advantage over the permuted condition (with a Bonferroni correction yielding a p_{crit} of .0056) showed that it was again larger in the 1vs3 group than in the 3vs3 group, $t(9) = 5.44$, $p < .001$, $d = 2.43$. No other comparisons reached significance (all *p*s ≥ .024).

Discussion

Experiment 1B, with static versions of the stimuli used in Experiment 1A, provides a direct comparison with the dynamic occlusion results of Experiment 1A. If the visual system is able to accumulate shape information over space and time in the manner predicted by STR, then we would expect performance to show a pattern of data similar to the dynamic task. Indeed, the results of Experiment 1B parallel the findings of Experiment 1A in every major respect. Relatability of visible fragments under occlusion exerted a strong effect on object formation, as indicated by a clear advantage in discrimination performance for the relatable condition. This outcome verifies that well-known object formation effects for static displays (e.g., Behrmann et al., 1998; Kellman, Garrigan, Shipley, et al., 2005; Ringach & Shapley, 1996) were observable with this task and stimulus set.

Rounding the corners of image fragments to remove tangent discontinuities at the points of occlusion lessened, but did not eliminate, discrimination performance. This result is similar to the findings of Experiment 1A and also to several results from the literature on static interpolation effects indicating the importance of tangent discontinuities for initiating contour interpolation (Albert, 2001; Guttman & Kellman, 2004; Shipley & Kellman, 1990). The data of Experiment 1B make the comparison more exact, as only the mode of presentation differed between the two experiments. Taken together, the results of the two experiments support the interpretation that the intermediate level of performance for rounded displays derives from the relatively local rounding in the stimuli reducing, but not eliminating, unit formation effects (Shipley & Kellman, 1990).

The static, simultaneously visible, spatially assembled displays used in this experiment have an important relationship to the spatially and temporally displaced fragments shown in the dynamic conditions of Experiment 1A. If the bits and pieces of

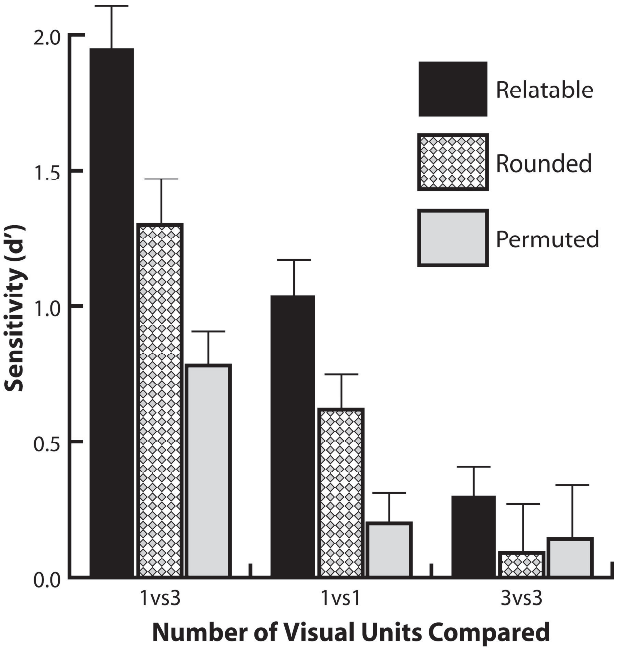

Figure 12. Sensitivity (d') to misalignment as a function of the number of visual units in the display and the type of comparison being performed in Experiment 1B (static occlusion). Error bars indicate ±1 standard error of the mean.

dynamic displays could be accumulated and position updated as we have theorized, the resulting displays would match the static, simultaneous displays of Experiment 1B. The fact that the dynamic and static cases show highly similar, robust, and statistically reliable effects of relatability, tangent discontinuities, and misalignment suggests comparable object formation processes for the two cases.

One fortuitous aspect of the data of Experiment 1B is that absolute levels of performance were similar to those in the dynamic conditions of Experiment 1A. Obviously, the relative performance among the relatable, rounded, and permuted displays was the focus of our predictions based on object formation and relatability. The similar absolute levels of performance overall seem to us to be coincidental but interesting. The 493-ms exposure duration was chosen after pilot work indicating that this presentation time would produce performance between floor and ceiling. Initially, we were concerned that simultaneous presentation of object fragments in a task requiring judgment of their spatial relations would simply be too easy in all conditions. Apparently, this is not the case, at least for half-second exposures.

One might ask how a 493-ms exposure of object fragments compares with total exposure in the dynamic occlusion displays. Recall that the occluder had four apertures in each row (see Figures 4 and 5). Because all object motion in the experiment was horizontal, any visible point on an object fragment on a single trial was projected four times through these apertures when traversing left to right and four times again when returning from right to left. Our data show comparable patterns for participants' performance with dynamically and statically occluded displays. It may be, however, that if one added up the total exposure time of image data in the dynamic condition, one would find that equivalent performance for dynamic displays comes only from a much longer total exposure duration. In fact, we would expect this result if the processes of persistence and position updating are subject to sources of noise or error under the conditions we tested.

The results of this comparison of total exposure time are surprising and remarkable. It turns out that in the dynamic displays, any image point was visible for a total of just under 152 ms—less than a third of the 493-ms stimulus duration for static displays. Yet performance in the dynamic case was as good as in the static case. These data suggest amazing precision of the persistence and position updating processes under the conditions we tested.[5] Per unit of physical stimulus exposure time, processing of positional relations in the displays presented seems to be more efficient in the dynamic than in the static case. In Experiment 2, we made more thorough comparisons between dynamic and static displays in an attempt to better understand the time course of dynamically occluded object perception.

Experiment 2: Temporal Attributes of Spatiotemporal Object Formation

Experiments 1A and 1B demonstrated that several stimulus manipulations known to affect contour interpolation in static displays have comparable effects in dynamic displays. A surprising result from those studies is that performance on dynamic displays with a physical exposure time of 152 ms was statistically equivalent to performance on static displays with an exposure time of 493 ms.

The persistence hypothesis of STR proposes that when parts of an object move behind an occluding surface, they nonetheless remain perceptually available for a short time after their disappearance. If this idea is true, then perhaps it is not surprising that performance in the dynamic displays of Experiment 1A was equivalent to performance in the static displays with a much longer exposure time in Experiment 1B. Although the physical exposure time of the dynamic displays was only 152 ms, persistence could have made the effective exposure time—or, more precisely, effective availability in a visual representation—much greater. Mere availability, however, would not suffice to produce discrimination performance equivalent to a longer static exposure. In the dynamic case, position updating, as hypothesized by STR, would also be required for coherent object formation.

In Experiment 2, we sought a better understanding of the relationship between the physical exposure time of fragments in a dynamic display and an effective equivalent exposure time, as measured in two different static displays for the same group of participants. For a single pass of an object behind the occluder used in Experiment 1A, any given pixel was physically visible for about 76 ms. However, if the persistence and position updating notions of STR are correct, then the effective availability for object formation processes might have been as high as 436 ms. This latter number comes from the amount of time it would take a single pixel to travel across the entire region of the occluder in which there are apertures in any row (see Figure 13).

Given these reference values, we asked in Experiment 2 how performance on dynamic displays would compare with performance on static displays with exposure durations of either 80 ms or 440 ms.[6] If performance on dynamic displays exceeded performance on static 80-ms displays, that would suggest that the effective perceptual availability for dynamically occluded objects is longer than their physical exposure durations. The 440-ms exposure time represents the amount of time that an observer would see the displays if persistence were optimal. If performance on dynamic and 440-ms static displays were identical, that not only would support the notions of persistence and position updating but would suggest in addition that these processes function (under the conditions tested) with little, if any, loss of information. What we expected, based on the theory and results described so far, was that persistence and position updating would do a good, but not perfect, job of preserving and positionally updating previously seen object fragments. If so, then performance on dynamic displays was predicted to fall somewhere between the 80-ms and 440-ms static exposure conditions. In this experiment, then, we hoped to estimate the effective exposure duration of dynamic displays relative to matched static controls within the same observers.

Besides being matched to the physical exposure duration of these dynamic displays, the 80-ms condition fulfilled another purpose in evaluating the claims of STR. We interpreted the

---

[5] In other stimulus configurations, position updating is far from veridical, resulting in robust visual illusions (E. M. Palmer & Kellman, 2001, 2002, 2003). We will report these data in forthcoming work.

[6] The slightly different exposure times are due to the refresh rate of the monitor used in the experiments. For our 75-Hz monitor, a new frame could be drawn every 13.3 ms, so the nearest exposure times to 76 ms and 436 ms were 80 ms and 440 ms, respectively.

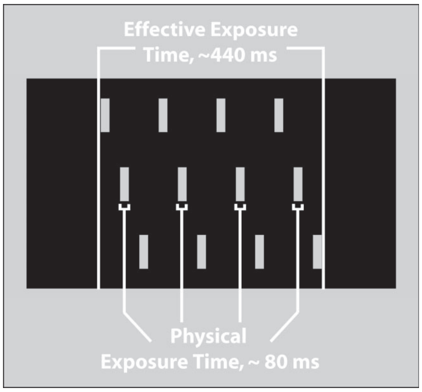

Figure 13. Physical versus effective exposure time for objects that pass once behind the occluding surface. Any given pixel on the moving object is physically visible for a total of roughly 80 ms (20 ms per aperture) through four apertures in the occluder. If persistence and position updating operate, then the effective exposure time might be more like 440 ms, which is the amount of time it takes for a single pixel to traverse the entire extent of apertures.

patterns of data observed in Experiments 1A and 1B as being due to contour interpolation and consequent visual unit formation effects between object fragments. If true, we expected to see a different pattern of data in displays in which the contour interpolation process was interrupted before it had time to occur fully.

Previous work on the time course of contour interpolation in static displays suggested that the precision of those connections increases over time. Using a dot localization task, Guttman and Kellman (2004) showed that some contour interpolation effects are clearly manifest by 60 ms but that such effects continue to increase (as measured by the precision and localization accuracy of a dot probe relative to a perceived contour) up to 120 ms. Other data (e.g., Ringach & Shapley, 1996) are consistent with even longer intervals (e.g., up to 160 ms). For the shorter exposure duration of 80 ms in the static displays of Experiment 2, we expected to observe partially developed contour interpolation effects. However, the static exposure duration of 440 ms should have been long enough, based on any existing estimates, to allow fully developed contour interpolation between visible regions. What differences in patterns of data might be expected if interpolation effects are not asymptotic? One clear prediction is that the absolute level of performance for relatable displays should have been lower in the 80-ms static condition than in the 440-ms and perhaps the dynamic conditions. Another prediction is that of an interaction: Greater exposure time should not simply improve performance on all display types (relatable, rounded, and permuted) in the same fashion. If object formation effects are involved, we would expect the gains of longer exposure time to be greater in the relatable condition than in the control display types.

