Motion Perception and Mid-Level Vision

Josh McDermott and Edward H. Adelson

Dept. of Brain and Cognitive Science, MIT

Note: the phenomena described in this chapter are very difficult to understand without viewing the moving stimuli. The reader is urged to view the demos when reading the chapter, at:
http://koffka.mit.edu/~kanile/master.html

Like many aspects of vision, motion perception begins with a massive array of local measurements performed by neurons in area V1. Each receptive field covers a small piece of the visual world, and as a result suffers from an ambiguity known as the aperture problem, illustrated in Figure 1. A moving contour, locally observed, is consistent with a family of possible motions (Wallach, 1935; Adelson and Movshon, 1982). This ambiguity is geometric in origin - motion parallel to the contour cannot be detected, as changes to this component of the motion do not change the images observed through the aperture. Only the component of the velocity orthogonal to the contour orientation can be measured, and as a result the actual velocity could be any of an infinite family of motions lying along a line in velocity space, as indicated in Figure 1. This ambiguity depends on the contour in question being straight, but smoothly curved contours are approximately straight when viewed locally, and the aperture problem is thus widespread. The upshot is that most local measurements made in the early stages of vision constrain object velocities but do not narrow them down to a single value; further analysis is necessary to yield the motions that we perceive.

It is possible to resolve the ambiguity of local measurements by combining information across space, as shown in Figure 2. The motion of 2-D features, such as the corner marked 2, is unambiguous, and can be combined with the contour information to provide a consistent velocity estimate. On the other hand, some 2-D features are the result of occlusion, such as the T-junction (marked 3) that occurs where the two squares of Figure 2(a) overlap. The motion of such features is spurious and does not correspond to the motion of any single physical object; in Figure 2 the two squares move left and right but the T-junction moves down. Distinguishing spurious features from real ones requires the use of form information, as the motion generated by such features does not in itself distinguish them.

An alternate way of extracting 2-D motion is to combine the ambiguous information from different contours of the same object, as shown in Figure 2(c). In velocity space, the constraints from contours 4 and 5 intersect in a single point (Adelson and Movshon, 1982), which represents the correct leftward motion of the diamond on the left. Similarly, contours 6 and 7, when combined, signal the rightward motion of the other diamond. However, it is important that the constraints that are combined originate from the same object. If the constraints from contours 5 and 6 are combined, for instance, they will lead to a spurious upward motion estimate. Thus it is critical to combine information across space, but it is also critical to do it

correctly. In the motion domain, however, it is not obvious that contours 4 and 5 belong together but that 5 and 6 do not. This again means that motion perception is inextricably bound up with form perception and perceptual organization.

In this chapter we review some of our work on the relationship between form, motion, occlusion, and grouping. We will consider these issues from two points of view. Sometimes it is most helpful to discuss them in terms of processes that act on local features, while in other cases explanations in terms of cost functions and optimization principles are most natural.

Motion Interpretation: Features

The most well-known consequence of the aperture problem is the barberpole illusion, which was studied by Wallach and many others since (Wallach, 1935; Wuerger, Rubin, & Shapley, 1996). A tilted grating moves behind a rectangular aperture, as shown in Figure 3. In the version shown in Figure 3(a), the aperture is the same color as the grating background, so there is no visible frame. Because of the aperture problem, the grating is consistent with various motions, including rightward or downward or diagonal motion. When the aperture is wider than it is high, as in this example, the grating will generally be seen as moving to the right. One way of explaining this effect is as follows. The grating is ambiguous, but the line terminators (the endpoints) are unambiguous 2-D features. There are more rightward terminators than downward ones, and so the rightward interpretation wins.

An interesting variant of the barberpole illusion is shown in Figure 3(b). Visible occluders are added on top and bottom. Now the same grating frequently appears to move downward. This could be explained as follows. The terminators along the top are now T-junctions, which could be the result of one contour being occluded by another. Since T-junctions can be created this way, their 2-D motions are often the spurious product of occlusion and therefore should be ignored. This means that the only reliable moving features in Figure 3(b) are the terminators along the left and right edges of the grating, and these are moving downward. Thus downward motion is seen.

The idea of junctions being detected, labeled, and possibly discounted, is well-established in the motion literature (e.g. Stoner et al., 1990; Lorenceau and Shiffrar, 1992; Trueswell and Hayhoe, 1993; Stoner and Albright, 1998; Rubin, 2000). Shimojo and Nakayama (1989) distinguished intrinsic

features, which really belong to an object, from extrinsic features, such as the T-junctions that are side-effects of occlusion. Nowlan and Sejnowski (1995), Liden and Pack (1999), and Grossberg and colleagues (2001) have all discussed models in which T-junctions are detected and discounted. Indeed, it can be said that this is the standard view of how many motion phenomena work. However, our research shows that the actual rules governing form influences on motion are subtle and complex, and, more surpisingly, that junction categories may have very little explanatory power. But before getting to the experimental results, let us consider representational issues that arise in these displays.

Motion Interpretation: Layers

The percepts associated with overlapping diamonds and moving barberpoles involve much more than just motion. The diamonds of Figure 2 are seen as two opaque objects, one occluding the other, and both occluding the background. Even if we cannot tell the exact depths of the various parts of this scene, we can tell their depth ordering and their opacity. A representation with motion, depth-ordering, and opacity, is known as a layered representation (Wang and Adelson, 1994), and it offers a basic tool for discussing the motion phenomena in this chapter.

Let us consider some layered decompositions associated with the barberpoles of Figure 4(a) and (b). Figure 4(a) shows a decomposition corresponding to rightward motion. First there is a background layer, then a moving strip in the next layer, and then a pair of occluders in the top layer. (The moving strip is shown as extended beyond the occluders for illustrative purposes). The main colors of all three layers are the same, so that only the black lines are visible in the actual display. This is referred to as a case of "invisible occluders," because the bounding contours that would normally demarcate the occluders cannot be seen. Figure 4(b) shows a decomposition corresponding to vertical motion. Now the invisible occluders are horizontal. Figure 4(c) shows a case where the occluders are visible because they are a different color.

How can we connect these decompositions to what is observed perceptually? For the basic barberpole of Figure 3(a), there are two interpretations that involve invisible occluders, shown in Figure 4(a) and (b). The one with the shorter invisible occluders is preferred by the visual system. We believe that this reflects a widespread principle of avoiding interpretations that involve illusory edges. Since an accidental match between an occluder and its

background (such as occurs when an illusory edge is perceived) is a rare event, the visual system prefers not to assume it has occurred. Given the choice of a longer or shorter stretch of invisible occluder, it will prefer the shorter stretch, leading it to choose the decomposition of Figure 4(a). In the case of Figure 3(b) and the dark rectangles, however, there is no need to posit invisible occluders, since visible rectangles are clearly present, with their boundaries indicating possible points of occlusion. There is little cost to placing the rectangles in a separate layer, leading to the preferred decomposition of Figure 4(c), and therefore leading to the downward motion percept.

Complementary Approaches to Understanding Perception

Observe that we have just walked through two very different explanations of the same barberpole phenomena. The first time through, we spoke of identifying, tracking, and discounting local features such as terminators and T-junctions. The second time through, we spoke of layered decompositions and accidental matches, but made no mention of terminators or T-junctions. The first, feature-based explanation could be used in developing a bottom-up model, in which various local image operations are combined in successive steps to build up a motion percept. The second, layer-based explanation could be used in a top-down model that sought an optimal solution to a stated problem, such as finding the most probable interpretation of motion given some assumptions about the statistics of the world. We have found both of these approaches to be useful in our thinking about motion phenomena.

The bottom-up approach is more popular in motion modeling. This is understandable, since modelers are often trying to determine the stages of neural processing that underlie the motion percepts, and these stages are usually conceived of as primarily feed-forward. Optimization is also commonly considered to be difficult to implement, because it is often necessary to search through a large space to find the optimum. However, it is worth noting that the optimization approach has many advantages. The idea of minimizing a cost function has a long history in perception. Helmholtz advocated the idea of finding the most likely interpretation of the sensory data, and others (e.g. Hochberg, 1953; Attneave, 1954; Leeuwenberg, 1969) have proposed that humans seek to minimize the complexity of image descriptions. In motion perception, Restle (1979), Hildreth (1984), Grzyawacz and Yuille (1991), and others have had success with various minimization rules. Recently, Weiss et al (2002) have shown that many

phenomena related to the aperture problem can be understood in terms of a Bayesian framework that finds the most likely single motion consistent with image data. Their results are noteworthy because they do not depend upon the usual explicit mechanisms such as feature-tracking, intersection of constraints, or vector averaging. Rather, they apply a unified principle that automatically captures the uncertainty associated with the aperture problem and with noise. In another paper, Weiss and Adelson (2000) show that similar minimization principles, when coupled with a layered decomposition, can account for a wide range of phenemona associated with rotating and distorting ellipses.

In this chapter we will discuss a range of phenomena involving moving figures and occlusion. We cannot offer a single minimization principle to cover all phenomena, and some of the phenomena are indeed suggestive of feature-based processes. However, we feel that cost functions provide a promising approach, and we hope that it will be possible to blend the feature-based descriptions with minimization principles in the future. In the present discussion will use both ways of thinking, as seems appropriate.

The Cross Stimulus

Our explorations of junction-based rules began with the cross stimulus, shown in Figure 5. The cross, derived from Anstis' chopsticks illusion (Anstis, 1990), consists of two bars that move sinusoidally, 90 degrees out of phase with each other (McDermott and Adelson, 2003). When the bars are combined to form a cross, their intersection point traces out a circle, and if the cross is viewed within an occluding frame, as in Figure 5(c), the cross bars appear to cohere and move together in a circle. Without the frame in place, as in Figure 5(d), the bars appear to move separately, in the linear direction orthogonal to their orientation, even though the image motion is unchanged.

What accounts for the effect of the frame? As discussed earlier for the barberpole stimulus, the usual explanation involves tagging and discounting certain kinds of junctions. The bar endpoints provide unambiguous two-dimensional motion signals, and without the frame, these are believed to determine the motion percept. The endpoints move linearly, and each bar follows along. When the frame is present, however, T-junctions are formed at the bar endpoints. These junctions provide a cue that the endpoint motions are the spurious result of occlusion. Accordingly, standard models discount motions that occur at T-junctions (Nowlan and Sejnowski, 1995; Liden and

Pack, 1999; Grossberg et al., 2001). With the T-junctions in the cross stimulus discounted, the circular motion of the bar intersection determines the motion percept, as all the local motions in the stimulus apart from those of the endpoints are consistent with such a circular motion.

Elements of this story may be on the right track, but the reality is more complex, and more interesting, as we learned when we took a closer look at the influence of junctions. We were surprised to find that the feature-based descriptions are of limited value, and in particular that the notion of tagging and discounting T-junctions can explain surprisingly little. While certain other features may be important, as we will describe later in the chapter, we have also found that the optimization approach has the potential to explain quite a bit, although it does not offer a process-based explanation of the percepts.

Junctions and Cost Functions

To test for the presence of junction-dependent form constraints, we examined the effects of changing T-junctions to L-junctions in the cross stimulus, by matching the luminance of the occluders with that of the moving bars. If the T-junctions that are formed where the bars and occluders overlap play any role in the interpretation of motion in the display, one would expect a change to the junctions to alter perceived motion. As shown in Figure 6, we either held the bar contrast or the occluder contrast fixed, and swept the other through the point of accidental match (the point where the bars and occluders have the same luminance), observing the effect on coherence. Given that L-junctions are thought to be weaker cues to occlusion than T-junctions, we expected to see a dip in coherence when the bars and occluders matched in luminance.

In the first experiment the bar contrast was fixed and 9 different occluder contrasts were tested (Figure 6a), running through the point of accidental match. In the second experiment the occluder contrast was fixed and 8 different bar contrasts were tested (Figure 6b), again running through the point of accidental match. Observers were shown short clips of each stimulus, and were asked to judge whether it was coherent, incoherent, or somewhere in between (for other details of the methods, see McDermott et al., 2001). These ratings were converted into a coherence index plotted in Figure 7.

As shown in Figs. 7a and 7b, the dominant effect was an overall shift in coherence with contrast: coherence increased with occluder contrast and decreased with bar contrast. Shapley and colleagues (1995) obtained similar results with the barberpole stimulus; these contrast effects appear to be a general property of occlusion/motion interactions. We believe the effects are in part due to the role that contrast plays as a depth cue (O'Shea et al., 1994; Stoner and Albright, 1998; Rohaly and Wilson, 1999), but we will not discuss it further in this chapter – we simply accept that the contrast effect is present.

