

Junctions and cost functions in motion interpretation

Josh McDermott

Department of Brain and Cognitive Science, MIT,
Cambridge, MA, USA



Gatsby Computational Neuroscience Unit,
University College, London, London, UK

Edward H. Adelson

Department of Brain and Cognitive Science, MIT,
Cambridge, MA, USA



Form, motion, occlusion, and perceptual organization are intimately related. We sought to assess the role of junctions in their interaction. We used stimuli based on a cross moving within an occluding aperture. The two bars of the cross appear to cohere or move separately depending on the context; in accord with prior literature, motion interpretation depends in part on whether the bar endpoints appear to be occluded. To test the importance of junctions in motion interpretation, we explored the effect of changing the junctions generated at the occlusion points in our stimuli, from T-junctions to L-junctions. In some cases, this change had a large effect on perceived motion; in others, it made little difference, suggesting junctions are not the critical variable. Further experiments suggested that what matters is not junctions per se, but whether illusory contours are introduced when the junction category is changed. Our results are consistent with an optimization-based computation that seeks to minimize the presence of illusory contours in the perceptual representation. Although it may be possible to explain our results with interactions between junctions, parsimony favors an explanation in terms of a cost-function operating on layered surface interpretations, with no explicit reference to junctions.

Keywords: motion, aperture problem, occlusion, junctions, surface, genericity, mid-level vision, optimization

Introduction

Although the anatomical pathways for motion and form are largely separate in the early stages of visual processing, it is clear that interactions between motion and form are important. Because of the aperture problem, local motion measurements are inherently ambiguous, and must be combined across space. However, this combination cannot occur blindly - some motions arise from distinct objects and must be segregated; others are the spurious artifacts of occlusion and must be discounted, as shown in [Figure 1](#). In the motion domain, though, spurious features are not obviously distinguishable from veridical ones, and it is not obvious which local motions are due to the same object. Form analysis seems necessary in both cases.

The influence of form and occlusion on motion may be studied with stimuli whose motion is perceptually ambiguous. Wallach (1935; Wuerger, Shapley & Rubin, 1996) adopted this approach with the barber pole stimulus, as have a number of researchers since (Adelson & Movshon, 1984; Shimojo, Silverman, & Nakayama, 1989; Vallortigara & Bressan, 1991; Lorenceau & Shiffrar, 1992; Bressan, Ganis, & Vallortigara, 1993; Trueswell & Hayhoe, 1993; Shiffrar, Li, & Lorenceau, 1995; Shiffrar & Lorenceau, 1996; Stoner & Albright, 1996; Anderson & Sinha, 1997; Castet & Wuerger, 1997; McDermott, Weiss, & Adelson, 1997; Stoner & Albright, 1998; Liden & Mingolla, 1998; Castet, Charton, & Dufour, 1999; Anderson, 1999; McDermott, Weiss, & Adelson, 2001). As shown in [Figure 2](#), in this work, we make use of a stimulus derived

from Anstis's (1990) chopsticks illusion, consisting of two orthogonal bars that move sinusoidally, 90 deg out of phase ([Figure 2a](#) and [2b](#)). When presented together within an occluding aperture ([Figure 2c](#)), the bars perceptually cohere and appear to move in a circle as a solid cross. However, when presented alone ([Figure 2d](#)), they appear to move separately (the horizontal bar translates vertically and the vertical bar translates horizontally), even though the image motion is unchanged.

In either stimulus condition, both percepts are legitimate interpretations of the image motion. In [Figure 2c](#), the bars could be translating separately within the aperture, as in [Figure 2d](#), and in [Figure 2d](#), they could be executing the circular motion of [Figure 2c](#), their endpoints hidden by invisible occluders. Yet a single interpretation is predominantly seen in each case. Of course, this makes sense: For the bars to be moving as a solid cross, they must be occluded, and the presence of visible occluders in the image makes this situation in the world more plausible. But how do the occluders exert their effect? We explore the nature of the form analysis involved.

Most previous theoretical work has supposed that the form influences on motion would be simple and local in nature, and previously documented phenomena are generally consistent with this notion. The form constraints in standard motion models are typically limited to discounting the motions at junctions formed at points of occlusion or transparency (Nowlan & Sejnowski, 1995; Liden & Pack, 1999; Grossberg, Mingolla, & Viswanathan, 2001). Junctions have been suggested as important components of