Experiment 2 was designed to allow a more sensitive assessment of the predictions of STR relative to static displays. The comparison of performance for Experiments 1A and 1B was between two groups of participants, with 20 participating in the former experiment and 10 participating in the latter. Because the statistical comparisons were between groups, perhaps they lacked the power to detect a difference between the dynamic and static presentation modes. Note that we hypothesized above that finding dynamic condition performance equivalent to the 440-ms condition would suggest ideal, noise-free performance of STR (i.e., no information loss relative to simultaneous and continuous exposure of object parts). Recall that Experiment 1B showed statistically indistinguishable performance between dynamic displays and a 493-ms static presentation. We did not consider it plausible that dynamic object formation based on STR would exceed the ideal of completely visible object fragments. Thus, although the result of Experiment 1B clearly puts dynamic interpolation in the ballpark of 500-ms static interpolation, we believed a more discerning test might be provided by a within-subjects design. Accordingly, all participants in Experiment 2 participated in the static 80-ms, static 440-ms, and dynamic conditions.

Method

The method in this experiment was the same as in Experiment 1B, except as noted.

Participants. Twenty participants (6 women and 14 men, mean age 25.8 years) from the Brigham and Women's Hospital Visual Attention Laboratory subject pool participated in this study and received $10/hr compensation. All observers reported normal or corrected-to-normal visual acuity, passed an Ishihara color test, and were naïve to the purposes of the experiment.

Design. There were one dynamic presentation condition and two static presentation conditions in this experiment. The dynamic condition was the same as in Experiment 1A, except that the object passed only once behind the occluding surface instead of twice. The direction that the object moved was chosen randomly on every trial. The static presentation conditions were the same as in Experiment 1B, except that they had durations of either 80 ms or 440 ms. The two static conditions were intermixed and presented in random order. For all three presentation conditions, only the misalignment pairings of 0/5, 0/15, and 0/25 were tested.

Procedure. Participants first completed the dynamic presentation condition and then the two static presentation conditions. The dynamic condition was always tested first so that it would not benefit from any practice effects in the task. Before the dynamic phase, participants were given the same instructions and practice as used in Experiment 1A. Before the static phase, participants were given the same instructions and practice as used in Experiment 1B. A mandatory 60-s break was administered after every 100 trials throughout both phases of the experiment.

Results

Participants' sensitivity (d') to misalignment as a function of display condition and presentation mode is depicted in Figure 14. As in Experiment 1B, patterns of performance in the dynamic condition and in the longer exposure static (440-ms) condition were highly similar. Consistent with incomplete emergence of interpolation effects, performance in the short (80-ms) static exposure condition differed both in overall level and in the patterns shown for different display types. Specifically, relatable displays did not attain levels in the 80-ms condition that were as high as the dynamic or 440-ms conditions. Moreover, at 80 ms, there was a smaller difference between relatable, rounded, and permuted displays than in the other conditions.

These patterns were confirmed by the analyses. The data were analyzed with an omnibus 3 × 3 × 3 (Presentation Mode [dynamic, static 80-ms, static 440-ms] × Display Type [relatable, rounded, permuted] × Misalignment) within-subjects ANOVA.

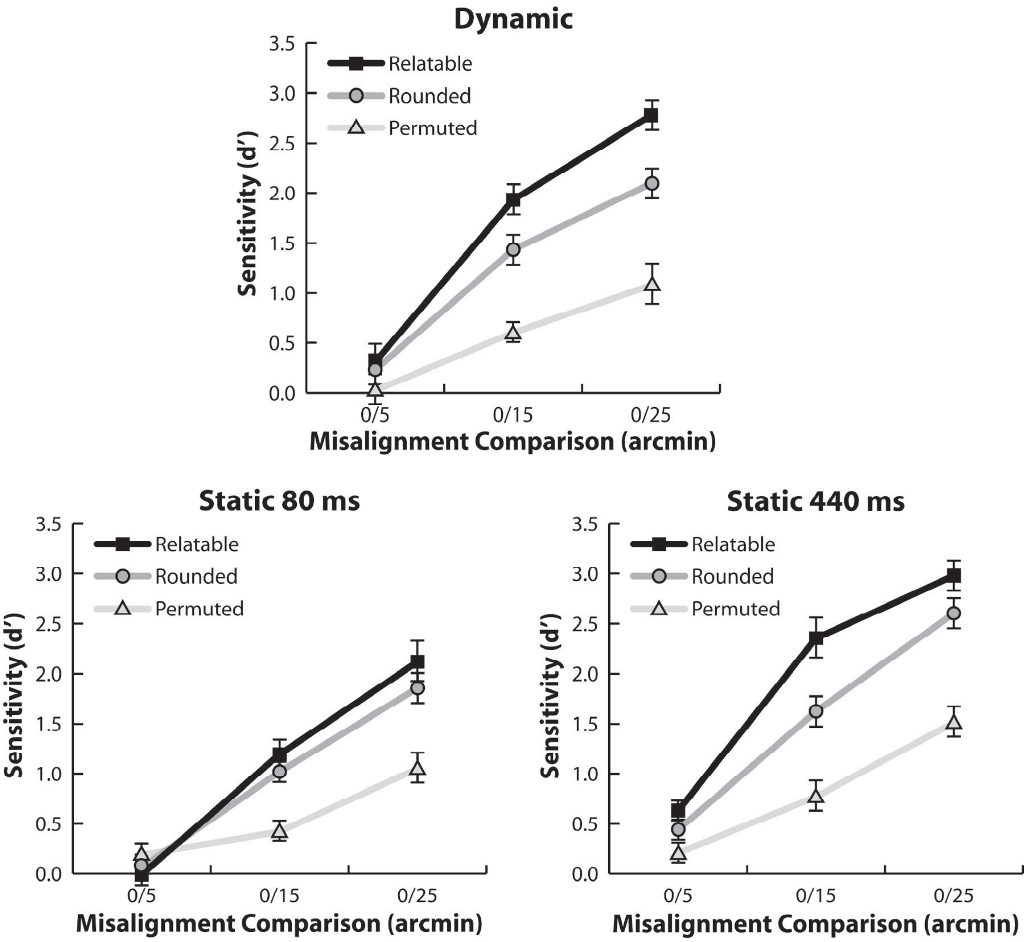

Figure 14. Main results for Experiment 2: Sensitivity (d') as a function of misalignment comparison in the dynamic, 80-ms static, and 440-ms static conditions of Experiment 2. Error bars indicate ±1 standard error of the mean.

There were reliable main effects of presentation mode, $F(2, 38) = 24.19$, $p < .001$, $\eta_p^2 = .56$; display condition, $F(2, 38) = 81.37$, $p < .001$, $\eta_p^2 = .81$; and misalignment, $F(2, 38) = 341.23$, $p < .001$, $\eta_p^2 = .95$. Interactions of Presentation Mode × Display Type, $F(4, 76) = 6.78$, $p < .001$, $\eta_p^2 = .26$, and Display Type × Misalignment, $F(4, 76) = 23.67$, $p < .001$, $\eta_p^2 = .56$, were also significant. There were no other main effects or interactions in the omnibus ANOVA (all *p*s > .10).

Further exploration of these data using paired *t* tests revealed that overall sensitivity to displays was greater for dynamic presentation than for 80-ms static presentation for both relatable, $t(19) = 3.89$, $p = .001$, $d = 1.23$, and rounded displays, $t(19) = 2.33$, $p = .03$, $d = 0.74$. For permuted displays, there was no significant difference between dynamic and 80-ms static presentation ($p > .10$). Performance in 440-ms static presentations was reliably better than performance in dynamic presentations for relatable, $t(19) = 2.31$, $p = .03$, $d = 0.73$, rounded, $t(19) = 2.35$, $p = .03$, $d = 0.74$, and permuted displays, $t(19) = 3.04$, $p = .007$, $d = 0.96$. Finally, performance in 440-ms static presentations was reliably better than in 80-ms static presentations for relatable, $t(19) = 10.49$, $p < .001$, $d = 3.32$, rounded, $t(19) = 5.64$, $p < .001$, $d = 1.78$, and permuted displays, $t(19) = 2.99$, $p = .008$, $d = 0.95$.

Of particular relevance to the primary experimental questions was the significant Presentation Mode × Display Type interaction, which indicates that the pattern of performance between relatable, rounded, and permuted displays was modulated by presentation mode/duration. A follow-up ANOVA comparing the dynamic condition with the 80-ms static condition showed a reliable Presentation Mode × Display Type interaction, $F(2, 38) = 10.56$, $p < .001$, $\eta_p^2 = .36$. Critically, however, the follow-up ANOVA comparing dynamic presentation with static presentation at 440 ms indicated no reliable Presentation × Display interaction, $F(2, 38) = .077$, $p = .93$, $\eta_p^2 = .004$, *ns*. These analyses indicate that the 80-ms static exposure condition yielded a different pattern of performance from either the 440-ms static or dynamic conditions.

As noted above, Figure 14 suggests that one key difference in the patterns of performance between presentation conditions is that relatable and rounded displays produced similar performance at 80-ms exposures but different levels of performance for 440-ms and dynamic exposures. Individual comparisons between relatable and rounded displays showed reliable differences for dynamic,

$t(19) = 3.77$, $p = .001$, $d = 1.19$, and 440-ms static presentations, $t(19) = 3.88$, $p = .001$, $d = 1.23$, but not for 80-ms static presentations ($p > .10$).

Given that the overall level of performance for dynamically presented relatable displays fell between the static 80-ms condition and 440-ms condition relatable displays, we conducted an exploratory analysis of the effective exposure time of dynamic and static presentations. For this analysis, we plotted average sensitivity as a function of exposure time for static displays for the 0/5, 0/15, and 0/25 misalignment conditions. We then fit linear functions to the data relating sensitivity to time of exposure and determined the slopes and intercepts for the three functions. Once these linear functions were determined, we used the average sensitivity on dynamic displays to solve for the effective exposure time (see Table 1). The effective exposure durations ranged from 260 ms to 355 ms depending on the misalignment condition in question.

Discussion

There are three major results of this experiment. One is that performance on dynamic displays reliably exceeded performance on static displays having the same physical exposure time (80 ms). This result shows that performance in the dynamic occlusion task is better than one would expect based on the total exposure duration of dynamically occluded objects visible through apertures in the occluder. This finding supports the persistence and position updating notions of STR, both of which would be necessary to keep information available and spatially coherent for discriminating positional relations.