The important point for our purposes is that there was no obvious drop in coherence at the point where L-junctions were generated at the bar endpoints. The curves passed smoothly through the match point, and the category of the junction generated at the bar endpoints had little to no effect on the coherence of the cross.

We also tested the role of the junctions at the center of the cross rather than at the bar endpoints. By changing the luminance of one of the bars we could change the L-junctions to T-junctions, as shown in Fig. 8. In this situation one would expect the L-junctions at the match point to produce an *increase* in coherence relative to stimuli with T-junctions at the center, since the L-junctions increase the likelihood that the two bars are a single, coherently moving object. We varied the luminance of one of the two moving bars while holding the luminance of everything else fixed, looking for an effect at the match point.

Curiously, in this case the match point did produce an obvious effect: coherence was highest where the bars matched in luminance, producing a "blip" in the graph of Fig. 8. We again observed the expected effect of bar contrast; coherence decreased with increasing bar contrast (although here the contrast varied for only one of the bars). But superimposed on this decreasing curve was a pronounced effect of the match point, consistent with what one would expect if junctions were important.

This effect of junction categories at the center intersection seems hard to reconcile with the previous experiment, in which the category of the junctions at the bar endpoints apparently had little to no effect on the extent to which the endpoint motions were discounted. What could explain this pattern of results?

One possibility is just that the junctions we varied at the bar endpoints were too small for the relevant visual processes to resolve. Although these junctions were clearly visible in our stimuli (it was easy to distinguish T's from L's), it is conceivable that the mechanisms that analyze them for motion interpretation operate at coarse resolution. To test this idea, we made the cross bars thicker, effectively enlarging the pair of junctions formed where the cross bars meet the occluders.

The problem with simply thickening the bars of the cross is that the intersection of the crossbars is also altered. When the crossbars are the same luminance, as in our original stimulus, the length of the contours that have to be completed when the bars are incoherent increases as the bar width is increased. Presumably because of this, the bars are much less likely to appear fully incoherent when they are thick. To avoid ceiling effects, we used a version of the stimulus in which one of the bars was lower or higher in luminance than the other, which was fixed at the match point luminance (see Fig. 9a). As we saw in the previous experiment, this results in somewhat lower overall coherence, but the stimulus otherwise behaves like the original cross. As a result of the luminance difference between the bars, however, the width of the bars can be changed without obviously changing basic aspects of the stimulus percept.

We varied the contrast of one pair of the occluders in this stimulus for two different bar thicknesses, again looking for an effect at the point where the occluders matched the bar in luminance and generated L-junctions instead of T-junctions. In the thin bar conditions, the bars were the same thickness as before; in the thick bar conditions the bars were 3.5 times as wide.

For the thin bars, there was again no apparent effect of junction category, as shown in Fig. 9b. With thick bars, there was a slight drop in coherence at the match point, but it was quite small. The dominant effect was that of bar contrast, as before. Even when the junctions were separated by large distances and were thus easy to resolve, their category was of little consequence.

Illusory Edges

To understand this apparently puzzling set of results, we must consider how different types of junctions are associated with occlusion in the first place. As shown in Fig. 10a, T-junctions are produced whenever an occluder’s color is different from that of the surface it occludes. We can say that

occlusion *generically* produces T-junctions because almost all combinations of surface colors produce the T. In contrast, an L-junction can only result from occlusion when the two surfaces involved accidentally match in color, as in Fig. 10c. Because an accidental match is involved, this interpretation involves postulating an "illusory" edge – an edge in the world (part of the occluding contour) where there is none in the image.

On grounds of parsimony alone, one would expect the visual system to minimize the number of surface edges in its perceptual interpretation that do not project to intensity edges in the image. If this were the case, then the visual system ought to be biased to interpret L-junctions as corners (Fig. 10b) rather than occlusion points, and T-junctions, which do not require postulating such edges, would clearly be the stronger occlusion cue.

Since the coherence of the cross seems to depend on evidence for occlusion, we had expected lower coherence at the point of accidental match, where L-junctions are generated at the bar endpoints. Upon inspection, however, both the coherent and incoherent percepts of the cross necessitate a discontinuity between the occluders and bars. As shown in Fig. 11a, this is because the occluders are static and the bars are moving, so regardless of whether the bars cohere and move under the occluders, there must be a surface discontinuity where they meet. When the bars are the same luminance at the match point, this discontinuity takes the form of an illusory edge. If the visual system is attempting to minimize such illusory edges, the coherent interpretation of the cross should in fact be no less likely at the match point despite the presence of L-junctions.

At the bar intersection, in contrast, the situation is different. When coherent, the bars are stuck together as one surface and there is no discontinuity at their intersection. Thus illusory edge minimization makes a different prediction, again correct, for the junctions at the bar intersection - coherence should be more likely when the bars match in luminance and generate L-junctions than when they differ in luminance and produce T-junctions. What appeared to be incompatible results actually provide evidence for a single, sensible computation based on the notion of optimization discussed earlier.

To put this notion to the test, we altered the cross stimulus once more. Our aim was to take the stimulus with matching bar and occluder luminances, shown in Fig. 11a, and selectively remove the endpoint discontinuity in the incoherent motion interpretation, to see if this might then produce a match

point effect at the bar endpoints. In the stimulus of Fig. 11b, the white occluders have been extended to cover the horizontal occluders (whose luminance is varied in the experiment). As a result, the horizontal occluders need not be stationary, and can be seen to move with the vertical bar as a single I-shape. Thus in addition to the two standard cross percepts, this new stimulus has a third perceptual interpretation, depicted in Fig. 11b (far right), in which the I-shape is seen to move back and forth without any discontinuity between the bar and the occluders. In our experience this percept is difficult to imagine from the static figures, but is readily experienced when viewing our online demos. The incoherent interpretation thus does not necessitate an illusory edge at the match point, because the bar and its occluders can be seen as part of the same surface. When coherent, in contrast, the bars still must move under the occluders, generating the illusory discontinuity. Illusory edge constraints might therefore predict a drop in coherence at the match point, since there would be reason to prefer the incoherent interpretation. We therefore conducted another match point experiment with both configurations of Fig. 11, varying the luminance of one pair of the occluders and looking for an effect where they matched the bar luminance.

As shown in Fig. 12, the new configuration indeed resulted in a pronounced effect of the match point; there was a large decrease in coherence, comparable to the increase in coherence observed in Figure 8, for the match at the bar intersection. We again observed a very small effect of the match point in our original configuration, but it was dwarfed by the big effect in the new configuration. This result is just that predicted by a computation minimizing the number of illusory edges in the perceptual interpretation. The visual system seems to try to avoid postulating surface discontinuities in the absence of visible edges.

The upshot of this series of experiments is that we have no evidence that there are form constraints on motion interpretation that are specifically tied to junctions. Instead, the behavior of the visual system seems well-characterized by an optimization-based form computation that tries to minimize the presence of illusory edges in the perceptual representation. This explanation is much the same as that suggested earlier for the barberpole illusion. As before, the cost function is easy to describe qualitatively, but its implementation is probably quite complex. It is not obvious how one could account for these effects with processes acting on local features. However, our description says nothing about what is involved

mechanistically, and it is possible that junctions play a role at this level. But there is no simple account of the results that is based on junction categories, whereas there is a simple account based on the minimization of illusory edges.

Amodal Completion

Illusory edges are not the only things that figure into the cost function for motion. Consider the square stimulus of Figure 13, first introduced by Lorenceau and Shiffrar (1992). The stimulus is made of moving bars, just as before, except this time there are two pairs of bars. Each pair oscillates sinusoidally, 90 degrees out of phase with the other pair. When viewed alone, as in Figure 13(a), the pairs of bars appear to move independently, translating horizontally and vertically. However, when static occluders are added to the display, as in Figure 13(b), the percept is quite different – the two pairs of bars appear to move together in a circle, as a single solid square. As before, we can ask what is driving the percept, and ask whether it is fruitful to think of the computations involved as minimizing some cost function.

In the case of the square, observers commonly report that when the four bars of the diamond appear to move coherently in a circle, the diamond corners perceptually complete behind the occluders. We wondered if amodal completion was merely an incidental feature of the percept or whether it might play some more fundamental role in determining perceived motion. To address this issue we manipulated the shape of the occluders, in a series of experiments more fully described elsewhere (McDermott et al., 2001). We first compared the coherence obtained with full occluders, shown in Figure 14a, to that produced by the L-shaped occluders of Figure 14b. If the coherence of the fully occluded diamond is closely related to the amodal completion of the diamond contours, one might expect coherence to be lower with the L-shaped occluders, as they do not provide room for the contours to complete. The thin gray lines in the background help to ensure that the entire background is seen as a single surface, leaving no room for the diamond contours to complete.

Even though the L-shaped occluders have the same occluding contour as the full occluders, and produce similar T-junctions, they produce much lower levels of coherence. The stimulus of Figure 14b was almost always incoherent, almost as often as when the bars were presented alone on the

background (Figure 14c). We were able to restore coherence by closing the L-shapes as shown in Figure 14d, so that the Ls were seen as the borders of extended surfaces which provide room for the diamond contours to complete. The results are consistent with an important role for amodal completion in motion interpretation, and again underscore the conclusion that there is much more to the form computations than mere junction detection.

The sophistication of the form constraints is further shown with two manipulations of the background lines. As shown in Figure 14e, coherence is reduced when the background lines are extended through the occluder outlines, presumably because they are inconsistent with the presence of extended surfaces which could support completion. Moreover, removing the background lines from the L-shaped occluder stimulus, as shown in Figure 14f, increases coherence, presumably because without the lines the Ls are more likely to form the borders of extended surfaces. Gradually closing the L-shapes, as shown in Figure 15, further increases coherence, again consistent with the increased likelihood of an extended surface. Motion interpretation again seems to be privy to rather subtle aspects of spatial form, and junctions by themselves seem to have little predictive value.

To further test the importance of completion, we manipulated the position of the diamond contours in ways that affected their ability to amodally complete. As with many of the other effects described in this chapter, the manipulations much easier to understand if one views the moving demos, for which we refer the reader to our demo web page. Consider the contours of Figure 16, in which the line segments are shown through apertures. In Figure 16a, the line segments can be connected with a smooth contour to form a square. Kellman and Shipley have referred to such contours as ‘relatable’ (1991). Relatability depends on the geometric relationships between the contours. In Figure 16b, the horizontal segments have been moved inwards so that a simple completion with the vertical segments is impossible; these contours are nonrelatable. When the line segments were set in motion and shown to observers, we found dramatic differences in how the motion was interpreted; while the relatable contours almost always cohered, the nonrelatable ones virtually never did. Note that proximity biases on motion integration (e.g. Nakayama and Silverman, 1988) would, if anything, predict that the nonrelatable stimulus should cohere more, as the segments are somewhat closer to each other than in the relatable stimulus. Evidently any proximity biases are swamped by the effect of relatability. One might

nonetheless object that it is simply impossible to see the nonrelatable configuration as a single object in coherent motion. This is not the case. As shown in Figure 16(c), we added dots to the nonrelatable line segments and moved them with the same circular trajectory seen when the line segments cohered in Figure 16(a). With the addition of the dots, the line segments appeared to cohere, moving together as a single object. Apparently the moving dots captured the motion of each line segment, and the segments were then grouped together in accord with the Gestalt principle of common fate. Nonrelatability thus does not prevent coherence per se but rather the specific process of motion integration across contours. We suggest this is another example of a completion constraint – local motions seem to be preferentially integrated when the contours that give rise to them can amodally complete.

We can think of these completion-related effects as the product of a cost function as well. Motion interpretations appear to be penalized when they involve integrating the motion of contours that are separated in space but which do not amodally complete.

However, a third example of the role of completion-related processes in motion interpretation is less conducive to such an explanation. Inspired by an experiment done by Shimojo, Silverman, and Nakayama (1989), we compared the motion seen in the single barberpole to that seen when identical barberpoles are added to the top and bottom of the original one. The top and bottom barberpoles tend to amodally complete with the middle one, and we thought this might increase the tendency of the visual system to discount the horizontal line endings, as amodal completion only occurs between occluded contours. Indeed, as shown in Figure 17, we find the triple barberpole to be roughly twice as likely to be seen moving vertically than is the single barberpole, suggesting that the presence of relatable contours in the adjacent gratings causes the occluded line endings to be discounted to a greater extent. Note that the relative proportion of different motion signals (horizontal line endings, vertical line endings, and line segments) is constant across the two stimuli, as the top and bottom barberpoles are identical to the middle. Thus it is not clear how to account for the result other than by supposing that the horizontal motion signals are discounted to a greater extent in the triple configuration. Completion-related constraints again seem to be exerting their influence, but in this case the most intuitive explanation is process-based, related to the weight given to particular motion signals as a function of the stimulus configuration.