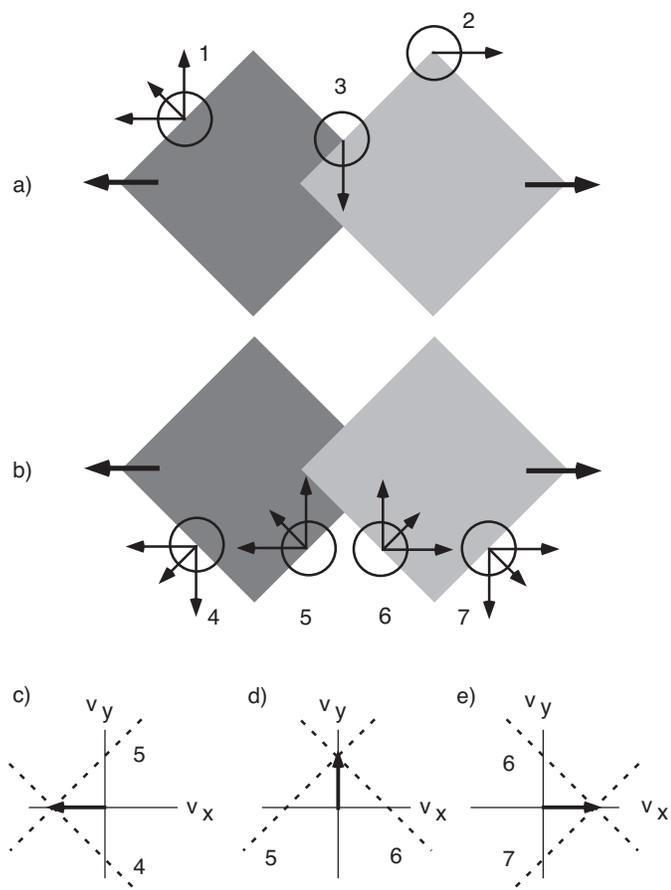


Figure 1. Example illustrating two problems that occur in motion interpretation. In a and b, two squares translate horizontally. The edge motions (e.g., 1) are ambiguous, due to the aperture problem, whereas the corner motions (e.g., 2) are unambiguous. The T-junction motions (e.g., 3) are also unambiguous, 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. Click on the link for a [demo](#).

many aspects of mid-level vision (Guzman, 1969; Stoner & Albright, 1996; Zaidi, Spehar, & Shy, 1997; Saund, 1999; Adelson, 2000; Rubin, 2001), and have therefore seemed a plausible basis for the form constraints on motion perception. In a previous work, we presented several demonstrations that the form constraints on motion interpretation involve amodal completion, border ownership, and depth segregation - considerably more than isolated junctions (McDermott, Weiss, & Adelson, 2001). It nonetheless seemed likely that junctions play an important role, albeit supplemented by more subtle and sophisticated processes. The goal of the present study was to explore the presumed role of junctions in motion interpretation.

Experiment 1: Endpoint junctions

The change in perceived motion that occurs from Figure 2c to 2d is easy to explain in terms of junctions. In Figure 2c, T-junctions are formed where the occluders overlap the crossbars, and offer a plausible cue that the motions of the bar endpoints are spurious and should be discounted. One could suppose that the motions of the bar endpoints are simply ignored by the visual system when the occluders generate T-junctions at those locations. When the endpoints are suppressed, all of the remaining local motions (of the bar edges and intersection) are consistent with a single circular motion, which is what is seen. Without the occluders and the T-junctions they produce, the endpoint motions are not ignored, and two motions, one for each bar, are necessary to explain the image data.

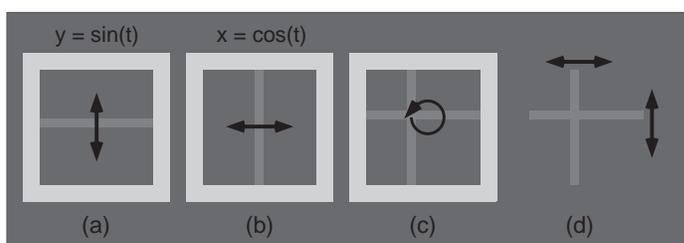


Figure 2. The cross stimulus is generated from two bars that move sinusoidally, 90 deg out of phase. The presence of occluding surfaces alters the interpretation of the motion. Arrows denote perceived motion. Image motion in c and d is identical. Click on the link for a [demo](#).

We attempted to test this story by manipulating the junctions at the bar endpoints. We wondered what would happen if the T-junctions became L-junctions due to matches in luminance between the cross bars and occluders. As shown in Figure 3, we held either the bar contrast or the occluder contrast fixed, and swept the other through the point of accidental match, 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 decrease in the tendency to cohere when the bars and occluders matched in luminance.

In the first experiment, the bar contrast was fixed and nine different occluder contrasts were tested (Figure 3a), running through the point of accidental match. In the second experiment, the occluder contrast was fixed and eight different bar contrasts were tested (Figure 3b), again running through the point of accidental match.

Methods

Stimuli were presented on a Hitachi monitor controlled by a Silicon Graphics Indy R4400. Viewing distance was approximately 95 cm. Subjects were instructed to freely view the experimental stimuli while confining their gaze to the central region of the display. This policy was adopted because (1) subjects found it unnatural and difficult to