The second major result from this experiment is that the pattern of performance observed for static displays lasting 80 ms differed from both static displays of 440 ms and dynamic displays in several ways. This outcome was predicted based on the use of an exposure duration at which interpolation effects were not expected to be asymptotic (Guttman & Kellman, 2004; Ringach & Shapley, 1996). The superiority of performance for the dynamic condition over the 80-ms static condition was found only for relatable and rounded displays. Dynamic and 80-ms static presentations did not differ for permuted displays, in which object formation was not predicted to occur in any condition.

These data support the idea that partial interpolation effects occurred in the 80-ms static condition. We have already noted that the level of performance for static relatable displays at 80 ms was not as high as in either dynamic presentation or the longer static presentation. Another interesting finding in the 80-ms presentation condition was the lack of a difference between relatable and rounded displays. Performance for both of these display types exceeded that for permuted displays. However, whereas differences were modest for rounded displays between the 80-ms condition and the other two conditions, differences were substantial for relatable displays. What might explain these aspects of the data?

These results cohere with other work on the time course of contour interpolation effects and the notion of coarse-to-fine processing in contour completion. Recall the results of Guttman and Kellman (2004) showing that interpolation effects are evident by 60 ms but not asymptotic until at least 120 ms. Specifically, they found that dot localization accuracy and precision (relative to a perceived interpolated edge) is not as good at 80 ms as at 120 ms or longer durations. These data are consistent with the idea that contour information is registered in a coarse-to-fine manner across spatial scales. Research indicates that coarse (lower spatial frequency) information influences visual processing earlier than fine (higher spatial frequency) information in a variety of visual tasks, including spatial discriminations (Watt, 1987), processing of hierarchical displays (Navon, 1977), and stereoscopic depth processing (Glennerster, 1996; Marr & Poggio, 1979; Menz & Freeman, 2004). A plausible explanation of our results is that at 80 ms, high-spatial-frequency information is incompletely registered, causing tangent discontinuities at low frequencies to be processed similarly for the rounded and relatable conditions. At longer exposures, when higher spatial frequencies register, the attenuating effect of rounding on interpolation is seen (because tangent discontinuities have been removed at the finer scales). The fact that the difference between relatable and rounded conditions in our studies involved a mere 6% of pixels removed from corners of the displays is consistent with the need to process fine-scale information to see differences between these conditions. If this interpretation is correct, it tends to support the idea that the weakening, rather than elimination, of interpolation effects in previous rounding studies (Albert, 2001; Shipley & Kellman, 1990; Wurtz & Lourens, 2000) is due to the survival of tangent discontinuities at low frequencies.

The third major result is the close match between patterns of performance for the 440-ms static exposure and dynamic displays. A slightly higher overall level of performance was found in the 440-ms condition, but none of the effects of relatability, rounding, permuting, or misaligning object fragments was found to interact with the presentation mode (dynamic vs. 440-ms static). These results and those of Experiment 1B support the theory of STR:

Table 1
Effective Exposure Time of a Dynamic Display Estimated From Performance Levels in Static Displays

Misalignment comparison	Average sensitivity (d')			Slope	Intercept	Effective exposure time of dynamic (ms)
	Dynamic	Static 80 ms	Static 440 ms			
0/5	0.31	−0.01	0.63	.0018	−0.15	259.80
0/15	1.93	1.20	2.36	.0032	0.94	308.51
0/25	2.77	2.12	2.98	.0024	1.93	355.22

Object formation effects in dynamic displays match those in much longer static exposure conditions.

An exploratory analysis to quantify this effect showed that observers behaved as if the dynamic displays were visible for 260–355 ms. These estimates are three to four times as long as the actual physical exposure time of 76 ms for dynamic displays. This analysis indicates that although the persistence and position updating processes are not perfect at preserving the perceptual availability of displays during occlusion (because dynamic performance was lower than in the 440-ms condition), they nonetheless greatly increase the effective availability of visual information during dynamic occlusion. At least one caveat applies to this analysis. The analysis assumed a linear relationship between exposure time and sensitivity values. At the highest misalignment comparison (0/25), there appear to be ceiling effects in the data, which amount to a nonlinearity. A similar argument could be made for floor effects in the lowest misalignment comparison. The 0/15 estimate of an equivalent exposure of 309 ms may be best for avoiding these concerns. We offer this estimate as a first approximation of the effective exposure time of these dynamic occlusion displays.

In sum, the results of Experiment 2 show that sensitivity to misalignment in dynamic displays is considerably better than would be expected from the mere physical exposure time of the pieces through the apertures. Moreover, differences in the pattern of effects at 80 ms from those in the dynamic and 440-ms cases suggest that the method of Experiment 2 can be used to probe the microgenesis of interpolation effects, such as the time courses of the influence of information at different spatial scales. These results, along with the strong correspondence between relatable, rounded, and permuted display effects in the dynamic case and in simultaneous, stationary displays at a long (440-ms) exposure, support the persistence and position updating hypotheses of STR.

Experiment 3: The Perception of Shape Relations in Dynamic Illusory Objects

In object formation under dynamic occlusion, the visual system assembles information over time and uses it to make connections across spatial gaps. However, this is not the only phenomenon in which shape representations result from connecting contours and surfaces across intervals in space and time. In illusory contour perception, interpolated contours and surfaces are formed, and these appear in front of surrounding inducing elements. Although most often studied in static, 2D displays, illusory contours can also be formed from information given sequentially in time by motion (Bruno & Bertamini, 1990; Kellman & Cohen, 1984).

Perhaps the most common way of inducing static illusory contours is to use patterns that appear to have sections cut out from them (as in the partial circles used in the well-known Kanizsa triangle). The cut-out sections become the physically defined parts of the illusory object that forms. More generally, it has been found that static illusory contours are created by the presence of tangent discontinuities and relatable contours leading into them (Kellman, Garrigan, & Shipley, 2005; Kellman & Shipley, 1991). In dynamic illusory objects, such physically specified edge fragments change over time, as by progressive partial occlusion of a form by a moving figure (which is otherwise invisible because it matches the background lightness and color). Such occlusion events may be arranged so that momentary views are inadequate for static interpolation processes to produce contour completion and object formation. That kinetic illusory contours form under such circumstances was discovered by Kellman and Cohen (1984) and later investigated by Bruno and colleagues (Bruno, 2001; Bruno & Bertamini, 1990). Processes of dynamic contour formation have also been investigated in dynamic displays having no oriented edge fragments in the stimulus (Bruno, 2001; Shipley & Kellman, 1994, 1997).

In nature, dynamic illusory objects correspond to situations such as an owl gliding across a starry sky or a camouflaged stingray moving across the ocean floor. In these situations, the moving animal is not occluded, but its bounding contours are only partially specified because of camouflage or a lack of contrast against the background. Dynamic illusory object perception is similar to dynamically occluded object perception in that edge information is revealed piecemeal over time at different places in the visual field, and interpolation across gaps occurs to form objects.

Until recent years, many investigators considered completion of partly occluded objects and of illusory objects to be distinct processes. For example, Michotte et al. (1964), although arguing that both involve processes of perceptual organization, distinguished these as *modal* and *amodal* completion (where modal refers to the presence of sensory attributes or modes in the resulting percept). In subsequent work, illusory contours have received widespread attention from researchers studying basic visual processes, but unit formation and representation of hidden parts under occlusion have sometimes been considered more cognitive phenomena. Because illusory contours have a sensory presence, they can be squarely addressed as sensory or perceptual. Representations of occluded regions, however, seem odd in that the regions in question pass behind other surfaces. If a contour or surface is out of sight, in what sense can one be said to see it? Even as interpolation of occluded contours and surfaces has received more attention in recent years, it is still common for some researchers to treat occlusion as involving grouping of visible areas rather than establishment of contour and surface representations involving hidden regions. This view contrasts with that of earlier investigators who believed amodal completion to be just as perceptual as modal completion (Kanizsa, 1979; Michotte et al., 1964).

As we have proposed elsewhere (Kellman et al., 1998; Shipley & Kellman, 1992), modal and amodal completion may name not separate perceptual processes but rather separate modes of appearance (specifically, whether interpolated areas appear in front of or behind other surfaces in a scene). Modal and amodal completion may rely on a common contour interpolation process (Kellman & Shipley, 1991; Ringach & Shapley, 1996; Shipley & Kellman, 1992). In final scene representations, interpolated boundaries and surfaces may appear behind other objects (amodal) or in front (modal). The idea of a contour interpolation step that is shared by both illusory and occluded object formation processes has been called the *identity hypothesis* in contour interpolation (Shipley & Kellman, 1992).

The identity hypothesis does not claim that there are no differences between processing of occluded and illusory displays. Rather, they may share a common interpolation step, but each is subject to other factors that may restrict or gate interpolation processes (Kellman, 2003b; Kellman, Guttman, & Wickens, 2001; for a recent discussion, see Kellman, Garrigan, & Shipley, 2005). It is likely that early visual processes produce a number of edge interpolations prior to having information about higher level constraints on scene representations—certain information about depth order, transparency (Anderson, Singh, & Fleming, 2002), and consistency of boundary assignment. Because of such constraints, not all interpolated edges appear in final scene representations (Guttman & Kellman, 2005; Kellman et al., 2001), and the exact shape of interpolated contours may be influenced by the presence or absence of perceived occlusion (Guttman & Kellman, 2004; Singh, 2004).

In stationary displays, a number of studies have shown strong empirical similarities for occluded and illusory stimuli having the same physically specified contours and gaps in terms of the determinants of interpolation (Gold, Murray, Bennett, & Sekuler, 2000; Shipley & Kellman, 1992), the time course (Guttman & Kellman, 2004), and strength of interpolation (Kellman et al., 1998; Ringach & Shapley, 1996). Other research has suggested differences in constraints on illusory and occluded appearance (Anderson et al., 2002), the precise shape of interpolated contours (Singh, 2004), or their neural substrates (von der Heydt, Peterhans, & Baumgartner, 1984). If, as has been proposed (e.g., Kellman et al., 2001), there is an early, common interpolation step followed by other constraints relating to final scene appearance, then the empirical results indicating strong commonalities and some differences between amodally and modally appearing interpolations are not in conflict.