Border Ownership

As a further test of the importance of nonlocal cues to occlusion, we devised stimuli such as those in Figure 18 (McDermott et al., 2001). The stimuli of Figure 18a and 18b have identical junctions at the bar endpoints, but differ globally in the extent to which the bars appear to be occluded. As shown in Figure 18c, observers reported the second stimulus to be far less coherent than the first, consistent with the weaker impression of occlusion that it conveys. Again, the T-junctions alone do a poor job of predicting motion interpretation, since the same T-junctions are present in both cases. What is the nature of the process or computation that is responsible for this effect?

The stimuli of Figure 18 differ in a number of ways, but we wondered whether the geometry of the occluding contour might be important. Note that in the stimulus of Figure 18a, the occluding contour abutting each moving bar is convex, whereas in Figure 18b, it is concave. Contour convexity is a well-known cue to border ownership (Stevens and Brooks, 1988; Pao et al., 1999), so it seemed possible that this might have something to do with the different motion seen in the two displays. To probe the role of convexity we conducted some experiments with outline stimuli, shown in Figure 19, which allow for some interesting manipulations (McDermott and Adelson, 2004). Figure 19a shows the diamond with outline occluders; this stimulus cohered most of the time as one would expect. In the stimulus of Figure 19b, we removed most of the occluding contour, leaving just the T-junctions at the bar endpoints. This stimulus generated intermediate levels of coherence. In the stimuli of Figure 19c and 19d, we added short line segments to the T-junctions to produce local convexities and concavities, respectively. The convexities increased the level of coherence relative to the T-junctions alone, while the concavities decreased it. Note that no occluders are visible in these stimuli; there are just isolated pieces of contour. Nonetheless, manipulating the local concavity produced a sizeable effect.

Can convexity predict perceived coherence in other stimuli as well? We compared the coherence obtained for the occluded diamond with that for an identical square viewed through apertures with the same occluding contours as the occluders, as shown in Figure 20. The apertures produced substantially lower levels of coherence than do the occluders, consistent with the notion that the degree of coherence is determined in part by the local

convexity, and perhaps the strength of occlusion, which may derive from the convexity.

We also wondered whether additional T-junctions along the occluding contour might influence border ownership and hence motion interpretation. The stimuli of Figure 21 were designed to address this issue. The round apertures of Figure 21a alone produced moderate levels of coherence, as did the oddly shaped occluders of Figure 21b. But when combined in the stimulus of Figure 21c, coherence was substantially lower than in either stimulus alone, consistent with the weak percept of occlusion that most observers report. Here the weak coherence cannot be attributed merely to the shape of the occluding contour. Something happens specifically when the two contours are combined. One appealing explanation is that the T-junctions of Figure 21(c) modulate the strength of border ownership, which in turn influences motion interpretation. The control of Figure 21d is further consistent with this notion.

These last examples of the effects of border ownership cues are most suggestive of processes acting on sets of local features. By themselves the T-junctions at the bar endpoints seem to predict very little, but if we consider the junctions along with the geometry of the occluding contour in a region surrounding the junction, we can account for much more. The results suggest that local cues such as contour convexity and junctions are combined to yield an estimate of the likelihood of occlusion, which then may be used to determine the motion interpretation. Note that this explanation has a very different flavor from that which we offered of the cross experiments, in which we proposed a cost function which could be applied to each of the candidate perceptual interpretations. The cost function didn't involve local image features, being a function only of the layered representation derived from the image data. Here, in contrast, it is hard to explain the phenomena without direct reference to particular critical image features. It remains a challenge for future research to show if and how these phenomena related to border ownership may be described as minimizing some cost function on perceptual interpretations.

Regardless of the kind of explanation adopted for the various phenomena in this chapter, certain general conclusions emerge. First, the form influences on motion serve to solve fundamental computational problems in motion interpretation introduced by occlusion. Feature motions are discounted when they are likely to be the spurious product of occlusion, and distant motions

are integrated only if they are likely to be due to the same object. Second, the popular view that the form constraints on motion can be accounted for with isolated processes operating on junctions has little merit in the phenomena we have examined. Motion interpretation is influenced by a variety of nonlocal form computations, and the effect of these computations is quite powerful. They can effectively switch between different motion interpretations depending on the stimulus configuration, even when the junctions are unchanged. The complexity of these interactions would appear to implicate substantial cross-talk between the motion and form pathways, which may be another fruitful avenue for future investigation.

Summary

Motion, form, occlusion, and perceptual organization are intimately related, and ambiguous moving stimuli provide powerful tools to investigate their relationship. We have described phenomena involving moving crosses and squares that suggest a number of subtle and sophisticated links between motion and form. The simplest story one could tell about motion and form interactions, involving local processes based on junctions, bears surprisingly little resemblance to the various form processes that we find to be at work. Two general sorts of explanations are suggested by our phenomena, process-based and optimization-based. In some cases the phenomena are best explained with reference to processes that act on local features, such as the convexity of the occluding contour. In other cases the simplest explanation is in terms of a cost function that is minimized, for instance one which penalizes illusory edges. In all cases isolated junctions have little explanatory power, and we must appeal to more complex and interesting form computations to account for the ease and accuracy with which we perceive motion in real-world scenes.

References

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

Anstis S, 1990. Imperceptible intersections: The chopstick illusion. In *AI and the Eye*, A Blake and T Troscianko, eds. New York: John Wiley.

Attneave, F, 1954. Some informational aspects of visual perception. *Psychological Review* 61, 183-193.

Grossberg, S., Mingolla, E., and Viswanathan, L. 2001. Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research* 41:2521-53.

Grzywacz, N.M. and Yuille, A.L. 1991. Theories for the visual perception of local velocity and coherent motion. In *Computational models of visual processing*, J. Landy and J. Movshon, eds. Cambridge, Massachusetts: MIT Press.

Hildreth, E.C. 1984. *The Measurement of Visual Motion*. Cambridge, Massachusetts: MIT Press.

Hochberg, J. & McAlister, E. 1953. A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, 46:362-364.

Kellman P, Shipley T, 1991. A theory of visual interpolation in object perception. *Cognitive Psychology*, 23:141-221.

Leeuwenberg, E. 1969. Quantitative specification of information in sequential patterns. *Psychological Review*, 76:216-220.

Liden L, Pack C, 1999. The role of terminators and occlusion cues in motion integration and segmentation: A neural network model. *Vision Research*, 39:3301-3320.

Lorenceau J, Shiffrar M, 1992. The influence of terminators on motion integration across space. *Vision Research*, 32:263-273.

Lorenceau J, Zago L, 1999. Cooperative and competitive spatial interactions in motion integration. *Visual Neuroscience*, 16:755-770.

McDermott, J. and Adelson, E.H, 2003. Junctions and cost functions in motion interpretation. To appear in the *Journal of Vision*.

McDermott, J. and Adelson, E.H, 2004. The geometry of the occluding contour and its effect on motion interpretation. Submitted.

McDermott, J., Weiss, Y., and Adelson, E.H, 2001. Beyond junctions: Nonlocal form constraints on motion interpretation. *Perception*, 30: 905-923.

Nakayama K, Silverman G H. 1988. The aperture problem II: Spatial integration of information along contours'. *Vision Research*, 28 747-753.

Nowlan S, Sejnowski T, 1995. A selection model for motion processing in area MT of primates. *Journal of Neuroscience*, 15:1195- 1214.

O'Shea, R. P., Blackburn, S. G., & Ono, H. 1994. Contrast as a depth cue. *Vision Research*, 34:1595-1604.

Pao, H., Geiger, D. and Rubin, N. 1999. Measuring convexity for Figure/Ground separation. *Proc. 7th IEEE Intl. Conf. Comp. Vision*, 948-955.

Restle, F. 1979. Coding theory and the perception of motion configurations. *Psychological Review*, 86:1-24.

Rohaly, A. M. & Wilson, H.R. 1999. The effects of contrast on perceived depth and depth discrimination. *Vision Research*, 39:9-18.

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

Shapley, R., Gordon, J., Truong, C., & Rubin, N. 1995. Effect of contrast on perceived direction of motion in the barberpole illusion. *Investigative Ophthalmology & Visual Science* 36:1845.

Shiffrar M, Li X, Lorenceau J, 1995. Motion integration across differing image features. *Vision Research*, 35:2137- 2146.

Shiffrar M, Lorenceau J, 1996. Increased motion linking across edges with decreased luminance contrast, edge width and duration. *Vision Research*, 36:2061- 2067

Shimojo S, Silverman G H, Nakayama K, 1989. Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29:619- 626.

Shipley T F, Kellman P J, 1992. Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception and Psychophysics*, 52:97- 106.

Stevens, K.A. & Brookes, A. 1988 The convex cusp as a determiner of figure-ground. *Perception*, 17:35-42.

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

Stoner, G. R. & Albright, T. D. (1998). Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from occlusion cue. *Vision Research*, 38:387-401.

Wallach H, 1935. ë ber visuell wahrgenommene Bewegungrichtung *Psychologische Forschung*, 20:325- 380 [see also Wuerger et al (1996)].

Weiss Y, Adelson E H. 2000. Adventures with gelatinous ellipses: constraints on models of human motion analysis. *Perception*, 29:543- 566.

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

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

Figures

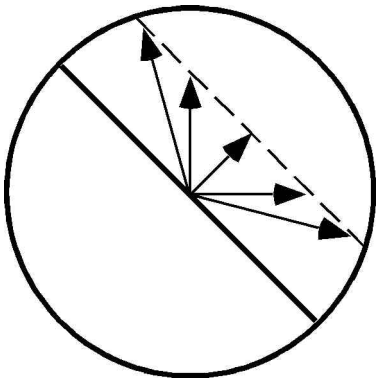

Figure 1. The aperture problem. Each of the motions (designated with arrows) on the depicted line in velocity space is physically consistent with the edge motion, as only the orthogonal component of its velocity can be detected.

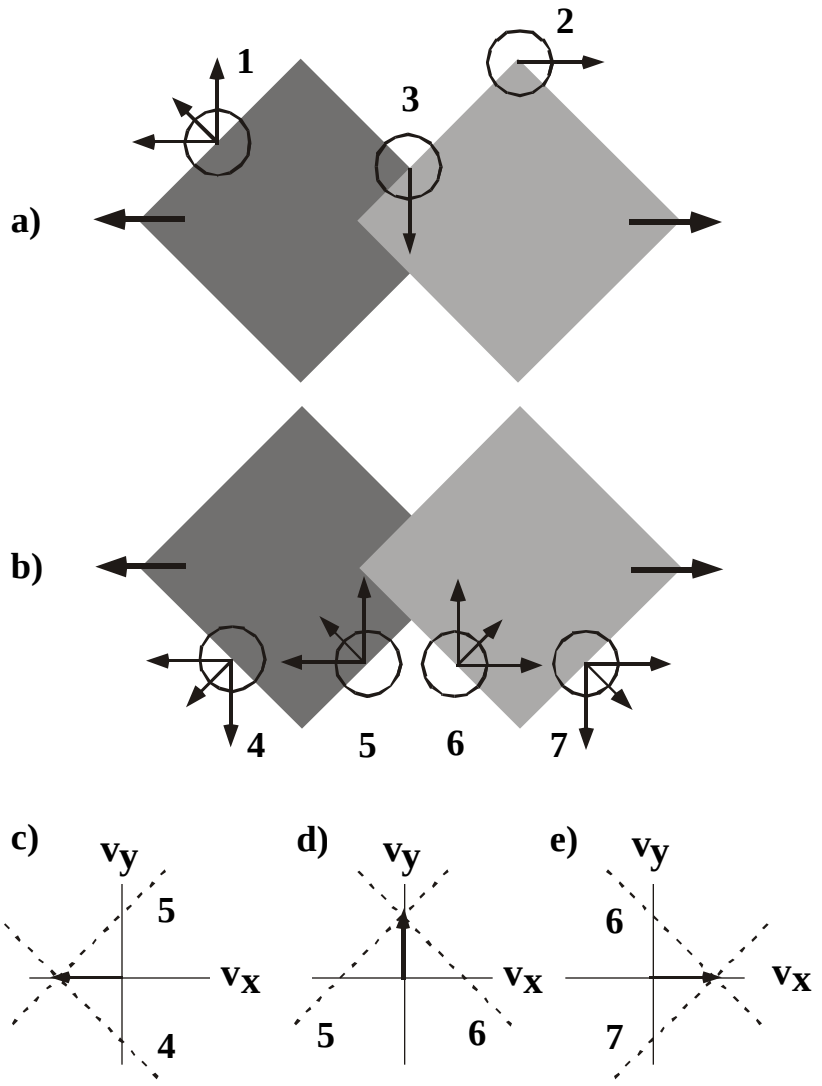

Figure 2. Example illustrating two problems that occur when integrating motion across space. In (a) and (b), two squares translate horizontally. The edge motions (e.g. 1) are ambiguous, while the corner motions (e.g. 2) are unambiguous. The T-junction motions (e.g. 3) are also unambiguous, but their motion is spurious and must somehow be discounted. Integration also poses a problem: (c), (d), and (e) show the velocity-space representations of the motion constraints provided by edges 4 and 5, 5 and 6, and 6 and 7, respectively. If the motion constraints from two edges of the same object are combined via intersection of constraints, as in (c) and (e), the correct horizontal motions result. If, however, motion constraints from edges of different objects are combined, as in (d), an erroneous upward motion is obtained. Note that the three pairs of local motions are separated by approximately the same distance, and are not distinguished on the basis of their motion. Form information is apparently needed to determine which measurements originate from the same object.