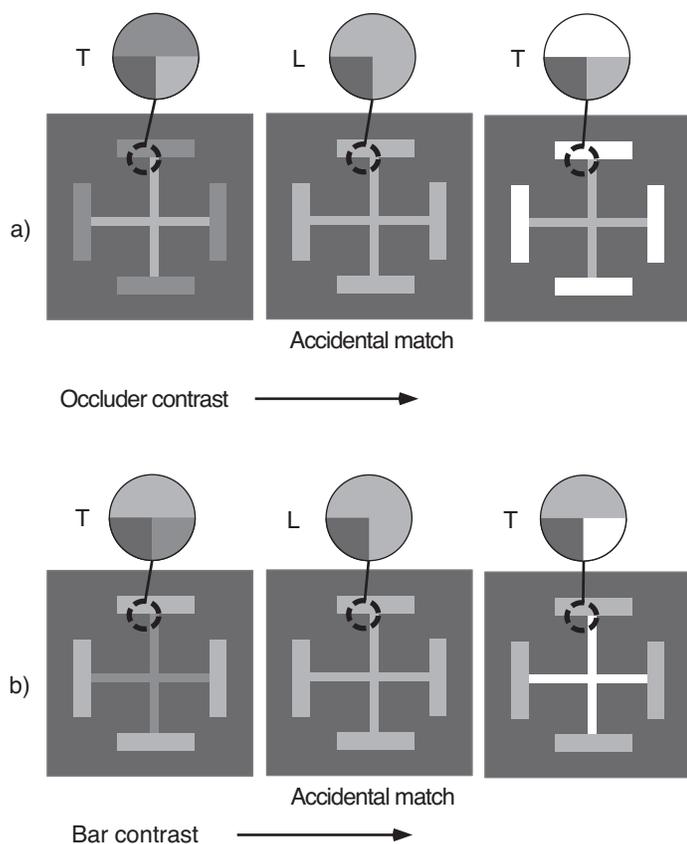


Figure 3. Stimuli for Experiment 1. The effect of junction category was tested by varying bar and occluder contrast and examining the effect of a match in contrast between bars and occluders. Click on the link for a [demo](#).

maintain fixation while the contours of the cross stimulus were moving underneath a fixation point, and (2) free viewing more closely approximates natural viewing conditions. Informal observation by the authors suggested that maintaining fixation would not have qualitatively changed any of the effects described herein.

We used a subjective measure of perception, perceived coherence, rather than the objective direction of rotation judgments that have been used in some past studies (Anstis, 1990; Lorenceau & Shiffrar, 1992). This is because in early pilot experiments, we found that some subjects could, over the course of an experiment, learn to perform the direction of rotation task even under conditions in which they perceived incoherent motion. Such subjects were presumably monitoring the relative phase of the motion of the two bars. Given that the objective task was not measuring the aspects of the percept that we were interested in, we adopted coherence judgments instead.

Subjects used the number pad on the keyboard to enter their responses. Subjects pressed 1, 2, or 3 following each trial to indicate, respectively, completely incoherent (bars moving separately), partially coherent, or completely coherent motion percepts. Subjects' responses were normalized to yield a coherence index ranging from 0 to 1. A coher-

ence index of 0 corresponds to a percept of completely incoherent motion on every single trial, whereas 1 indicates consistently coherent motion. Subjects completed several practice trials before beginning the experimental trials. We discarded the data from subjects who were at ceiling in two or more conditions. In all experiments, the order of stimulus presentation was randomized across trials.

In Experiment 1, the 11 occluder contrasts used were 0, 0.05, 0.125, 0.225, 0.325, 0.355, 0.375, 0.395, 0.425, 0.5, and 0.75. In the second experiment, the 10 bar contrasts used were 0.125, 0.25, 0.325, 0.35, 0.375, 0.4, 0.425, 0.5, .625, and 0.75. The background luminance was 2.5 ftL. Stimulus speed was 2.2 deg/s, and the extent of the stimulus motion was 40 pixels (0.65 deg). The bars were 250 by 20 pixels (4 by 0.32 deg). Each trial lasted 1.5 s, which allowed for approximately two revolutions of the cross. Eight naive MIT students participated in this experiment. Subjects completed 15 trials per condition in a single block.

In Experiment 2, the bars were 200 by 20 pixels (3.25 by 0.32 deg). The occluders were 140 by 60 pixels (2.28 by 1 deg), and their contrast was 0.2. The contrast of one bar was fixed at 0.375; the contrast of the other bar varied across conditions, taking on the values 0.125, 0.25, 0.325, 0.375, 0.425, 0.5, 0.6, and 0.7. All other parameters were as in Experiment 1.

In Experiment 3, the length of the bars was 200 pixels (3.25 deg). The bars in the thin conditions were 20 pixels (0.32 deg) wide; in the thick conditions, they were 70 pixels (1.12 deg). The contrast of one of the bars was adjusted for each subject to avoid floor and ceiling effects, but was always at least 5% above or below the contrast of the bar fixed at the match point of 0.375. One pair of occluders was fixed at a contrast of 0.75; the other varied with condition, taking on the values 0, 0.1, 0.225, 0.325, 0.375, 0.425, 0.5, and 0.75. Other parameters were as in Experiment 1.

In Experiment 4, stimuli in the short occluder conditions were identical to those in the thick bar conditions from Experiment 3; in the long occluder conditions, all parameters were the same except the white occluders were 200 pixels in length, such that they abutted the other occluders.