The strongest evidence for the identity hypothesis comes from cases in which it can be shown that interpolation must be determined prior to determination of the eventual modal (illusory) or amodal (occluded) appearance of interpolated contours (Albert, 1993; Kellman, 2003a; Kellman, Garrigan, & Shipley, 2005; Kellman et al., 1998). Moreover, the discovery of interpolated contours that form between an illusory-contour-inducing element on one end and an occluded contour inducer on the other (so-called *quasi-modal completion*) supports the idea of a common interpolation process that can lead to either modal or amodal appearances based on other characteristics of a scene (Kellman, Garrigan, & Shipley, 2005; Kellman et al., 1998). A common interpolation step appears to explain these phenomena, making the identity hypothesis highly likely in static interpolation (for recent discussions, see Albert, in press; Anderson, in press; Kellman, Garrigan, & Shipley, 2005; Kellman, Garrigan, Shipley, & Keane, in press).

Dynamic Illusory Contours and the Identity Hypothesis

Although a common interpolation step seems evident in static object formation, no research has explored the identity hypothesis in dynamic object formation. Accordingly, the main purpose of Experiment 3 was to test spatiotemporal interpolation in dynamic illusory object displays and compare it with the results for the dynamically occluded object displays in Experiment 1A.

Previous research on kinetic illusory contours has shown that sequential interruption of background elements could induce perception of a rotating form with illusory edges. Shape discrimination was better when the figures were stationary and background elements moved than when rotating illusory figures were induced in front of stationary background patterns (Bruno & Bertamini, 1990; Kellman & Cohen, 1984). These studies of dynamic illusory object formation have had certain limitations. Kellman and Cohen (1984) measured accuracy on a simple shape-identification measure but did not use signal-detection methods. Both studies used rotation (in a frontoparallel plane, around a stationary center) as the figural motion despite the fact that translation is more common in ordinary environments. Accordingly, in addition to testing the identity hypothesis, the motivations for Experiment 3 included using signal-detection measures and translatory motion to confirm and extend what is known about dynamic illusory object formation.

Experiment 3 also tested another question. Were the object formation effects found in earlier experiments due primarily to contour interpolation processes, or did surface spreading play an additional role? Yin et al. (1997, 2000) showed that surface spreading under occlusion can connect regions even when their contours are not relatable. In such cases, surfaces visible at occluding edges spread behind the occluder, confined by linear extensions of visible edges, if present (Yin et al., 1997). When surface spreading within linear extensions from two visible regions meets, a unit may be perceived with indistinct boundaries. Given these facts about the surface spreading process, it is possible that STR does not connect contours. Rather, the effects of relatable displays in Experiments 1A, 1B, and 2 could have depended on surface spreading. The displays of Experiment 3, however, provided a means of testing this possibility.

In 2D illusory object displays, surface spreading alone (without contour interpolation) along linear extensions from visible edges would meet identical surface properties of the background. Surface spreading would thus not serve to connect moving visible regions to each other because it could not segment them from the background. Surface spreading can occur, of course, in 2D illusory displays when regions are connected by contour interpolation. In that case, surface spreading is confined by real and interpolated edges (producing, e.g., the characteristic apparent lightness effects seen in Kanizsa-style illusory figure displays). The upshot of the preceding analysis is that if the dynamically relatable displays of Experiment 3 produced effects similar to those of Experiments 1A, 1B, and 2, it would implicate contour interpolation processes as the cause of those effects.

Testing the Spatiotemporal Identity Hypothesis

If the same interpolation process is responsible for contour completion in both modal and amodal objects, and if this same process underlies perception of dynamically occluded stimuli, we predicted that modal and amodal figures should yield similar patterns of performance in the dynamic occlusion task.

To evaluate this prediction, the amodal stimuli used in the earlier experiments were transformed into modal stimuli (see Figures 15 and 16; see also Kellman & Shipley, 1991, for a discussion of transforming stimuli from modal to amodal and vice versa). Both the object fragments and background were white. Twelve black rectangles were placed at the same locations as the apertures in the previous experiments. The moving white object fragments were visible only when they passed in front of the black rectangles. This resulted in the perception of a very strong illusory figure that modally completed across both the horizontally and vertically unspecified regions (see Figure 15).

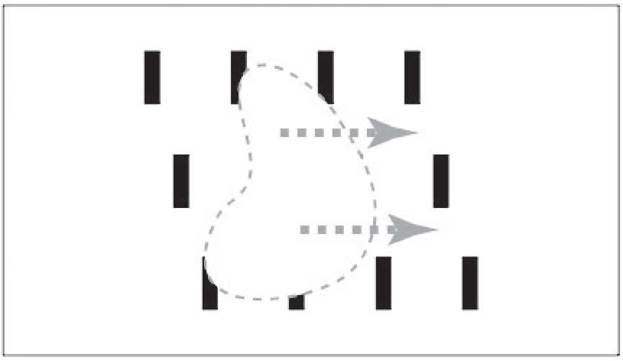

Figure 15. A schematic representation of the dynamic illusory displays used in Experiment 3. When three 0-arcmin-misaligned fragments passed in front of the 12 black rectangles, the resulting percept was of a completed illusory figure that spanned the horizontally and vertically undefined space between the black inducing regions.

Method

All aspects of the method were the same as in Experiment 1A, except where noted.

Participants. Twenty undergraduates (8 men and 12 women, mean age 20.1 years) from the University of California, Los Angeles, participated in the experiment in partial fulfillment of requirements for an introductory psychology class. All participants reported normal or corrected-to-normal vision and were naïve as to the purposes of the experiment.

Stimuli. The stimuli used in this experiment were illusory versions of the stimuli used in Experiments 1A, 1B, and 2 (see Figures 15 and 16, along with the description above).

Results

The ability of participants to discriminate between a 0-arcmin-misaligned stimulus and other values of misalignment is depicted in Figure 17. Participants performed best in the relatable condition overall, followed closely by the rounded condition. Participants were again least accurate in the permuted condition. These results with dynamic illusory objects closely resemble those of Experiments 1A and 2 for dynamically occluded objects.

We performed a 3 × 9 (Display Type × Misalignment) ANOVA on the sensitivity scores. The statistical analyses revealed main effects of display type, $F(2, 38) = 22.38$, $p < .001$, $\eta_p^2 = .54$, and misalignment pair, $F(8, 152) = 38.87$, $p < .001$, $\eta_p^2 = .67$, as well as a Display Type × Misalignment interaction, $F(16, 304) = 4.23$, $p < .001$, $\eta_p^2 = .18$. Planned comparisons between the display types established that sensitivity to misalignment was reliably greater in the relatable condition than in the permuted condition, $t(19) = 7.87$, $p < .001$, $d = 2.49$, and marginally greater than in the rounded condition, $t(19) = 2.06$, $p = .054$, $d = 0.65$, *ns*. Sensitivity was also greater in the rounded condition than in the permuted condition, $t(19) = 4.21$, $p < .001$, $d = 1.33$. Nineteen of 20 observers performed better in the relatable than in the permuted condition, whereas 18 of 20 performed better in the rounded than in the permuted condition. Thirteen of 20 participants were more sensitive to relatable than to rounded displays.

To test for differences among the data patterns for Experiments 1A, 1B, and 3, we conducted a 3 × 3 × 9 (Experiment × Display Type × Misalignment) mixed ANOVA. This analysis showed no reliable main effect of experiment, $F(2, 47) < 1$, $p = .40$, $\eta_p^2 = .04$, *ns,* nor did the experiment factor interact with any other factor (all *p*s > .20, *ns*).

As in Experiments 1A and 1B, we computed sensitivity to misalignment as a function of visual units, comparison task, and display type and then calculated the superiority in d' for the relatable and rounded conditions over the permuted condition (see Figure 18). Using *t* tests examining the advantage of relatable over permuted displays for the different comparison groups, we found that the relatable advantage was larger in the 1vs3 task than in the 3vs3 task, $t(19) = 3.77$, $p = .0013$, $d = 1.19$, and in the 1vs3 task compared with the 1vs1 task, $t(19) = 4.82$, $p = .001$, $d = 1.52$. There were no other reliable differences between groups (all *p*s > .0056), though the rounded condition showed a marginally significant advantage of the 1vs3 task over the 3vs3 task ($p = .0093$, *ns* relative to the Bonferroni corrected significance level of .0056). This post hoc analysis showed that the relatable condition benefited more from the 1vs3 task than the permuted condition did.

Discussion

Perceptual processing of dynamic illusory contour displays in Experiment 3 mirrored that of dynamic occlusion and static occlusion displays in virtually all respects. The superiority of spatiotemporally relatable displays provides evidence that the geometry of relatability applies to object formation processes regardless of whether the inputs are static or dynamic or whether the final appearance is either an illusory or occluded object. The fact that dynamic modal, dynamic amodal, and static amodal displays yielded similar patterns of performance in this paradigm provides evidence that the identity hypothesis holds for the spatiotemporal contour interpolation process. However, as with any theoretical assertion that relies on a null result, these conclusions should be accepted with some caution.

As noted above, the strongest arguments for a common interpolation component in static interpolation come from displays in which, logically, interpolation must precede determination of final modal/amodal appearance (Albert, 1993; Kellman, 2003b; Kellman, Garrigan, & Shipley, 2005; Kellman et al., 1998). The present results in dynamic interpolation involve empirical similarities; investigations of dynamic versions of phenomena involving

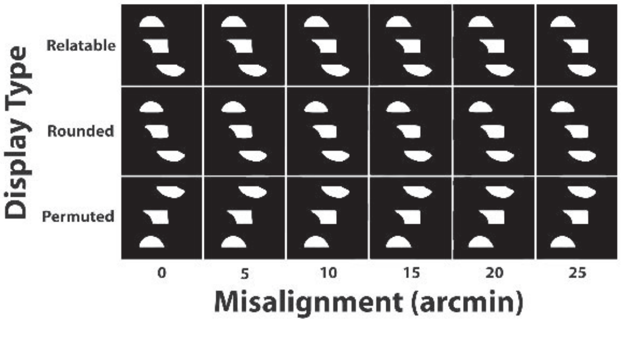

Figure 16. The modal versions of the stimuli used in Experiment 3. The white figures translated in front of small black rectangles on a white background (rather than a black background, as pictured here).