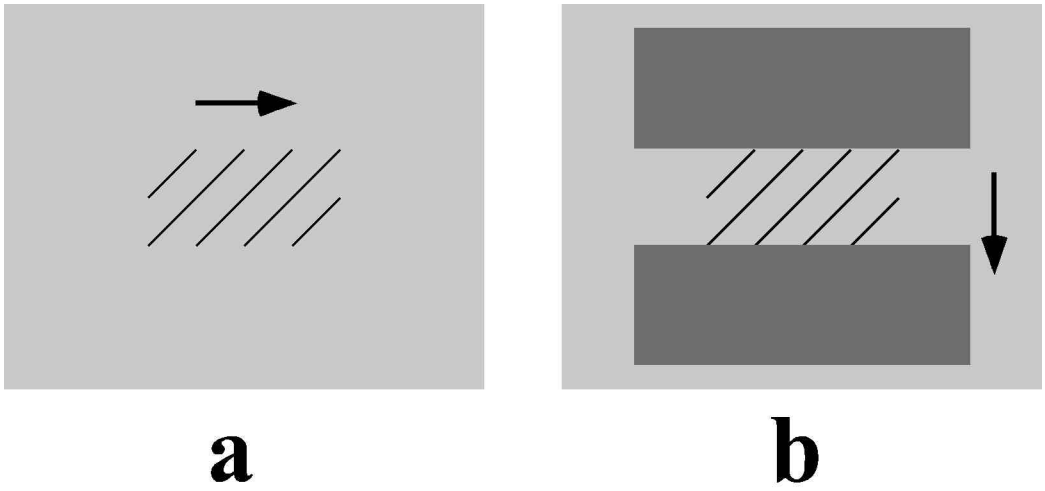

Figure 3. The barberpole illusion. (a) A gratings drifts behind an invisible rectangular aperture, and appears to move horizontally, along the long axis of the aperture. (b) When occluders are added at the top and bottom of the barberpole, vertical motion is often seen, even though the image motion is unchanged. Arrows denote perceived direction of motion.

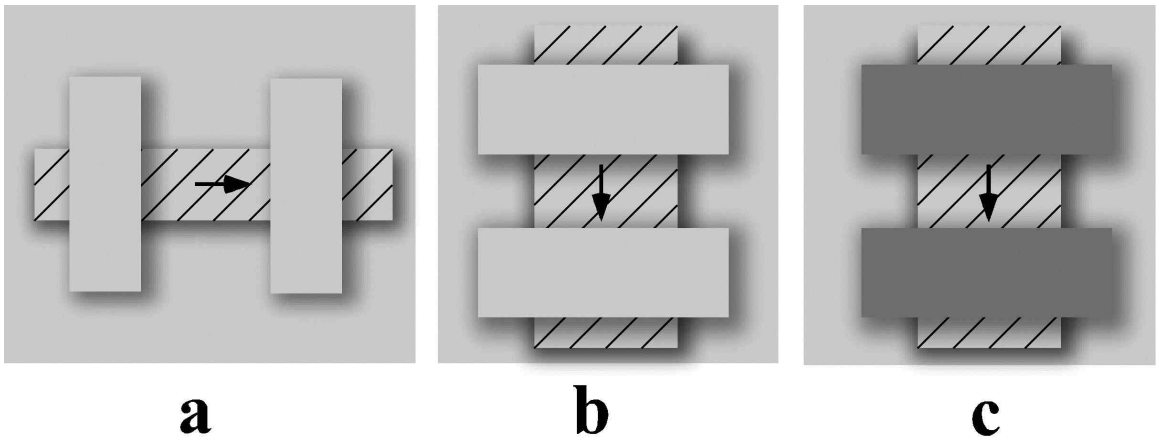

Figure 4. Layered interpretations of the barberpole stimuli of Figure 3.

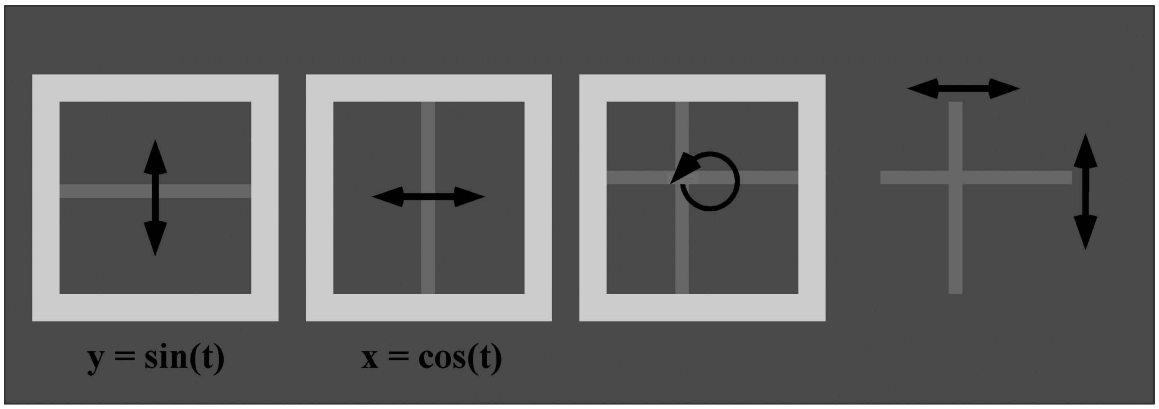

Figure 5. The cross stimulus. Two bars translate sinusoidally, 90 degrees out of phase, such that their point of intersection executes a circular trajectory. When viewed within an occluding aperture, the bars perceptually cohere and appear to move together with this circular trajectory. When the occluding aperture is removed, coherence breaks down and the bars are seen to move separately, even though the image motion is unchanged.

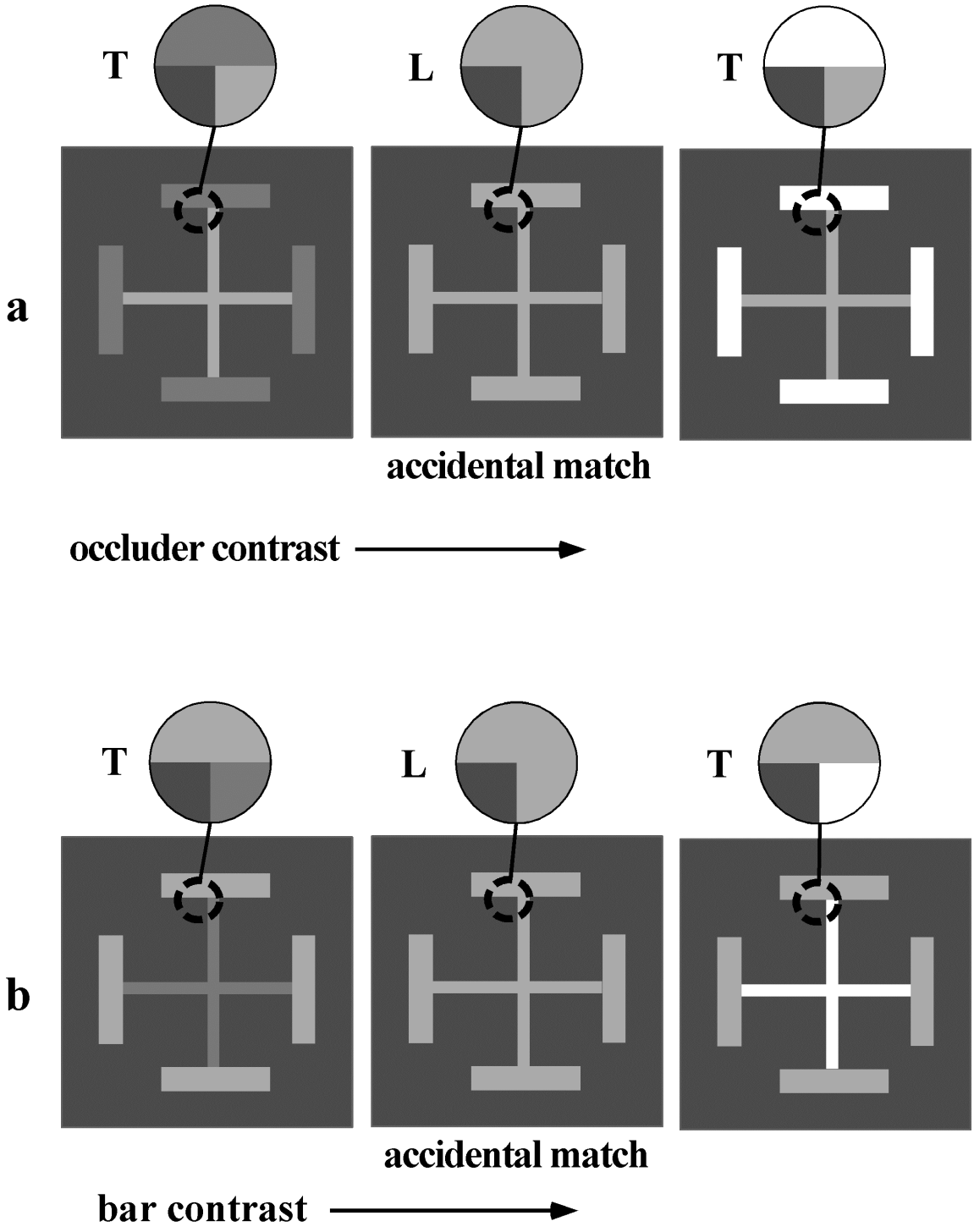

Figure 6. The effect of junction category was tested by varying bar (a) and occluder (b) contrast and examining the effect of a match in contrast between bars and occluders.