Results

As shown in Figure 4a and 4b, the dominant effect is an overall shift in coherence with contrast: coherence increases with occluder contrast and decreases with bar contrast. Shapley, Gordon, Truong, and Rubin (1995) obtained similar results with the barberpole stimulus; these contrast effects appear to be a general property of occlusion/motion interactions. The effects may be due to the role that contrast plays as a depth cue (O'Shea, Blackburn, & Ono, 1994; Stoner & Albright, 1998; Rohaly & Wilson,

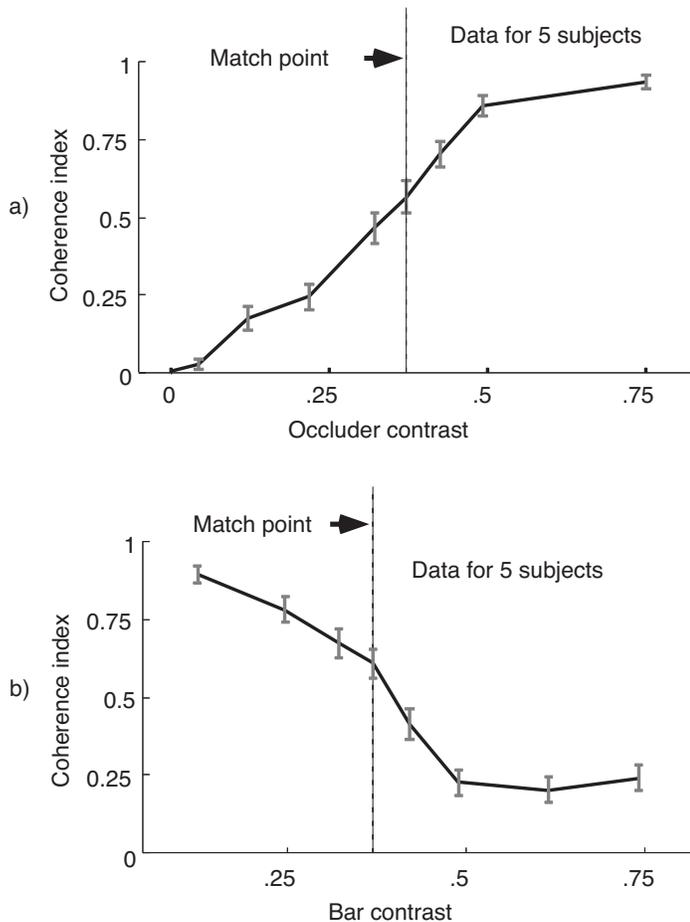


Figure 4. Results of [Experiment 1](#). Click on the link for a [demo](#).

1999), but for the purposes of this work, we simply note that these contrast effects are consistent with prior findings.

More importantly for our purposes, there was no obvious drop in coherence at the point where L-junctions are generated at the bar endpoints, as shown in [Figure 4a](#) and [4b](#). The curve passes smoothly through the match point, and the category of the junction generated at the bar endpoints seems to have little to no effect on the coherence of the cross. In fact, as the occluder contrast decreases (or as the bar contrast increases), the T-junction conditions actually become *less* coherent than the L-junction conditions, a result seemingly at odds with a junction-based mechanism.

Experiment 2: Intersection junctions

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 [Figure 5](#). 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, because 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

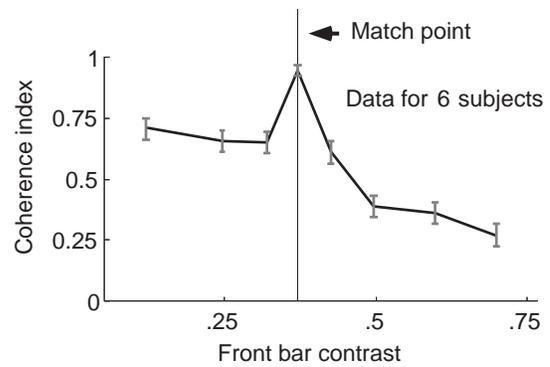


Figure 5. Stimuli and results of [Experiment 2](#). A match between the luminance of the two bars results in a pronounced peak in coherence. Click on the link for a [demo](#).

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 [Figure 5](#). 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 which motion interpretation was chosen. What could explain this pattern of results?

Experiment 3: Controlling for resolution

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 Ts from Ls), it is conceivable that the mechanisms that analyze them for motion interpretation operate at coarse resolution, in which case the change in junction category might not be detected.¹ To test this idea, we made the cross bars thicker, effectively enlarging the pair of junctions formed where the cross bars meet the occluders and degrading the large-scale T-shape formed by the junction configuration.

The problem with simply thickening the bars of the cross is that the cross becomes more coherent overall, particularly when both bars are the same luminance. One explanation is that the length of the contours that have to be completed when the bars are incoherent increases as the bar width is increased, and 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 Figure 6a). As we saw in Experiment 2, this results in lower levels of coherence, which allowed us to change the width of the bars while avoiding ceiling effects.