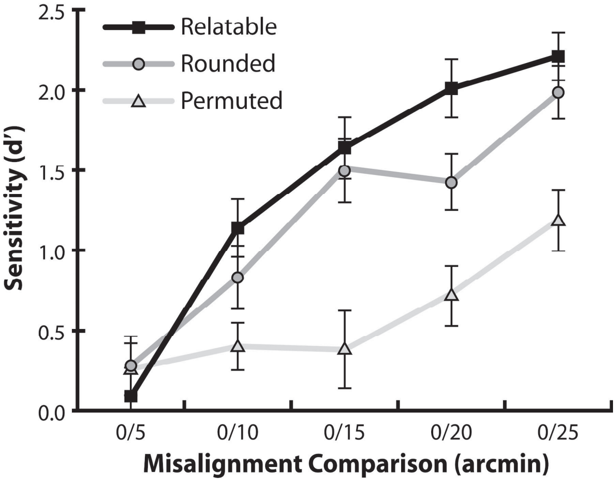

Figure 17. Main results for Experiment 3 (dynamic illusory): Sensitivity (d') by display condition as a function of misalignment difference between target and distractor images. Error bars indicate ±1 standard error of the mean.

these same logical dependencies (and of quasi-modal interpolations) would be useful. We are currently investigating these issues.

As with the earlier experiments, the results of Experiment 3 support the notion that STR allowed object formation, which in turn conferred an advantage for participants in performing the discrimination task. The results also indicate that contour interpolation, rather than surface spreading, produced the advantages in task performance observed in all four experiments for relatable displays. The surface spreading process is unlikely to operate in the illusory displays to form unitary objects except as a consequence of contour interpolation.

Experiment 3 also provides the first objective evidence that we know of for illusory contour formation from translatory motion where edge interpolation is based on nonsimultaneous information. The full range of and constraints on illusory object motions are not known and remain an important topic for future study.

In sum, the findings support the generality of dynamic illusory object formation. It works with translations, as well as rotations, and depends on similar variables as in the static case. The findings offer support for a unified treatment of contour interpolation for object formation in both dynamically and statically occluded, as well as kinetic illusory, displays.

General Discussion

The goal of the present work was to create and test a theory of the visual processes that allow for perception of unitary objects with determinate shape from information that is fragmentary in both space and time. STR explains dynamic object formation by a combination of persistence and position updating of sequentially viewed fragments, allowing application of the geometry of spatial relatability to the currently seen and previously seen (positionally updated) parts. To test this theory, we used an experimental paradigm that assessed observers' ability to perceive the alignment of object fragments that were visible through an occluder with multiple, narrow, spatially misaligned apertures. Under these conditions, very little figural information was available at any moment, and large regions of the occluded objects were never projected to the eyes. Specifically, the horizontally and vertically misaligned apertures allowed for projection over time of roughly 56% of the objects' shapes; the other 44% of each object was never seen. Moreover, those fragments that did project to the eyes did so sequentially in time and in spatially disconnected locations, making the task one of spatiotemporal interpolation. We hypothesized that displays that fulfilled the conditions of STR would lead to object formation, which in turn would produce superior performance in this task relative to two kinds of control groups.

Conditions that fulfilled the requirements of STR yielded discrimination performance markedly superior to conditions that did not. Experiment 1A showed that when relations of object fragments revealed over time satisfied the criteria for STR, performance was superior to a condition having the same object fragments but not satisfying STR. Experiment 1A also examined a separate manipulation known in static displays to reduce contour interpolation: the rounding of corners to eliminate tangent discontinuities. As predicted, rounded figures produced a reduced level of performance in dynamic occlusion displays whose fragments were otherwise relatable. Taken together, these results support the ideas that STR describes conditions for dynamic object formation, that object formation produces better performance in encoding object fragments and their positional relations, and that tangent discontinuities are an important ingredient in spatiotemporal interpolation.

In Experiment 1B, we performed a direct test to verify that well-known unit formation effects from the static domain would occur in the discrimination paradigm we used to assess dynamic object formation. We found the expected effects of static spatial

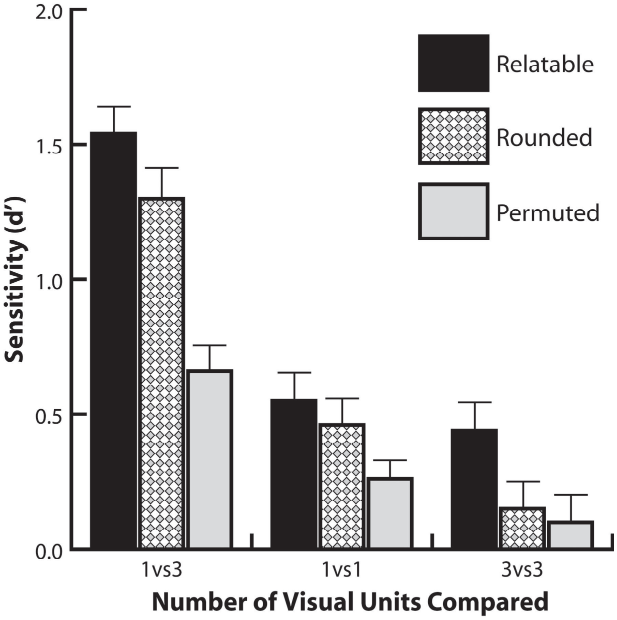

Figure 18. Sensitivity (d') to misalignment as a function of the number of visual units in the display and the type of comparison being performed in Experiment 3 (dynamic illusory). Error bars indicate ±1 standard error of the mean.

relatability, misalignment, and rounding of tangent discontinuities. Strikingly, comparisons of total time of exposure between Experiments 1A and 1B showed roughly equivalent performance for dynamic and static displays despite the dynamic presentations having less than a third of the total exposure time of the static presentations. The evidence suggests not only that STR operates to allow perception of objects from spatially and temporally fragmentary input but that it does so, under the conditions tested, with remarkable precision.

These issues were more deeply investigated in Experiment 2, in which we compared dynamic object formation with static object formation at two key exposure durations, 80 ms and 440 ms. These values corresponded to the cumulative physical exposure time for object parts or to the maximum potential perceptual availability of moving, occluded fragments (assuming STR), respectively. A major finding is that performance in dynamic displays was reliably better than one would expect from the corresponding physical exposure duration (~80 ms). Another important result is that the pattern of performance for dynamic displays matched the pattern observed in displays with an exposure time (440 ms), similar to what STR might ideally achieve.

As indicated in other research, the 80-ms static condition was expected to show partially developed, but non-asymptotic, interpolation effects. This expectation was confirmed in three ways. Relatable displays in the 80-ms condition showed an advantage over permuted displays but did not attain the levels of performance seen in dynamic or 440-ms static presentation. In 80-ms static presentation, permuted displays, which should not show interpolation effects in any condition, produced performance equivalent to permuted displays in the dynamic presentation case. Finally, at 80 ms of exposure, relatable and rounded displays did not reliably differ, unlike the dynamic and 440-ms conditions, in which they did differ. This last result also makes sense in terms of incomplete interpolation effects, especially if contour interpolation proceeds in a coarse-to-fine manner between 60 ms and 120 ms (Guttman & Kellman, 2004).

In contrast, dynamic presentation showed the same relative pattern of results for relatable, rounded, and permuted displays as found in the 440-ms presentation condition; there was no hint of any statistical Presentation Mode × Display Type interaction for these two conditions. The 440-ms condition involved an exposure duration well above that needed for asymptotic static interpolation effects (Guttman & Kellman, 2004; Ringach & Shapley, 1996; Sekuler & Palmer, 1992). The fact that the dynamic condition closely mirrored the longer static condition indicates that the persistence and position updating notions of STR allow object fragments captured over time to be preserved and coherently integrated, achieving a result that is much like simultaneous, continuous, static presentation of much longer durations.

Finally, Experiment 3 provides evidence that, as in static interpolation, the effects of STR are highly similar for dynamic occluded and illusory displays. As in the static case, the evidence from dynamic object formation is consistent with the identity hypothesis—the idea that interpolation in both occluded and illusory cases operates similarly and may depend on a common mechanism. Experiment 3 also exploited limitations of surface spreading effects in 2D illusory contour displays to reveal that the object formation effects in these experiments depend on contour interpolation processes.

Static and dynamic object formation may be different manifestations of a single, general process. In perception of objects under dynamic occlusion, processes of persistence and position updating allow moving fragments to be brought into spatial register and enter into the same spatial relatability computation as in static object formation. Although static spatial interpolation has received far more study than spatiotemporal interpolation, it may nonetheless be considered a limiting case of the latter, having zero values of persistence and position updating.[7] The unification of static and dynamic object formation emerges from the realization that nonzero values of persistence and position updating are possible in the formation of objects across gaps.

Processes of Spatiotemporal Relatability

The persistence and position updating hypotheses are consistent with research in related domains, such as the visual icon, anorthoscopic perception, and multiple object tracking. The theory of STR combines these with the known geometry of relatability in static unit formation and its ecological bases (Field et al., 1993; Geisler et al., 2001; Kellman & Shipley, 1991). Combining representations that preserve and positionally update information about object fragments with spatial relatability constraints yields a common geometry for both spatial and spatiotemporal interpolation and provides a unified account of visual object formation. Apart from this account, it is not clear how one might explain the results of the current experiments, as well as prior research on relatability and tangent discontinuities in the perception of occluded and illusory objects.

One question that may be asked about how these processes combine is whether, in a particular viewing episode, all visible areas must be registered and positionally updated prior to the operation of relatability. Some evidence suggests this is not the case. As soon as two object fragments that span a gap have been received, it is likely that relatability can act to produce interpolation. E. M. Palmer and Kellman (2001, 2002) performed a number of experiments with only two object fragments and two apertures. In these situations, object formation occurred, although it was subject to some interesting velocity-dependent illusions; these phenomena will be addressed in a future paper. Piecewise interpolation of fragments that later become integrated with others was suggested in a related context by Shipley and Kellman (1994).

Formalizing Spatiotemporal Relatability

We have considered the basic ideas of STR and their fit with present and prior experimental results. Available data support the explanatory value of STR in understanding dynamic object perception. Further developments in this area may benefit, however, from a more precise formulation of STR. Accordingly, we present here a formalization of the spatiotemporal edge relations giving rise to interpolation. We define, for a given edge and any arbitrary point, the range of orientations at that point that fall within the limits of relatability. Although recent work has indicated that relatability operates three dimensionally (Kellman, Garrigan, &

---

[7] We thank Myron Braunstein for suggesting that static interpolation may be best viewed as a limiting case, requiring minimal persistence and zero position updating.

Shipley, 2005; Kellman, Garrigan, Shipley, et al., 2005), for simplicity we describe STR in a 2D (x, y) coordinate system. This coordinate system is object centered, much like the *stable features frame* proposed by Feldman (1985) or the *distal reference frame* discussed by Shipley and Cunningham (2001). When an object is tracked by smooth pursuit eye movements, this frame also corresponds to a retinotopic frame, although, as we noted earlier, evidence indicates that such tracking is not a requirement for position updating.