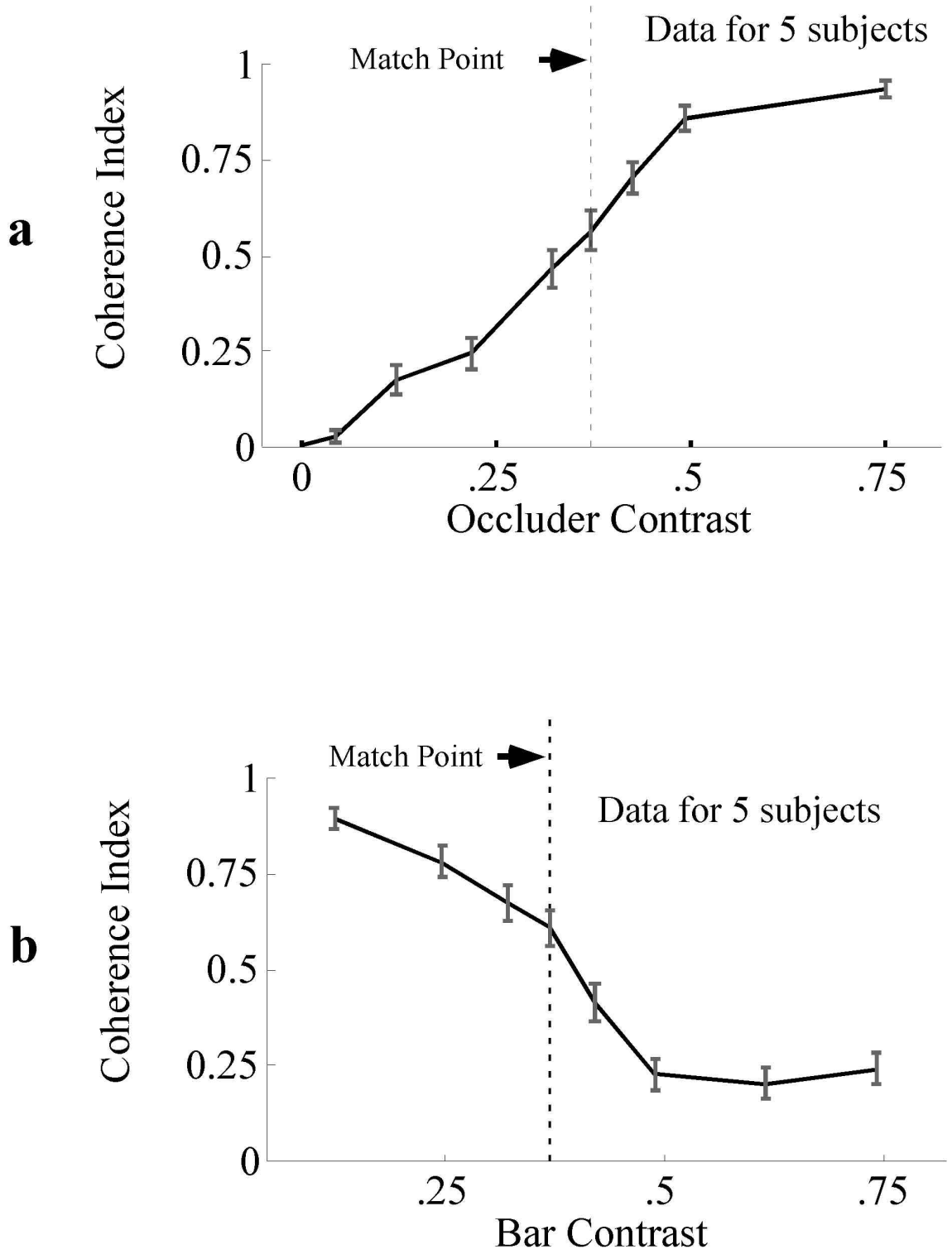

Figure 7. Results of the experiment schematized in Figure 6. Error bars in this and all other graphs denote standard errors.

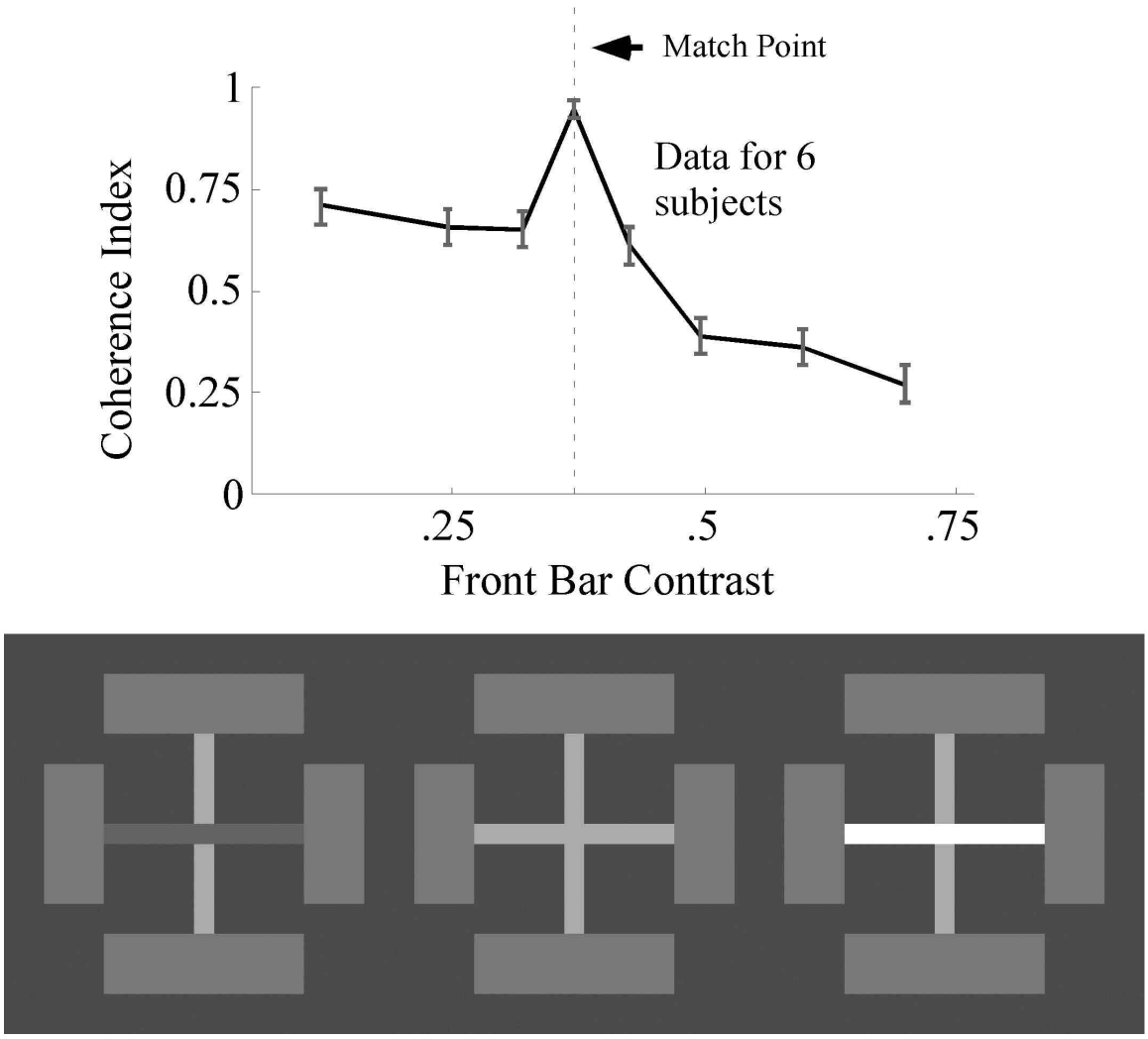

Figure 8. A match between the luminance of the two bars results in a pronounced peak in coherence.

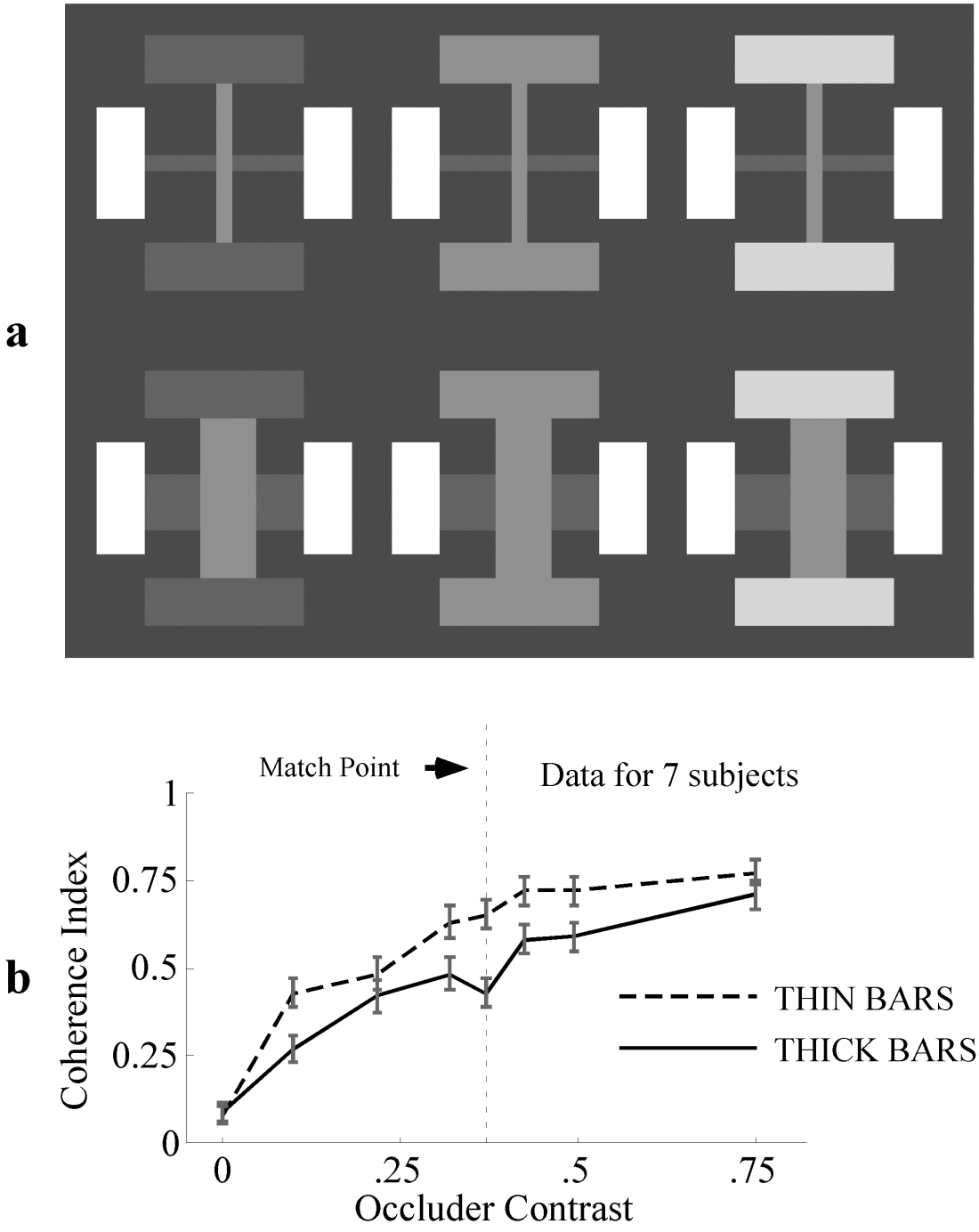

Figure 9. To control for resolution issues, we repeated the first experiment with bars that were 3.5 times as thick. Changing the junctions at the bar endpoints again has little to no effect.

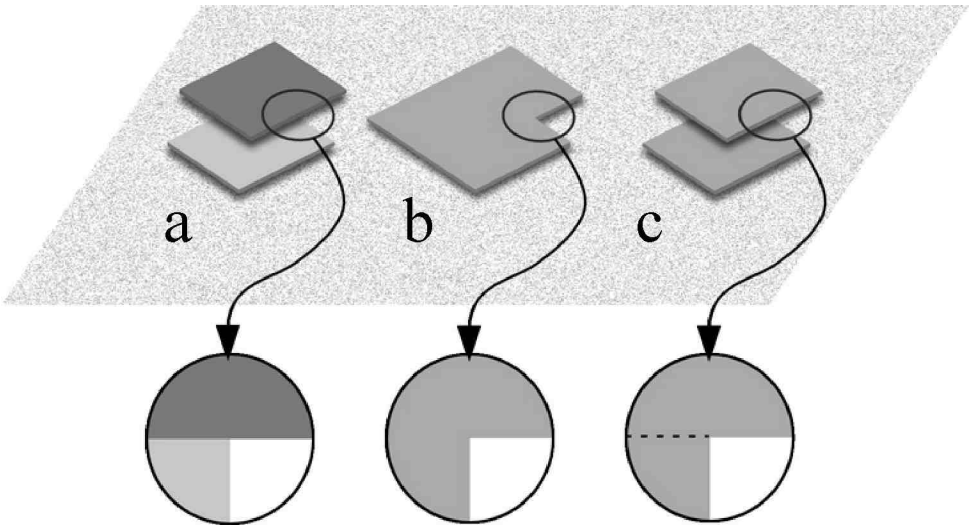

Figure 10. T-junctions are generically associated with occlusion, L-junctions are not. Interpreting an L-junction in terms of occlusion requires postulating an illusory edge – a surface discontinuity that does not correspond to a luminance edge in the image.