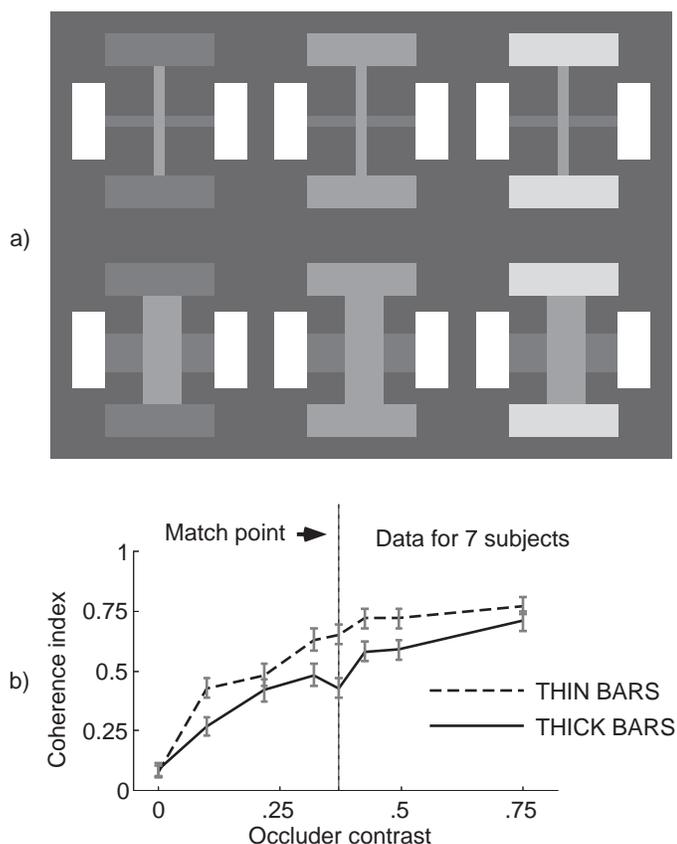


Figure 6. Stimuli and results of Experiment 3. Changing the junctions at the bar endpoints again has little to no effect. Click on the link for a [demo](#).

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 Figure 6b. With thick bars, there was a slight drop in coherence at the match point, but

it was quite small. The dominant effect is that of bar contrast, as before. Even when the junctions are separated by large distances and are easy to resolve, their category is of little consequence.

Experiment 4: 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 Figure 7a, 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 Figure 7c. 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 probability and parsimony, 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 (Figure 7b) rather than occlusion points, and T-junctions, which do not require postulating such edges, would clearly be the stronger occlusion cue.

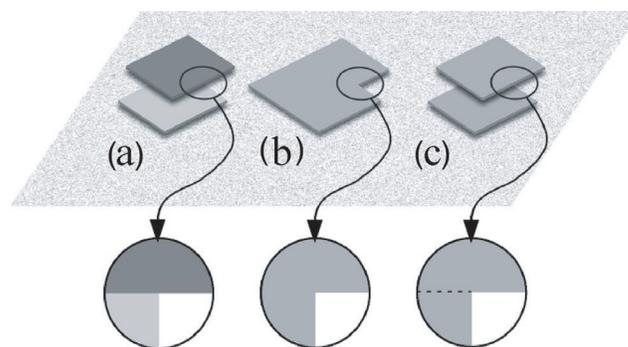


Figure 7. T-junctions are generically associated with occlusion; L-junctions are not.

Because the coherence of the cross seems to depend on evidence for occlusion, one might expect lower coherence at the point of accidental match, where L-junctions are generated at the bar endpoints. On inspection, however, both the coherent and incoherent percepts of the cross necessitate a discontinuity between the occluders and bars. As shown in Figure 8a, 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 be no less likely at the match point despite the presence of L-junctions.

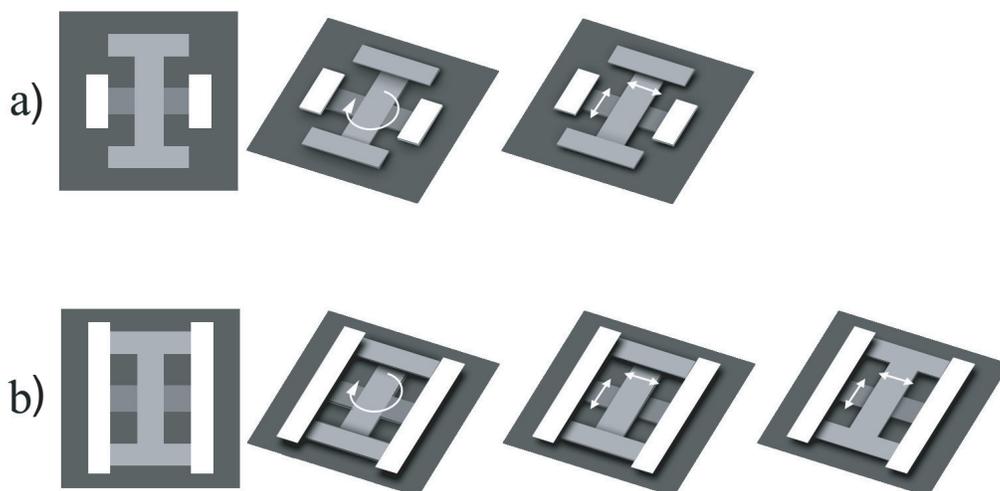


Figure 8. New and old stimulus configurations with their perceptual interpretations.

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.

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 Figure 8a, 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 Figure 8b, 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 Figure 8b (far right), in which the I-shape is seen to move back and forth without any discontinuity between the bar and the occluders. 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, on the other hand, 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, because there would be reason to prefer the incoherent interpretation. We therefore conducted another match point experiment with both configurations of Figure 8, varying the luminance of one pair of the occluders and looking for an effect where they matched the bar luminance.