For convenience, we choose the coordinate system such that the tip of one physically specified edge ends at the origin of the coordinate system (0, 0) and its tangent vector lies along the x-axis (see Figure 19). (This arrangement simplifies the description. Application to any edge in any orientation can be accomplished by translating and/or rotating the coordinate system to align the x-axis with that edge.) For such a surface edge, the possible 2D-relatable edges constitute a set of orientation and location combinations where the set of eligible locations is given by $x > 0$. Without loss of generality, we assume that $y \geq 0$ (if not, invert the y-axis). In this framework, 2D spatial relatability can be expressed as follows. For an input edge terminating at any eligible location (x, y), the set of relatable orientations at that location is given by

$$\arctan\left(\frac{y}{x}\right) \leq \Theta \leq \frac{\Pi}{2}, \qquad (1)$$

where Θ is the angle between the contour and the x-axis in the xy-plane. The lower bound of this equation expresses one limit of relatability: where the linear extension of the input edge intersects the tip of the reference edge. The upper bound expresses the 90° constraint. Within these limits, we would expect quantitative variation in strength of interpolation.

This basic definition can be extended to define STR for translations in the xy-plane given by the coordinate system. For an edge that has been visible and becomes occluded (or disappears against a background due to lack of contrast, as with dynamic illusory objects), the persistence and position updating hypotheses specify that the edge continues to be represented for some period of time and that its position is updated based on velocity information obtained while it was visible. Position updating requires only that spatial coordinates relevant to the relatability computation change over time; we find it convenient to formalize these changes in terms of the notion of *occlusion velocity,* V_{OCC}, although we do not intend any specific commitment to a particular form for this computation at this time. This velocity is not the physical velocity but a represented change in position for a previously seen edge (e.g., Shioiri et al., 2000). In other research (E. M. Palmer, 2003), we have determined that under the conditions of multiple apertures in the experiments here, V_{OCC} is very close to real velocity. Thus, we do not have much to say here about how V_{OCC} may differ from the actual velocity of a previously viewed object part. (In other circumstances, V_{OCC} differs strikingly from actual velocity and leads to perceptual illusions; E. M. Palmer & Kellman, 2001, 2002, 2003.)

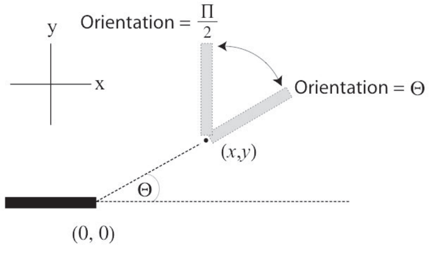

Figure 19. Spatial relatability: At position (x, y), limits of two-dimensional relatability for any oriented edge ending in a tangent discontinuity at that point are given by two orientations: Θ and $\Pi/2$, where $\Theta = \arctan(y/x)$. This geometry applies to simultaneously visible edges.

The velocity of an object behind an occluding surface (or in the absence of background contrast, for the illusory case), V_{OCC}, has both x and y directional components, which may be written as $V_{OCC(x)}$ and $V_{OCC(y)}$, respectively. Let the time that a portion of the object has been occluded be described by T_{OCC}. In general, the position updating component of STR can be described by $V_{OCC} * T_{OCC}$, and more specifically, $V_{OCC(x)} * T_{OCC}$ is the represented position change in the x direction, and $V_{OCC(y)} * T_{OCC}$ is the represented position change in the y direction (see Figure 20). Consequently, the set of spatiotemporally relatable edges is confined to points where $x + (V_{OCC(x)} * T_{OCC}) \geq 0$. Again, we consider for simplicity the case where $y + (V_{OCC(y)} * T_{OCC}) \geq 0$. For another edge that becomes visible at an eligible point at time T and has its endpoint at (x, y), the set of possible orientations at that point and that time satisfying STR is given by Θ, such that

$$\arctan\left(\frac{y - (V_{OCC(y)} * T_{OCC})}{x - (V_{OCC(x)} * T_{OCC})}\right) \leq \Theta \leq \frac{\Pi}{2}.$$

This definition of STR can be readily extended to motions including rotations or combinations of rotation and translation and, as mentioned above, to three-dimensional relations and motions. We should also note that T_{OCC}, as it is currently formulated, has no upper bound. Intuition (and work on the visual icon; e.g., Loftus, Duncan, & Gehrig, 1992) suggests that the persistence time of an occluded region is not infinite but rather must have a maximum value. The decay function of representations of occluded object fragments is a research issue we are currently investigating (see below).

Representations and Processes in Dynamic Object Formation

Although consistent with known visual persistence phenomena, the present results and the theory used to explain them seem to define a particular kind of representation used in visual processing. We suggest the name *dynamic visual icon* for this representation. The dynamic visual icon is an extension of the concept of the visual icon first demonstrated by Sperling (1960) and later named by Neisser (1967). The dynamic visual icon is dynamic because it not only maintains a representation of seen regions after occlusion but also updates their positions over time according to their previously observed velocities. As reviewed earlier, properties of the dynamic visual icon are implicit in other known phenomena. Our notion of the dynamic visual icon is therefore not a wholly new idea. The properties of the dynamic visual icon differ substantially, however, from representations that have been previously described to hold information beyond the duration of a stimulus. It is possible that the familiar notion of a visual icon is simply more versatile

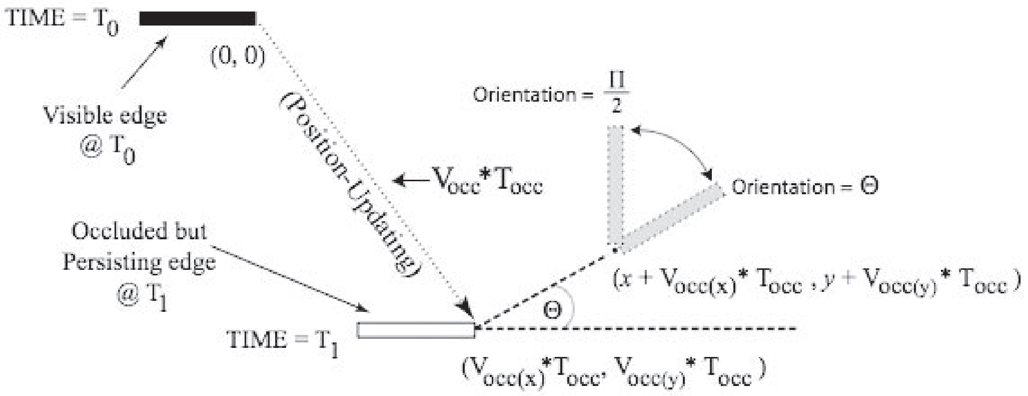

Figure 20. Spatiotemporal relatability: Persistence and position updating (based on occlusion velocity) produce geometrically appropriate relations between currently visible and previously seen edges. T_{OCC} = time that a portion of an object has been occluded; V_{OCC} = velocity of an object behind an occluding surface.

than usually described (encompassing position change for moving objects). Although this possibility cannot be ignored, we believe that a distinct label has at least heuristic value. It calls attention to the idea either that there is a distinct dynamic visual icon or that some already identified visual iconic representation is dynamic.

As we discussed earlier, the dynamic visual icon could accomplish position updating in some circumstances in connection with smooth pursuit eye movements. Stimulus velocity signals used to execute these eye movements carry the relevant information for position updating, and it is conceivable that reafferent feedback from the eye movements could provide information. The dynamic visual icon, however, is not intrinsically related to or dependent upon eye movements (Fendrich et al., 2005; Haber & Nathanson, 1968; Shipley & Cunningham, 2001). It is best thought of as a visual representation that incorporates extrapolation of trajectories in a distal reference frame of previously viewed fragments based on their velocity information, whether or not these fragments are optically pursued.

So far, we have not closely examined the inputs into the dynamic visual icon. Prior to representing parts that have gone out of sight, visual processes must extract a *dynamic visible regions* representation that includes information about contour shape, motion (speed and direction), junctions, and boundary ownership of the visible portions of the object. Surface properties are also likely part of this representation. In particular, contour junctions and boundary ownership can provide information about which contours in the dynamic visible regions and dynamic visual icon representations can be linked with other contours from the object that are revealed later in time. This proposal for a visible regions representation that includes observable regions and labeling of edge ownership as an input to interpolation processes has been suggested in the static case (Kellman, 2003b; Kellman et al., 2001) and has some similarities with the common region notion proposed earlier by S. E. Palmer (1992), although there are also important differences (for a discussion, see Kellman, 2003b).

A schematic of the STR process is presented in Figure 21. Object fragments and certain labeled characteristics (e.g., velocity, boundary ownership) in the dynamic visible regions representation are used to generate the dynamic visual icon. The visible persistence and position updating elements of the dynamic visual icon allow recently occluded regions of the object to remain in spatial alignment with currently seen regions of the object in the dynamic visible regions representation. One reason for separating the dynamic visible regions and dynamic visual icon representations is the phenomenological observation that although both appear to

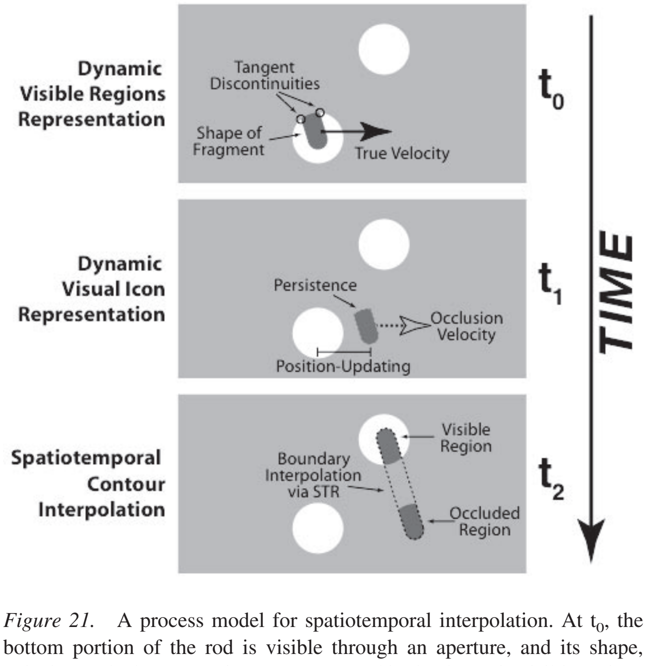

Figure 21. A process model for spatiotemporal interpolation. At t_0, the bottom portion of the rod is visible through an aperture, and its shape, velocity, and edge properties are represented in the dynamic visible regions representation. Next, at t_1, during the time that the rod is entirely occluded, its shape and position are maintained within the dynamic visual icon representation as it moves behind the occluding surface. When the top portion of the rod becomes visible at t_2, the shape and position of the occluded portion of the rod are available from the dynamic visual icon representation and may be combined with the top portion of the rod via the boundary interpolation processes proscribed by spatiotemporal relatability (STR).

have perceptual reality (E. M. Palmer, 2003; E. M. Palmer & Kellman, 2001, 2002, 2003), the observer is generally aware of whether parts of an object are occluded or not at any moment.