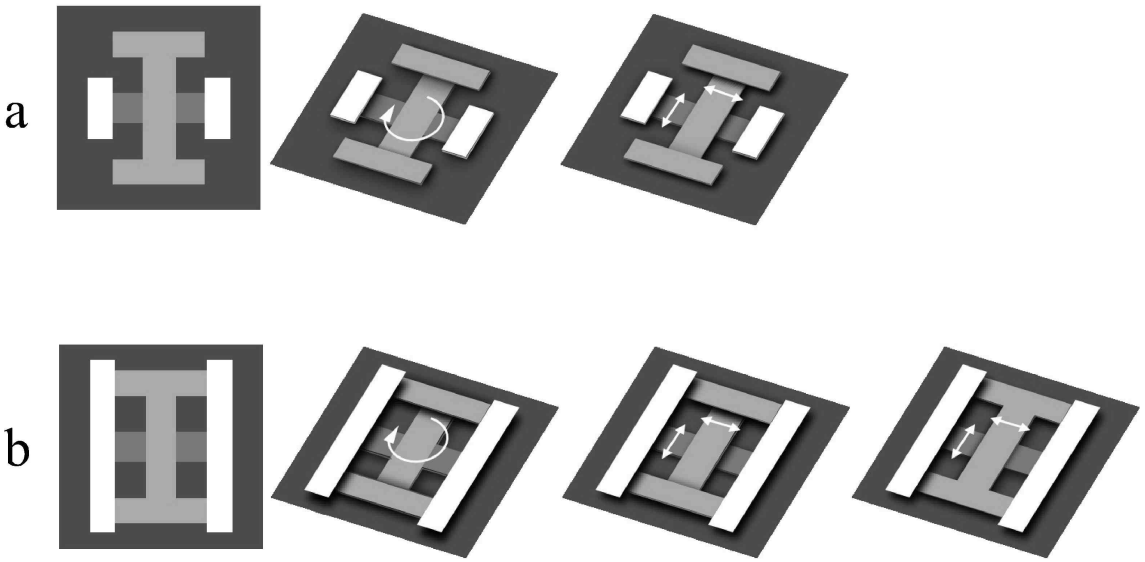

Figure 11. Two variants of the cross stimulus with their perceptual interpretations at the bar-occluder match point. The long occluders in the new configuration allow the horizontal occluders to slide back and forth with the bars, giving rise to a novel third interpretation in which the bars and occluders translate together as a single I-shape.

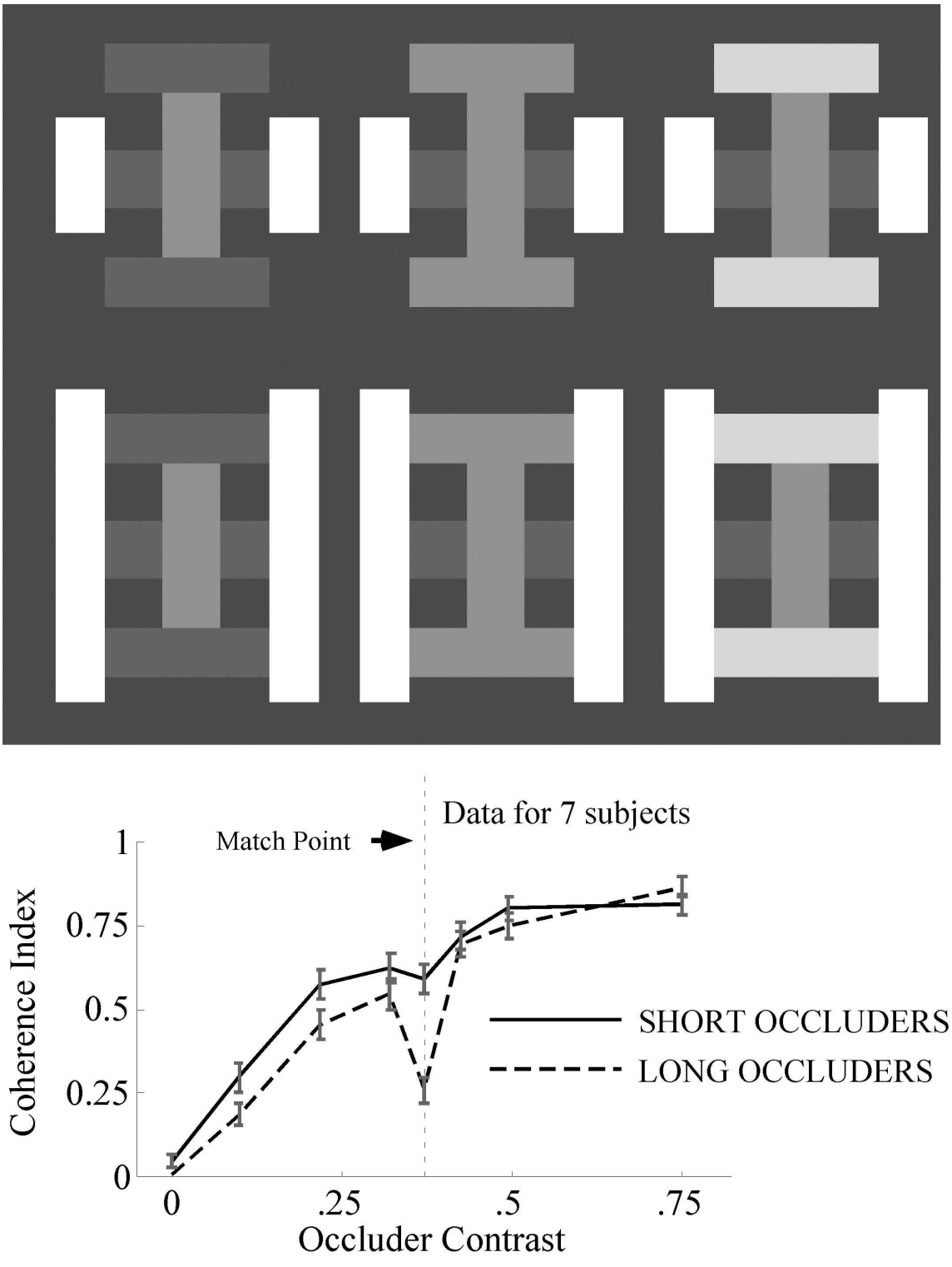

Figure 12. The match point produces a dip in coherence for the new configuration.

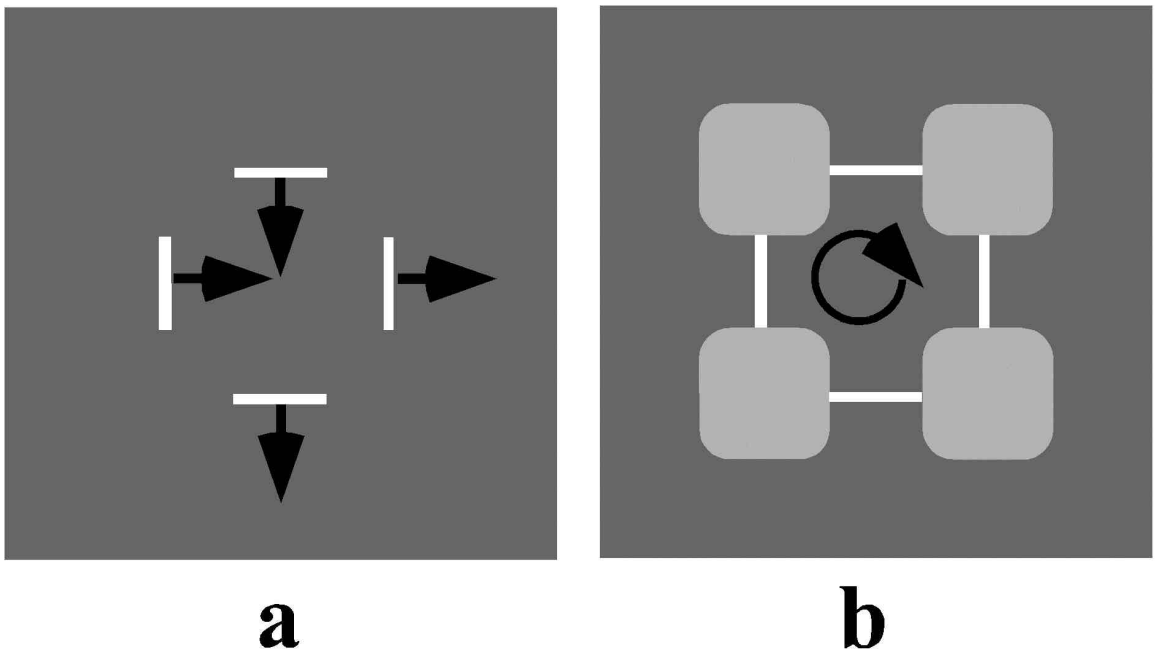

Figure 13. The basic diamond stimulus, generated by moving a diamond in a circle behind occluders, which can either be invisible (a) or visible (b). The arrows denote perceived direction of motion (the image motion is identical in the two stimuli).

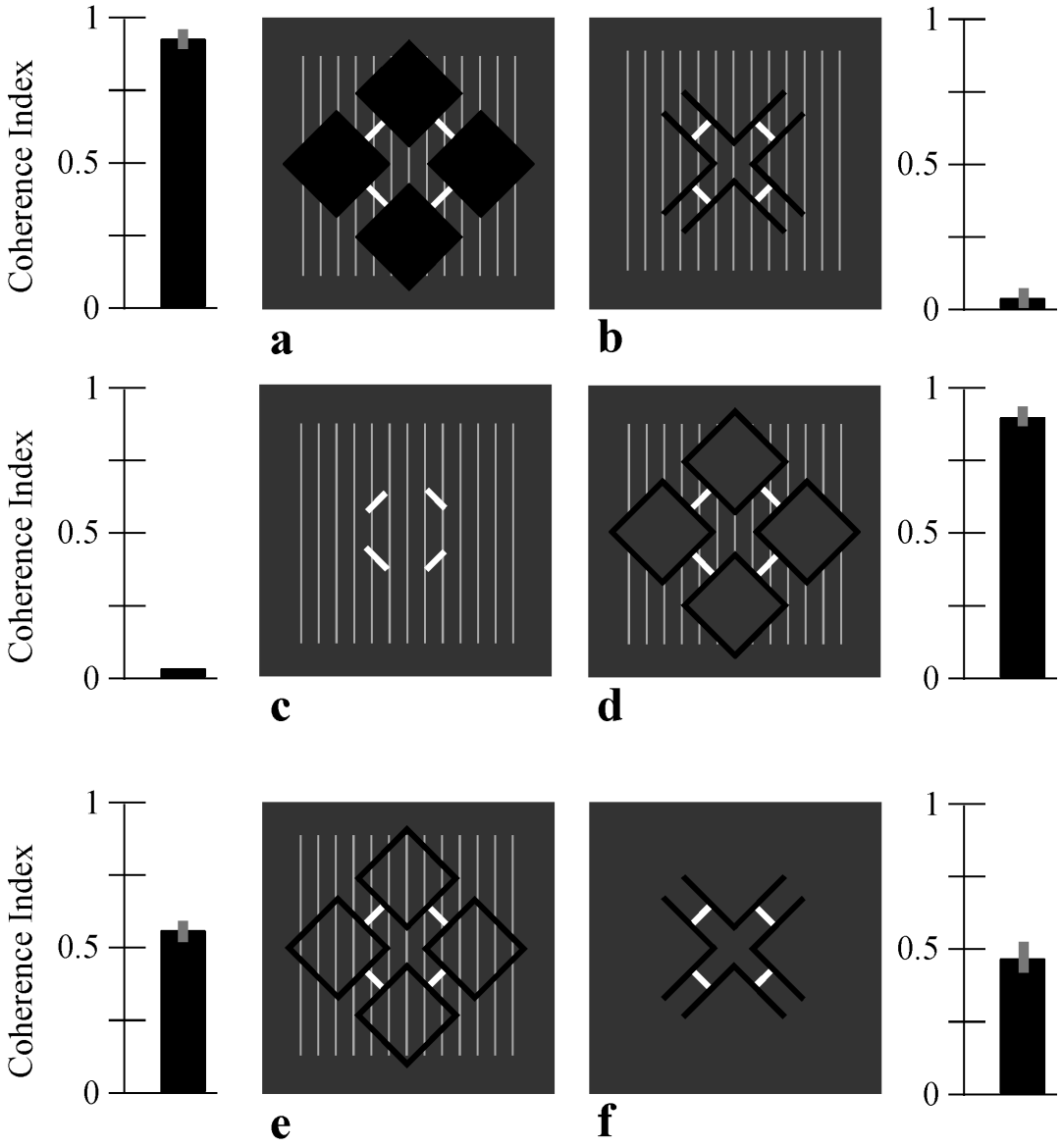

Figure 14. The influence of amodal completion on motion interpretation. (a) Diamond with thick occluders, supporting amodal completion. (b) Diamond with thin occluders, preventing amodal completion. (c) Diamond contours without occluders or T-junctions. (d) Diamond with outline occluders, restoring amodal completion and coherence. (e) Diamond with hollow outline occluders. Coherence is lower than for the solid outline occluders (d), presumably because there is less evidence for an extended occluding surface. (f) Diamond with thin occluders without background lines. Coherence is higher than when background lines are present (b), presumably because it is easier to interpret the Ls as borders of extended occluding surfaces. The results are for eight naive subjects.