As shown in Figure 9, 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 Experiment 2. 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 con-

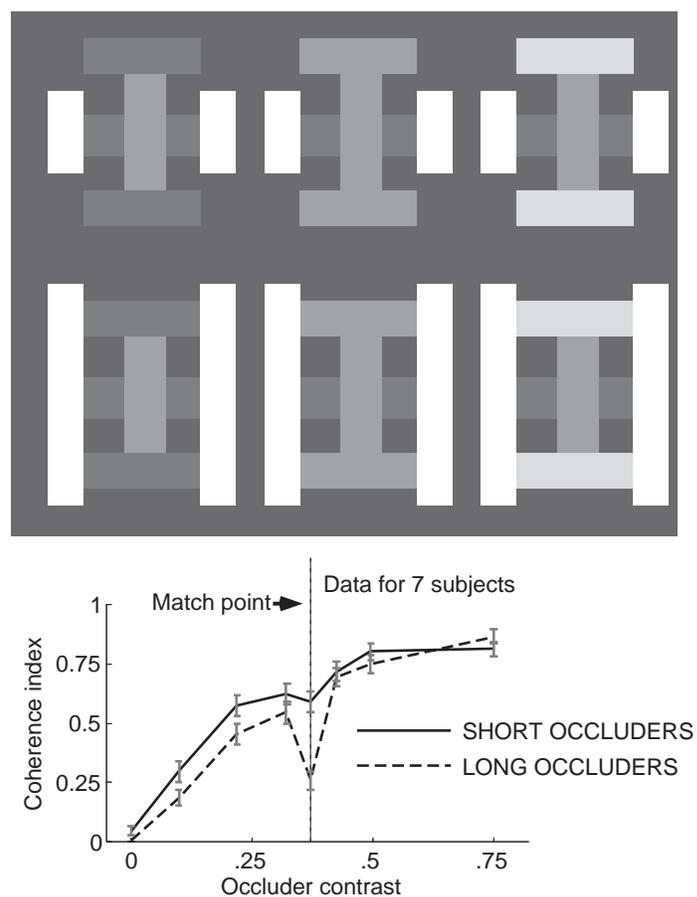


Figure 9. Stimuli and results of Experiment 4. The match point matters in the new configuration. Click on the link for a [demo](#).

figuration. This result is just that predicted by a computation minimizing the number of illusory edges in the perceptual interpretation. What seems to matter is the presence or absence of surface discontinuities, but only when they are not signaled by edges in the image.

Discussion

The experiments in this work were designed to test the role of local, junction-based computations in motion interpretation. We found that junction categories were of little value in predicting the motions that were seen. It seems instead that the visual system is executing a computation involving the minimization of illusory edges. A change in junction category leads to a change in motion percept only if it also leads to a change in illusory edge count; thus, the illusory edges, and not the junctions, are doing the explanatory work.

Figures 10 and 11 summarize the key stimuli from all of our experiments, and the various possible perceptual interpretations. All the effects, or lack thereof, can be predicted by considering the illusory edges generated in the different interpretations of a stimulus. For instance, in the basic occluded cross stimulus (Figure 10a), only the incoherent interpretation necessitates illusory contours (where the two moving bars overlap), and we correctly predict a preference for coherence for this stimulus. In contrast, when the occluding frame is removed in the stimulus of Figure 10b, both interpretations involve illusory contours, but the incoherent interpretation contains fewer of them, consistent with its status as the preferred percept.

To predict the results of the match experiments, we consider whether there is a difference in the number of illusory edges present in the coherent and incoherent interpretations. If this difference is different across stimuli, then we predict a change in the tendency to cohere from one stimulus to the other. In Experiment 1, when the bars and occluders matched in luminance (Figure 10c), both the coherent and incoherent percepts have discontinuities between the bars and occluders that are not present as edges in the stimulus itself. The incoherent percept also has illusory contours where the moving bars overlap, but these are also present in the nonmatched stimuli (Figure 10d). Thus, we correctly predict no effect of the accidental match on motion perception – coherence is no more likely at the match point than it is off of it, because the competing percept is equally penalized. In contrast, when the two bars are set to different luminance values (Figure 10e) as in Experiment 2, the illusory edges in the incoherent percept are only present for the matched stimulus (Figure 10d). Thus, we correctly predict a preference for the coherent percept at the match point, as it has fewer illusory edges than the incoherent interpretation. In Experiment 3, as in Experiment 1, the two percepts again both have illusory edges at the match point (Figure 11a and 11b), and we again correctly predict no drop in coherence. In the new

stimulus of Experiment 4 (Figure 11c and 11d), the coherent percept again has the illusory edges at the match point, but due to the stimulus manipulation, there are two incoherent percepts, one in which the occluders move as a single surface with the bar they are matched with. The incoherent percept thus need not have the illusory edge, and a computation attempting to minimize such edges would predict that incoherence would increase at the match point, which it does. To summarize, a computation seeking to minimize illusory edges correctly predicts the presence or absence of match point effects in each of our experiments, whereas junction category does not.

Before embarking on these experiments, we assumed, as others might have, that the coherence of our stimuli would depend mainly on the strength of local occlusion derivable from an analysis of junction category. In hindsight, this is plainly incorrect. Perceived motion appears to be determined by a comparison between different interpretations of the image motion (in this case, coherent vs. incoherent). If the coherent interpretation better satisfies some criteria (in this case, it appears to be one related to illusory edges), coherent motion can be seen even if the evidence for occlusion is otherwise weak, as it is when the bars and occluders match in luminance.