The third computational component of spatiotemporal interpolation is that currently visible regions and those in the dynamic visual icon, with updated spatial positions, enter into the standard spatial relatability computation, as described by Kellman and Shipley (1991) and recently generalized to three-dimensional object formation (Kellman, Garrigan, & Shipley, 2005). The addition of the dynamic visual icon representation to the static spatial relatability geometry allows this geometry to apply to dynamically occluded objects while at the same time preserving its ability to apply to static, occluded, and illusory objects.

We are currently agnostic about the nature of the representational format in the dynamic visual icon and dynamic visible regions representations. It is convenient to talk about these representations as continuous and analogue in nature. We have referred to the updated position of a fragment represented in a mental representation. In truth, our mathematical expression of STR is compatible with any implementation that accurately keeps track of time, velocity, and distance relationships in a spatial framework. So long as object interpolation is constrained to occur only with the proper spatial relations between previously viewed and currently viewed regions and interpolation produces a shape representation consistent with relatability and represented positions of parts, STR is satisfied. For example, a computation that simply increments certain numbers based on the previously given velocity signal could support STR without a truly analogue representation.

One reason we are neutral about the exact form of the representation is that our data do not much constrain the possibilities. Some ways of thinking about the dynamic visual icon raise issues that have become classic in debates on mental imagery (e.g., Kosslyn, 1983; Pylyshyn, 1973) and also certain disreputable ideas such as homunculi (who might view moving fragments in the Cartesian theatre of the mind; Dennett, 1991).

We do believe that phenomena suggesting the existence of the dynamic visual icon resemble a number of other phenomena in cognition that raise issues of analogue representation and continuity in position tracking. These phenomena include some that depend almost solely on mental representations (e.g., mental rotation) but also more mundane phenomena of seeing continuous trajectories of moving objects, tracking to reach for a moving object, or organizing more complex actions (e.g., hitting a tennis ball). In none of these cases are there precise accounts about the neural substrates or representational format, yet all serve the needs for spatially organized behavior. What these observations suggest is that the representation in the dynamic visual icon may be mysterious but no more mysterious than the representations people use to deal with motion and space in ordinary perception of fully visible objects. In fact, these phenomena involve the same issues as the dynamic visual icon—what representational system can perform anticipatory computations about spatial positions of moving objects? Whether by lists of numbers that increment or by some more intrinsically analogue system (Shepard & Chipman, 1970), these computations are important in perception, action, and cognition.

A particular issue related to analogue representation is the issue of continuity. Instead of continuous updating in the dynamic visual icon, why not describe our results in some more discontinuous fashion? For example, it would be possible to imagine that only at the moment some new fragment appears is the position of a previously seen fragment calculated. What may seem superfluous is the notion that previously seen and now occluded fragments occupy intermediate positions along the way. On this issue, the phenomena of these experiments and other data suggest that the notion of continuous updating in some representational format is the correct one. In the illusory object displays of Experiment 3 (in the relatable condition with 0 misalignment), for example, one sees the moving fragment continuously in between places where its contours are physically specified (see Figure 15). Moreover, other research has offered relatively direct tests of and support for the notion of occluded position (E. M. Palmer, 2003; E. M. Palmer & Kellman, 2001, 2002, 2003; Shioiri et al., 2000) and occlusion velocity (see also De Valois & Takeuchi, 2001).

Although the phenomenological persistence of a shape beyond and between inducing elements is especially striking in dynamic illusory contour displays, we believe that the same process occurs in dynamically occluded objects. As we have argued elsewhere (Kellman, 2003b; Kellman, Garrigan, & Shipley, 2005; Kellman & Shipley, 1991; Shipley & Kellman, 1992), this illusory/occluded difference in phenomenology derives not from lesser perceptual reality of an interpolated, occluded surface but from the visual system's encoding of whether surfaces are in front of or behind other surfaces. In our experimental task, participant performance on the two types of displays (occluded and illusory) showed all the same patterns and was not statistically distinguishable. As noted above, the simplest interpretation of this evidence is that the same contour interpolation processes are engaged by dynamic occluded and illusory objects despite the fact that the two types of displays differ in appearance.

Two aspects of the dynamic visual icon representation seem important to explore in future research. One involves the limits on visual persistence in the dynamic visual icon. How long after the dynamic visible regions representation ends does a fragment's representation in the dynamic visual icon last? In the related perceptual phenomena of spatiotemporal boundary formation, of which accretion and deletion of texture are the most common example (Gibson et al., 1969), a clear persistence limit of around 150 ms has been demonstrated (Shipley & Kellman, 1993). In the richer stimulus situation of dynamic occlusion, in which clear oriented edge inputs are given in the stimulus, the persistence limits may be longer, and they may depend on a variety of stimulus factors. Further research is needed to provide clear answers to these questions.

A final aspect of the dynamic visual icon that deserves comment is the fidelity of position updating. In the present work, we have assumed for convenience that the velocity used in updating a fragment's position is veridical (i.e., it corresponds to its real velocity at its last appearance). The data of Experiments 1A and 3, when compared with Experiments 1B and 2, indicate that persistence and position updating operated with a remarkable degree of precision under the conditions tested. In displays with multiple apertures, as in the present studies, the assumption of veridical updating appears to be accurate; however, clear decreases in occlusion velocities, relative to real velocities, occur when only a single pair of apertures is used. In other research, we measured occlusion velocities under various conditions and found that underestimation of velocity after occlusion produces robust perceptual illusions (E. M. Palmer, 2003; E. M. Palmer & Kellman, 2001, 2002). Although not elaborated here, results showing that represented velocity in some conditions is slow relative to real velocity (De Valois & Takeuchi, 2001) provide convergent support for the basic constructs of persistence and position updating in STR.

In sum, the theory and results put forth here shed light on processes of dynamic object formation, in which objects with well-defined shapes are perceived from spatiotemporally fragmented information. Because of the pervasiveness of occlusion and motion by objects and observers, accomplishing object perception under these circumstances is both crucial and challenging. Fortunately, the processes partially revealed in the present work are well designed to provide accurate perception of objects despite fragmentation in the input. These processes appear to depend on spatial and temporal relations expressed as STR and also on specific representations, such as the dynamic visual icon. The theory and results both suggest that static interpolation may be merely a limiting case of a more general spatiotemporal process—the static case is one in which persistence and positional updating have zero values. These more general and more powerful spatiotemporal object formation processes allow perceivers to cope with their own motion and motions of objects. Via STR, the visual system combines the previously visible with the currently visible and connects these across the invisible—making possible perception of coherent objects despite gaps in both space and time.

References

Albert, M. K. (1993). Segmentation, grouping and nonlocal cooperativity in stereopsis. *Investigative Ophthalmology & Visual Science, 34,* 1438.

Albert, M. K. (2001). Surface perception and the generic view principle. *Trends in Cognitive Sciences, 5,* 197–203.

Albert, M. K. (in press). The identity hypothesis: Answers to Anderson. *Psychological Review*.

Albert, M. K., & Hoffman, D. D. (2000). The generic-viewpoint assumption and illusory contours. *Perception, 29,* 303–312.

Albert, M. K., & Tse, P. U. (2000). The role of surface attraction in perceiving volumetric shape. *Perception, 29,* 409–420.

Anderson, B. L. (in press). The demise of the identity hypothesis, and the insufficiency and non-necessity of contour relatability in predicting object interpolation: Comment on Kellman, Garrigan, and Shipley. *Psychological Review*.

Anderson, B. L., & Barth, H. C. (1999). Motion-based mechanisms of illusory contour synthesis. *Neuron, 24,* 433–441.

Anderson, B. L., Singh, M., & Fleming, R. W. (2002). The interpolation of object and surface structure. *Cognitive Psychology, 44,* 148–190.

Anstis, S. M., & Atkinson, J. (1967). Distortions in moving figures viewed through a stationary slit. *Quarterly Journal of American Psychology, 80,* 572–585.

Banks, W. P., & Barber, G. (1977). Color information in iconic memory. *Psychological Review, 84,* 536–546.

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance, 242,* 1011–1036.

Bruno, N. (2001). Breathing illusions and boundary formation in space-time. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 531–556). Amsterdam: Elsevier.

Bruno, N., & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics, 48,* 331–342.

Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology, 4,* 121–138.

Clark, S. E. (1969). Retrieval of color information from preperceptual memory. *Journal of Experimental Psychology, 82,* 263–266.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998a). The dynamic specification of surfaces and boundaries. *Perception, 27,* 403–415.

Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998b). Interactions between spatial and spatiotemporal information in spatiotemporal boundary formation. *Perception & Psychophysics, 60,* 839–851.

Demkiw, P., & Michaels, C. F. (1976). Motion information in iconic memory. *Acta Psychologica, 40,* 257–264.

Dennett, D. C. (1991). *Consciousness explained.* Boston: Little, Brown.

De Valois, K. K., & Takeuchi, T. (2001). Judging the velocity of an invisible moving object [Abstract]. *Journal of Vision, 1,* 378a.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109,* 75–97.

Di Lollo, V., & Dixon, P. (1988). Two forms of persistence in visual information processing. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 671–681.

Dixon, P., & Di Lollo, V. (1994). Beyond visible persistence: An alternative account of temporal integration and segregation in visual processing. *Cognitive Psychology, 26,* 33–63.

Duysens, J., Orban, G. A., & Cremieux, J. (1985). Velocity selectivity in the cat visual system: II. Independence from interactions between different loci. *Journal of Neurophysiology, 54,* 1050–1067.

Efron, R. (1970). The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia, 8,* 37–55.

Fantoni, C., & Gerbino, W. (2003). Contour interpolation by vector-field combination. *Journal of Vision, 3,* 281–303.

Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Sciences, 8,* 265–313.