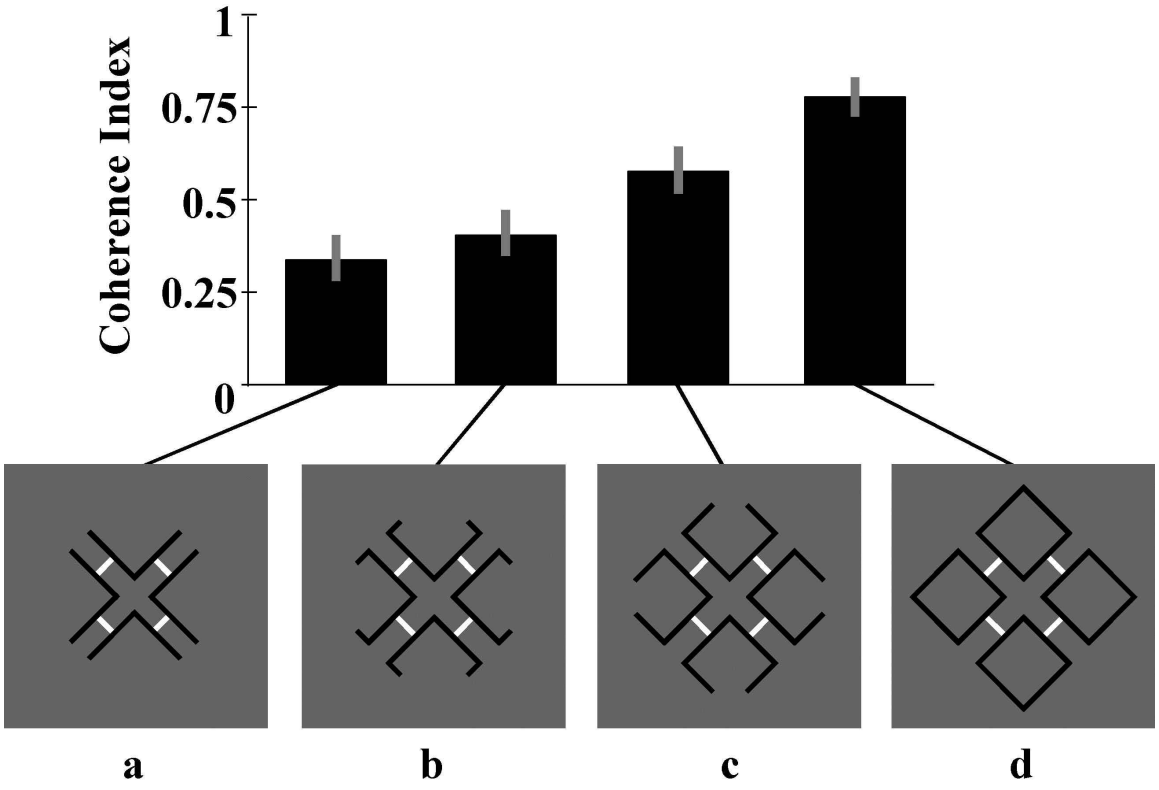

Figure 15. Closure. (a) Diamond with L-shaped occluders, preventing amodal completion. (b) - (d) Increasing closure increases coherence. The results are for five naive subjects.

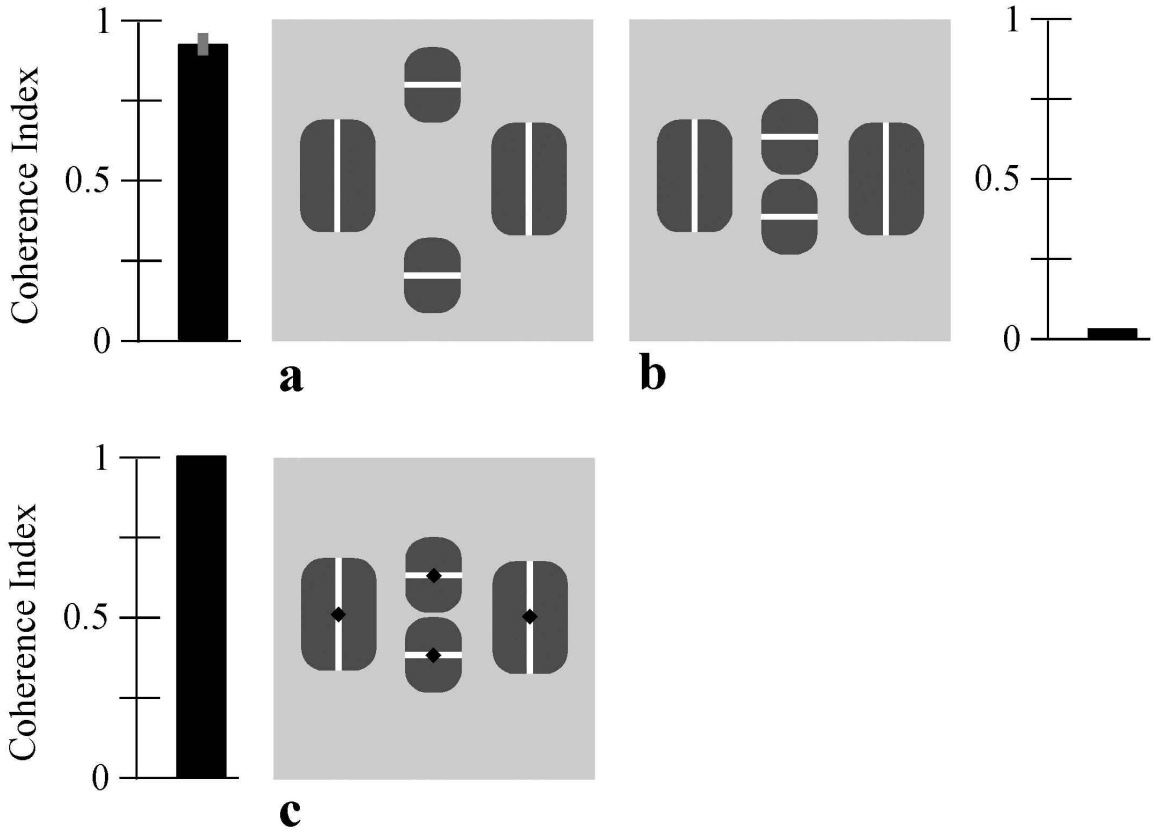

Figure 16. Relatability. (a) Relatable configuration, which generates high coherence. (b) Nonrelatable configuration, which never coheres. (c) Nonrelatable configuration with dots superimposed on the contours. The dots move in the direction of coherent motion, and with their addition the stimulus coheres. The results are for six naive subjects.

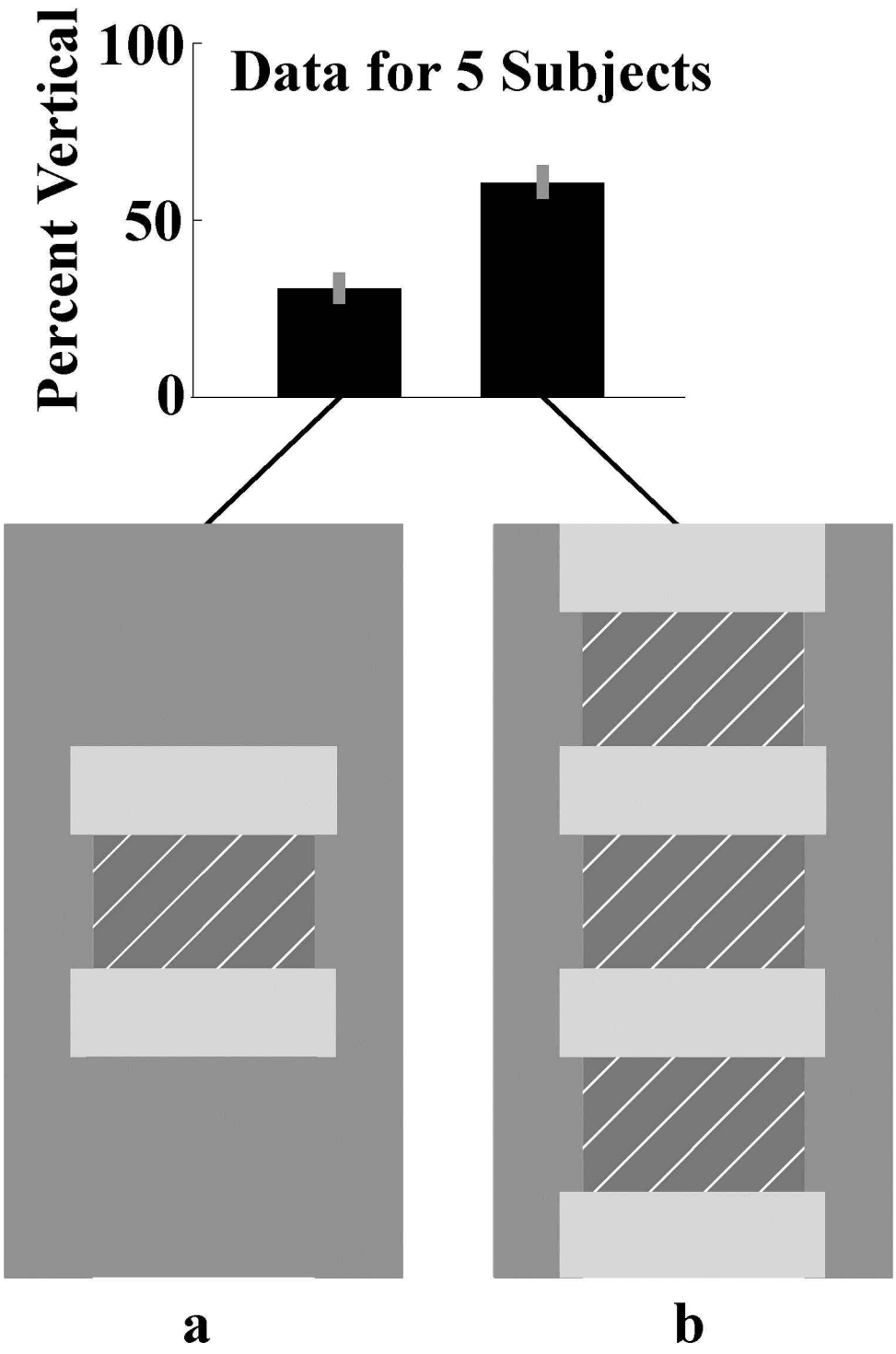

Figure 17. Triple barberpole experiment. The single barberpole appears to move vertically some of the time, but this tendency is enhanced in the triple barberpole.