It should also be noted that the illusory edges that seem to be affecting motion perception are not evident from a static analysis of the stimuli. A single stimulus frame is insufficient to determine the layered interpretations that define the illusory edges. The form computations involved are evidently reciprocally dependent on motion information (Wallach, 1935; Anderson & Sinha, 1997; Watanabe, 1997).

Surfaces appear to be the natural representation with which to think of these phenomena, because the discontinuities that seem to be critical are not defined unless the stimulus has been segmented into surfaces. One candidate computation would be a cost function on layered surface interpretations of the image motion that penalizes non-generic interpretations (i.e., those containing edges not present in the image). Coming up with the family of possible interpretations is another matter, but once they are available, a single, simple cost function may be able to predict what we see.

Explanations of perceptual phenomena in terms of optimization and cost functions have a long history in perception. Helmholtz (1867) advocated the idea of finding the most likely interpretation of the sensory data, and others (e.g., Hochberg, 1953; Attneave, 1954; Leeuwenberg, 1969; Mumford, 1995) have proposed that humans seek to minimize the complexity of image descriptions. In motion perception, Restle (1979), Hildreth (1984), Grzywacz and Yuille (1991), Weiss, Simoncelli, and Adelson (2002) and others have had success with various minimization rules. Another approach to perception is to describe processes that act on features of the stimuli. Most motion models that incorporate form cues are of this nature (e.g., Nowlan