Fendrich, R., Rieger, J. W., & Heinze, H. (2005). The effect of retinal stabilization on anorthoscopic percepts under free-viewing conditions. *Vision Research, 45,* 567–582.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research, 33,* 173–193.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41,* 711–724.

Gibson, J. J. (1966). The problem of temporal order in stimulation and perception. *Journal of Psychology, 62,* 141–149.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gibson, J. J., Kaplan, G. A., Reynolds, H. N., & Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. *Perception & Psychophysics, 5,* 113–116.

Glennerster, A. (1996). The time course of 2-D shape discrimination in random dot stereograms. *Vision Research, 36,* 1955–1968.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10,* 663–666.

Guttman, S. E., & Kellman, P. J. (2004). Contour interpolation revealed by a dot localization paradigm. *Vision Research, 44,* 1799–1815.

Guttman, S. E., & Kellman, P. J. (2005). *Promiscuous interpolation: Contour completion without perception of completed contours.* Manuscript submitted for publication.

Haber, R. N., & Nathanson, L. S. (1968). Post-retinal storage? Some further observations on Park's camel as seen through the eye of a needle. *Perception & Psychophysics, 3,* 349–355.

Haber, R. N., & Standing, L. G. (1969). Direct measures of short-term visual storage. *Quarterly Journal of Experimental Psychology, 21,* 43–54.

Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., & Kubler, O. (1998). Simulation of neural mechanisms: Representing anomalous contours. *Image and Vision Computing, 16,* 407–421.

Helmholtz, H. von (1962). *Treatise on physiological optics* (Vol. 3). New York: Dover. (Original work published 1867)

Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception* (pp. 309–331). New York: Holt, Rinehart & Winston.

Hunt, S. M. (1994). Macprobe: A Macintosh-based experimenter's workstation for the cognitive sciences. *Behavior Research Methods, Instruments, & Computers, 26,* 345–351.

Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 343–360.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development, 74,* 94–108.

Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception.* New York: Praeger Publishers.

Kellman, P. J. (2003a). Interpolation processes in the visual perception of objects. *Neural Networks, 16,* 915–923.

Kellman, P. J. (2003b). Segmentation and grouping in object perception: A four-dimensional approach. In M. Behrmann & R. Kimchi (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives—The 31st Carnegie Symposium on Cognition* (pp. 155–204). Hillsdale, NJ: Erlbaum.

Kellman, P. J., & Cohen, M. H. (1984). Kinetic subjective contours. *Perception & Psychophysics, 35,* 237–244.

Kellman, P. J., Garrigan, P. B., Kalar, D., & Shipley, T. F. (2003). Good continuation and relatability: Related but distinct principles. *Journal of Vision, 3,* 120.

Kellman, P. J., Garrigan, P. B., & Shipley, T. F. (2005). Object interpolation in three dimensions. *Psychological Review, 112,* 586–609.

Kellman, P. J., Garrigan, P. B., Shipley, T. F., & Keane, B. (in press). Interpolation processes in object perception: A reply to Anderson (2007). *Psychological Review.*

Kellman, P. J., Garrigan, P., Shipley, T. F., Yin, C., & Machado, L. (2005). 3-D interpolation in object perception: Evidence from an objective performance paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 558–583.

Kellman, P. J., Gleitman, H., & Spelke, E. S. (1987). Object and observer motion in the perception of objects by infants. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 586–593.

Kellman, P. J., Guttman, S. E., & Wickens, T. D. (2001). Geometric and neural models of object perception. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision.* (pp. 183–245). Amsterdam: Elsevier.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23,* 141–221.

Kellman, P. J., & Shipley, T. F. (1992). Perceiving objects across gaps in space and time. *Current Directions in Psychological Science, 1,* 193–199.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15,* 483–524.

Kellman, P. J., Yin, C., & Shipley, T. F. (1998). A common mechanism for illusory and occluded object completion. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 859–869.

Kerzel, D., & Ziegler, N. E. (2005). Visual short-term memory during smooth pursuit eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 354–372.

Kosslyn, S. M. (1983). *Ghosts in the mind's machine: Creating and using images in the brain.* New York: Norton.

Kratz, K. E., & May, J. G. (1990). Response persistence of cat retinal ganglion cells to the temporally discrete presentation of sinewave gratings. *International Journal of Neuroscience, 52,* 111–119.

Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 530–549.

Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research, 32,* 263–273.

Lorenceau, J., & Shiffrar, M. (1999). The linkage of visual motion signals. *Visual Cognition, 6,* 431–460.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London, Series B, 204,* 301–328.

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

Mateeff, S., Popov, D., & Hohnsbein, J. (1993). Multi-aperture viewing: Perception of figures through very small apertures. *Vision Research, 33,* 2563–2567.

Menz, M. D., & Freeman, R. D. (2004). Temporal dynamics of binocular disparity processing in the central visual pathway. *Journal of Neurophysiology, 91,* 1782–1793.

Michotte, A. (1950). On phenomenal permanence: Facts and theories. *Acta Psychologica, 7,* 293–322.

Michotte, A., Thinès, G., & Crabbé, G. (1964). *Les compléments amodaux des structures perceptives* [Amodal complements of perceptual structures]. Oxford, England: Publications University Louvain.

Morgan, M. J., Findlay, J. M., & Watt, R. J. (1982). Aperture viewing: A review and a synthesis. *Journal of Experimental Psychology: Human Perception and Performance, 34A,* 211–233.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Neisser, U. (1967). *Cognitive psychology.* East Norwalk, CT: Appleton-Century-Crofts.

Neumann, H., & Mingolla, E. (2001). Computational neural models of spatial integration in perceptual grouping. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 353–400). Amsterdam: Elsevier.

Palmer, E. M. (2003). *Spatiotemporal relatability in the perception of dynamically occluded objects.* Unpublished doctoral dissertation, University of California, Los Angeles.

Palmer, E. M., & Kellman, P. J. (2001). The aperture capture effect: Misperceived forms in dynamic occlusion displays [Abstract]. *Journal of Vision, 1,* 381a.

Palmer, E. M., & Kellman, P. J. (2002). Underestimation of velocity after occlusion causes the aperture-capture illusion [Abstract]. *Journal of Vision, 2,* 477a.

Palmer, E. M., & Kellman, P. J. (2003). (Mis)perception of motion and form after occlusion: Anorthoscopic perception revisited [Abstract]. *Journal of Vision, 3,* 251a.

Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology, 24,* 436–447.

Parks, T. E. (1965). Post-retinal visual storage. *American Journal of Psychology, 78,* 145–147.

Plateau, J. A. F. (1836). Anorthoscop. *Bulletin de l'Académie de Bruxelles, 3,* 364.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80,* 1–24.

Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research, 36,* 3037–3050.

Rock, I., & Halper, F. (1969). Form perception without a retinal image. *American Journal of Psychology, 82,* 425–440.

Rock, I., Halper, F., DiVita, J., & Wheeler, D. (1987). Eye movement as a cue to figure motion in anorthoscopic perception. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 344–352.

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception, 30,* 339–366.

Saidpour, A., Braunstein, M. L., & Hoffman, D. D. (1994). Interpolation across surface discontinuities in structure from motion. *Perception & Psychophysics, 55,* 611–622.

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

Sekuler, A. B., & Palmer, S. E. (1992). Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General, 12,* 95–111.

Shepard, R., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology, 1,* 1–17.

Shioiri, S., & Cavanagh, P. (1992). Visual persistence of figures defined by relative motion. *Vision Research, 32,* 943–951.

Shioiri, S., Cavanagh, P., Miyamoto, T., & Yaguchi, H. (2000). Tracking the apparent location of targets in interpolated motion. *Vision Research, 40,* 1365–1376.

Shipley, T. F., & Cunningham, D. W. (2001). Perception of occluding and occluded objects over time: Spatiotemporal segmentation and unit formation. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 557–585). Amsterdam: Elsevier.

Shipley, T. F., & Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Perception & Psychophysics, 48,* 259–270.

Shipley, T. F., & Kellman, P. J. (1992). Perception of partly occluded objects and illusory figures: Evidence for an identity hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 106–120.

Shipley, T. F., & Kellman, P. J. (1993). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision, 7,* 323–339.

Shipley, T. F., & Kellman, P. J. (1994). Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General, 123,* 3–20.

Shipley, T. F., & Kellman, P. J. (1997). Spatio-temporal boundary formation: The role of local motion signals in boundary perception. *Vision Research, 37,* 1281–1293.

Shipley, T. F., & Kellman, P. J. (Eds.). (2001). *From fragments to objects: Segmentation and grouping in vision.* Amsterdam: Elsevier.

Singh, M. (2004). Modal and amodal completion generate different shapes. *Psychological Science, 15,* 454–459.

Sohmiya, T., & Sohmiya, K. (1994). What is a crucial determinant in anorthoscopic perception? *Perceptual & Motor Skills, 78,* 987–998.

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs, 74*(11, Whole No. 498).

Sun, J. S., & Irwin, D. E. (1987). Retinal masking during pursuit eye movements: Implications for spatiotopic visual persistence. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 140–145.

Tse, P. U. (1999). Volume completion. *Cognitive Psychology, 39,* 37–68.

Turvey, M., & Kravetz, S. (1970). Retrieval from iconic memory with shape as the selection criterion. *Perception & Psychophysics, 8,* 171–172.

Vishton, P., Kellman, P. J., & Shipley, T. F. (1991). Scene perception under dynamic occlusion. *Bulletin of the Psychonomic Society, 29,* 458.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984, June 15). Illusory contours and cortical neuron response. *Science, 224,* 1260–1262.

Von Wright, J. (1968). Selection in visual immediate memory. *Quarterly Journal of Experimental Psychology, 20,* 62–68.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America, A: Optics, Image & Science, 4,* 2006–2021.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II [Laws of organization in perceptual forms]. *Psycologische Forschung, 4,* 301–350.

Wurtz, R. P., & Lourens, T. (2000). Corner detection in color images through a multiscale combination of end-stopped cortical cells. *Image and Vision Computing, 18,* 531–541.

Yin, C., Kellman, P. J., & Shipley, T. F. (1997). Surface completion complements boundary interpolation in the visual integration of partly occluded objects. *Perception, 26,* 1459–1479.

Yin, C., Kellman, P. J., & Shipley, T. F. (2000). Surface integration influences depth discrimination. *Vision Research, 40,* 1969–1978.

Zöllner, F. (1862). Uber eine neue Art anorthoskopischer Zerrbilder [On a new type of anorthoscopic distorted image]. *Annalen der Physic, 117,* 477–484.

Received November 14, 2005
Revision received June 5, 2006
Accepted June 6, 2006 ■