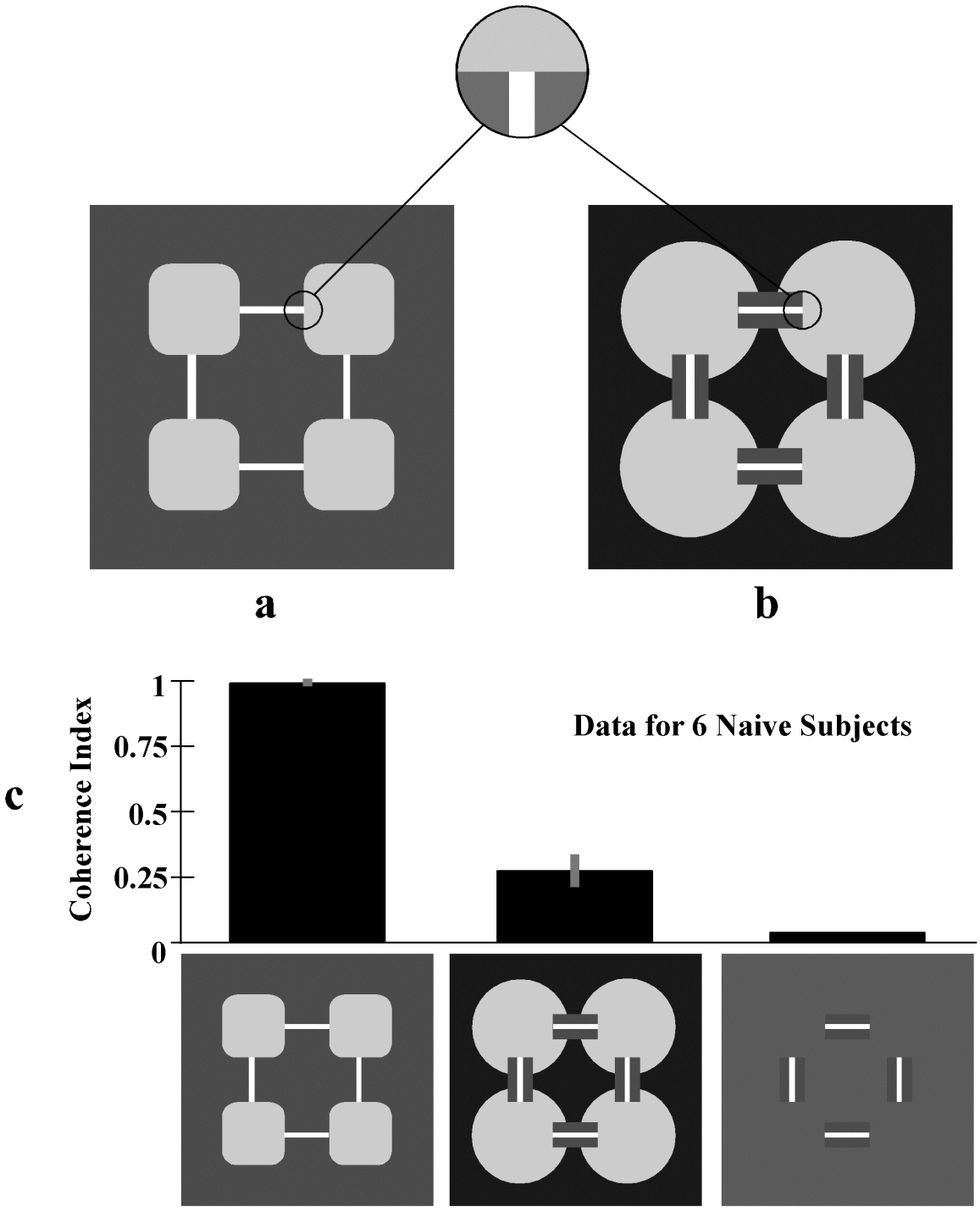

Figure 18. Influence of border ownership on motion interpretation. (a) and (b) Experimental stimuli that are identical in the local vicinity of the diamond contours but which differ globally in the extent to which they support occlusion. (c) Observed coherence levels for each stimulus, for six naive subjects.

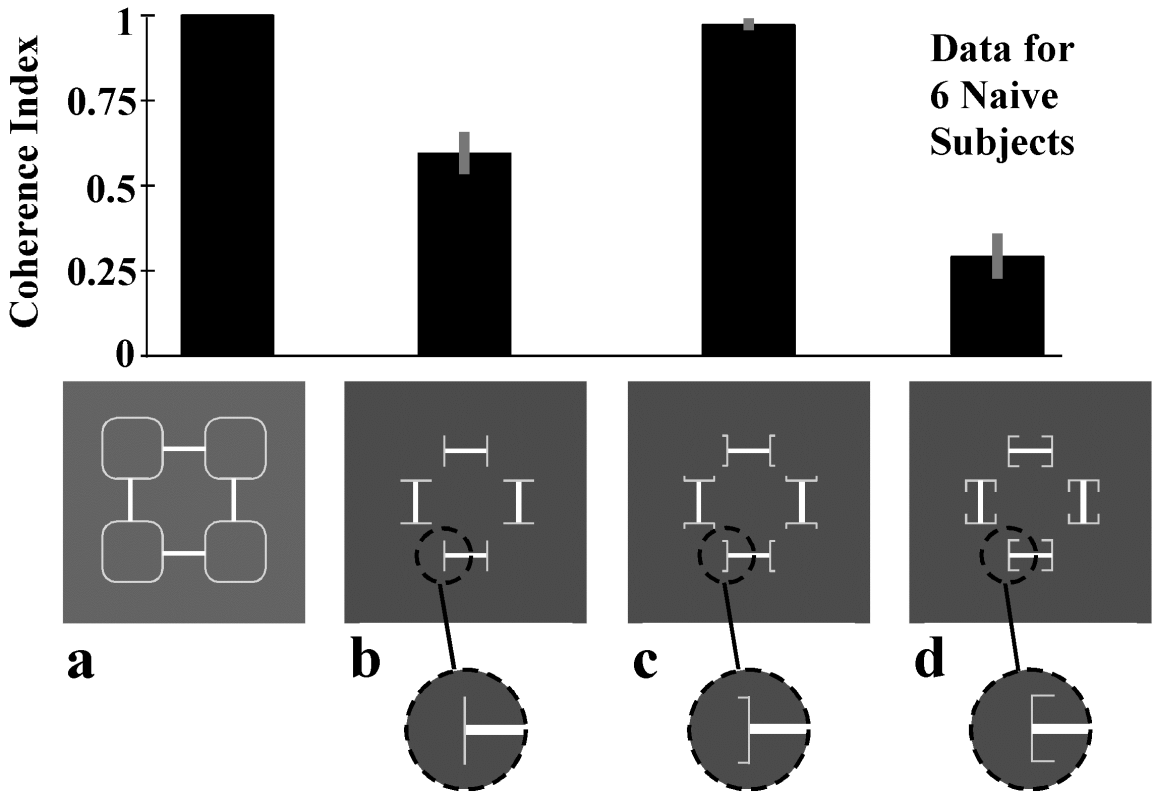

Figure 19. Contour convexity. (a) Outline occluders produce high levels of coherence. (b) T-junctions alone produce intermediate levels of coherence, which is increased by adding convexities (c) and decreased by adding concavities (d).

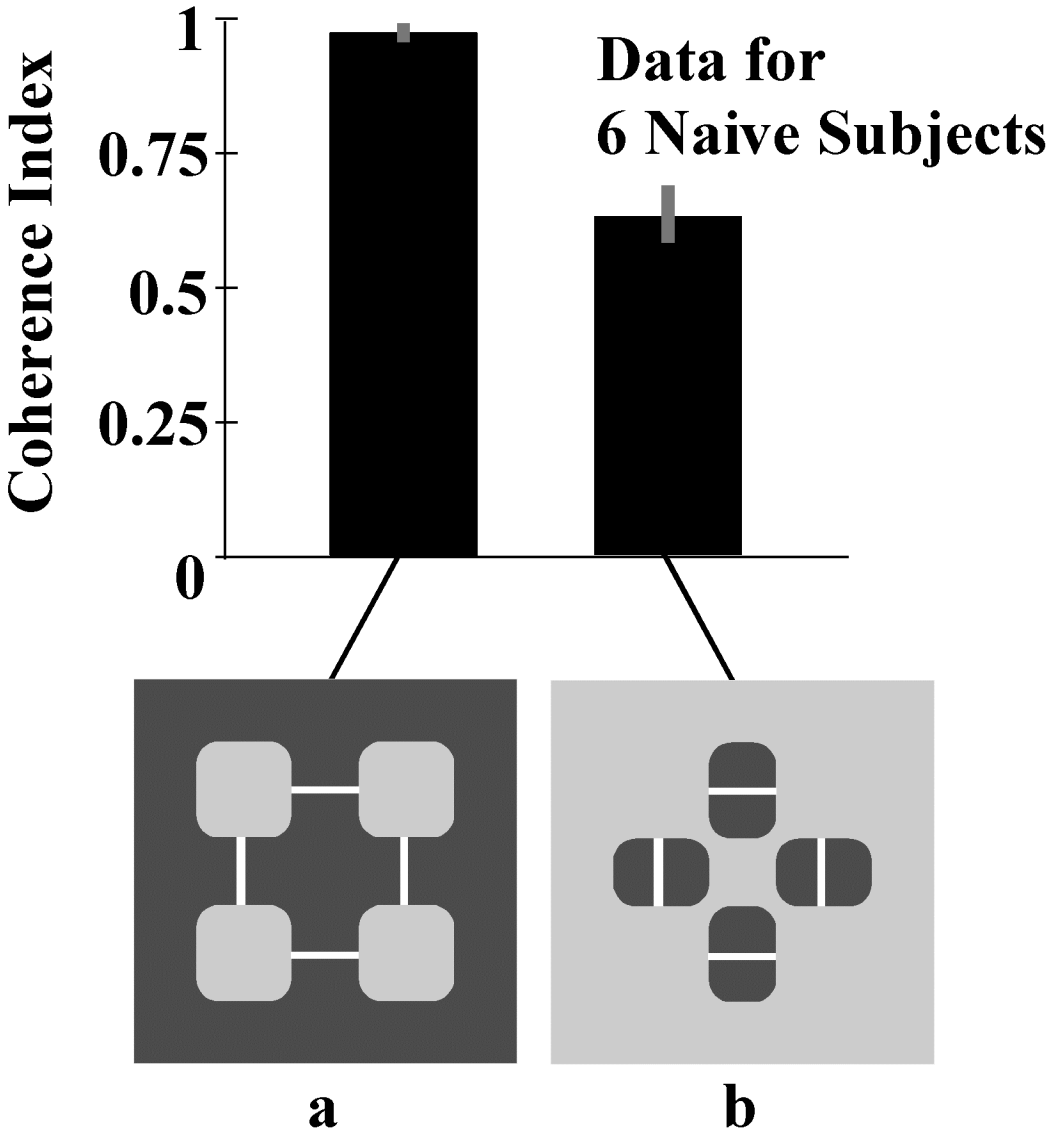

Figure 20. Occluders vs. Apertures. (a) With occluders the square is highly coherent. (b) Apertures with the same occluding contour produce lower coherence, perhaps because the occluding contour is concave.

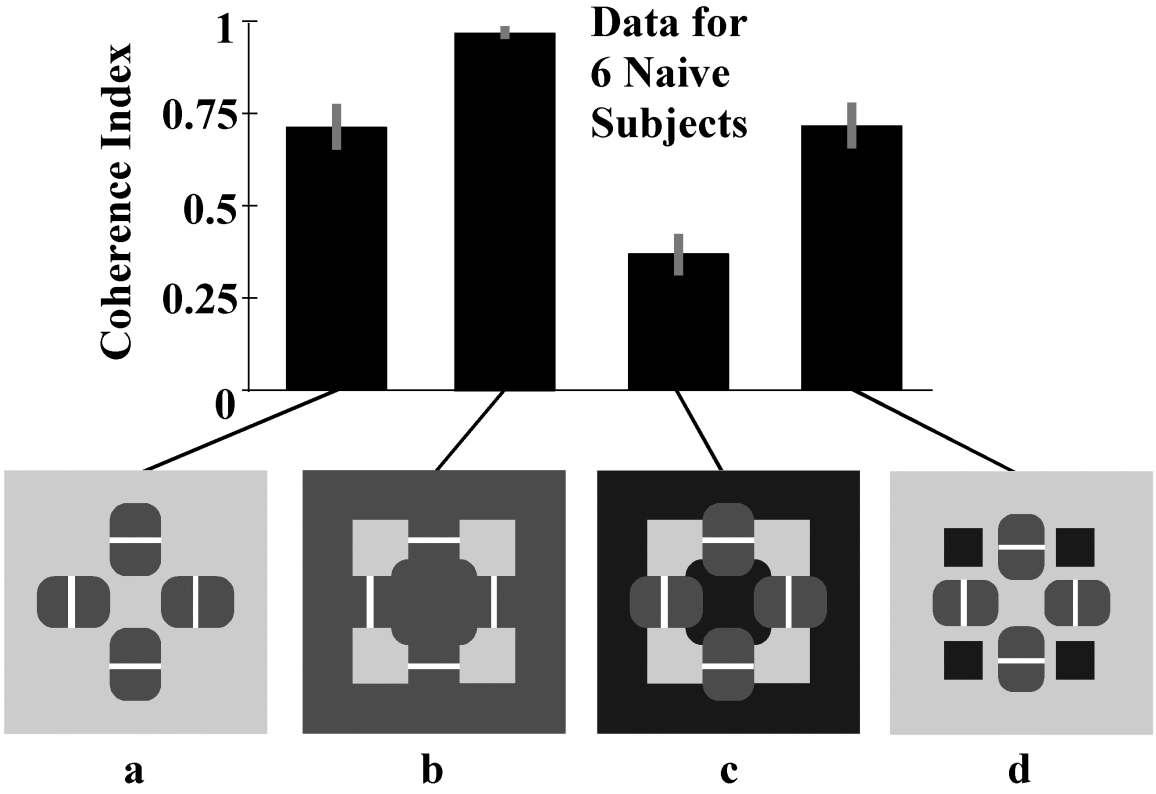

Figure 21. The role of static T-junctions along the occluding contour. When the round apertures of (a) and the occluders of (b) are combined in (c), coherence is lower than it is for either stimulus alone. The control condition in (d) suggests the T-junctions created in (c) are key.