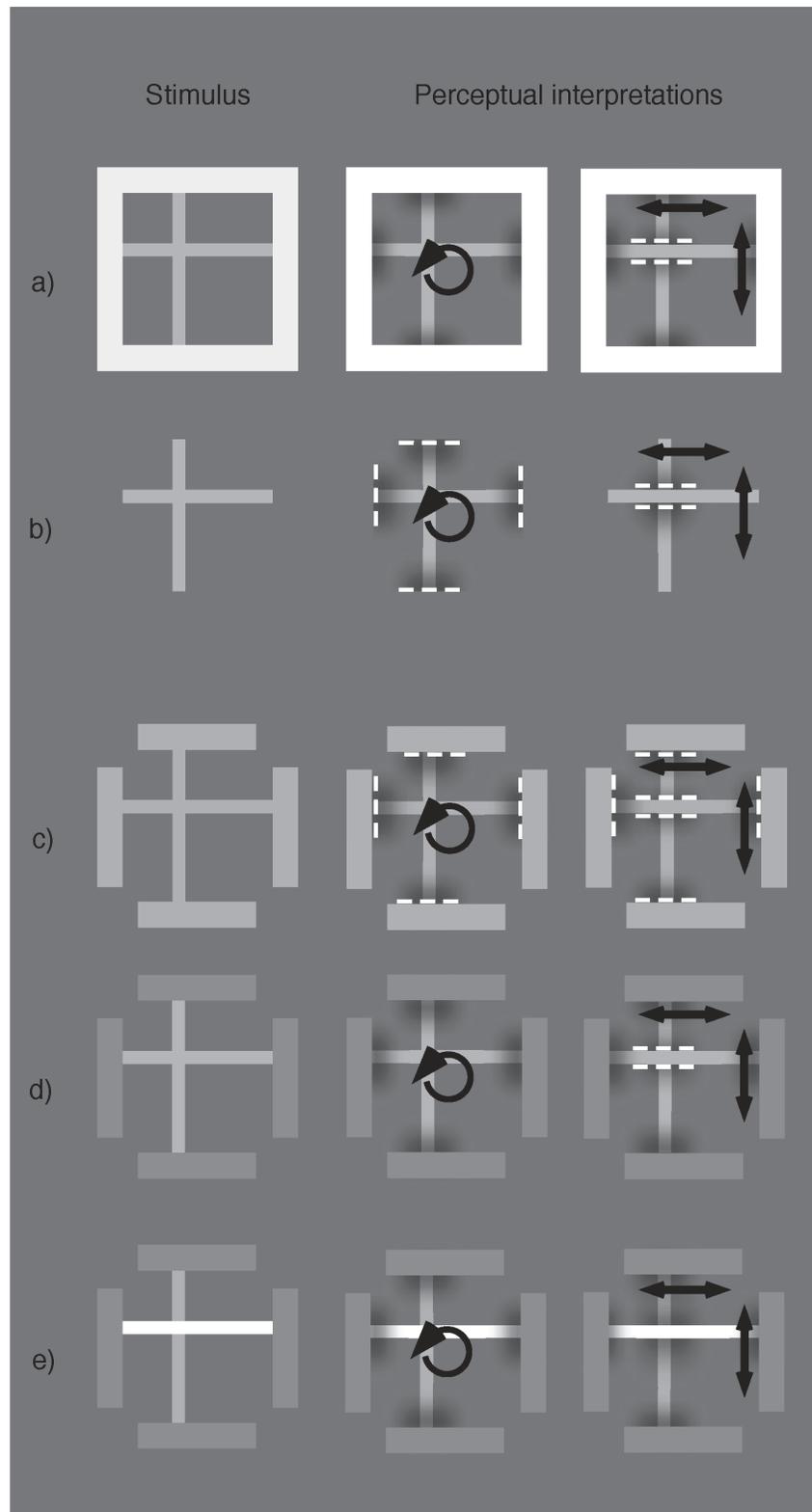


Figure 10. Summary of stimuli and their perceptual interpretations for the basic stimulus as well as [Experiments 1](#) and [2](#). Stimuli are in the leftmost column. Their perceptual interpretations in the two right columns are depicted with the use of drop shadows (to indicate depth discontinuities) and dashed lines (to indicate illusory contours). Arrows indicate perceived motion. [a](#) and [b](#) depict the basic effect of adding occluders to the cross bars. Without occluders, there are more illusory contours in the coherent interpretation than in the incoherent, but with occluders, the reverse is true. [c](#) and [d](#) depict the key conditions of [Experiment 1](#), which tested the effect of changing the endpoint junctions. [d](#) and [e](#) depict the key conditions of [Experiment 2](#), which tested the effect of changing the center junctions. See text for details.

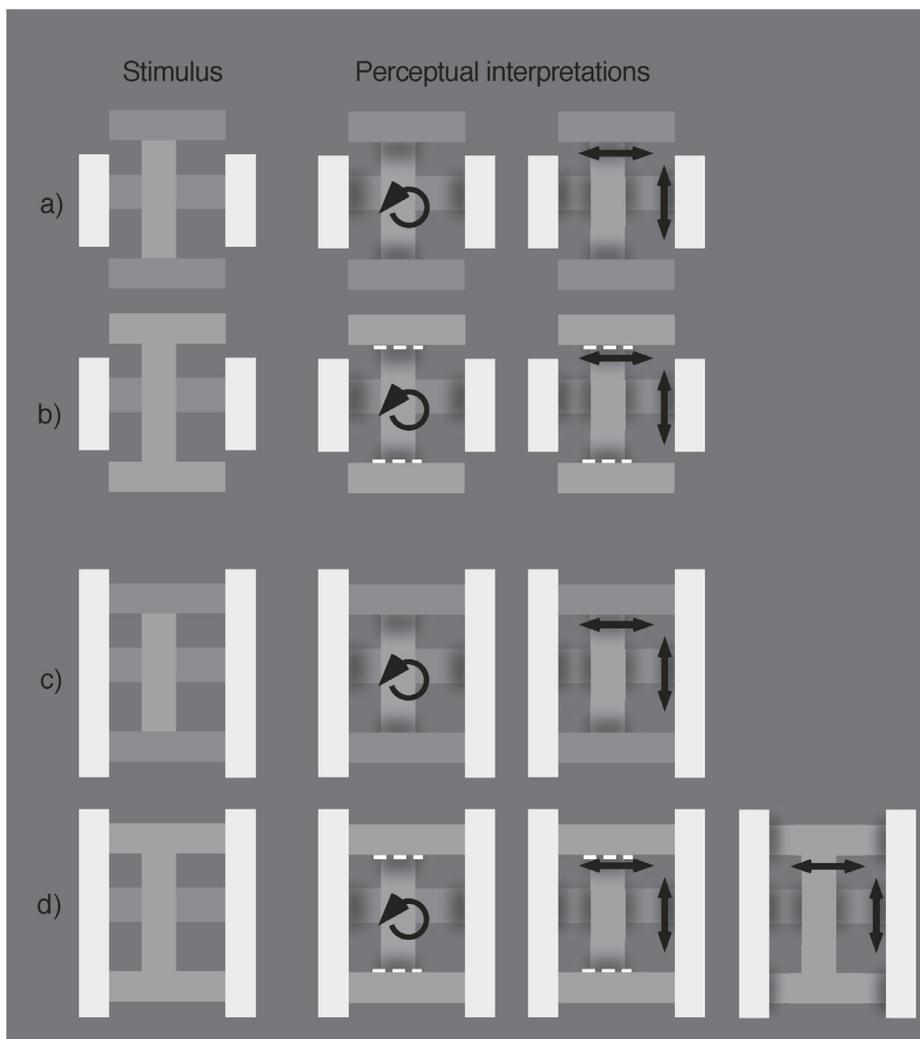


Figure 11. Summary of stimuli and perceptual interpretations for Experiments 3 and 4. Drop shadows and dashed lines are used as in Figure 10. a and b depict stimuli from Experiment 3, which again explore the effect of changing the junction category at the bar endpoints. The absence of an effect is well accounted for illusory edges, which are present in equal amounts in both perceptual interpretations. c and d depict stimuli from the new configuration introduced in Experiment 4, again with T-junctions (nonmatch) and L-junctions (match) at the bar endpoints. In the latter case, there is a distinct motion percept (far right), which lacks the illusory edges of the other percepts, and thus seems to be favored.

& Sejnowski, 1995; Grossberg et al., 2001), detecting junctions and altering motion analysis in some way as a function of the junction, usually by suppressing motions that occur at T- or X-junctions. For many basic stimulus manipulations, this approach may work, but our phenomena seem much easier to describe in terms of a cost function that operates on layered surface interpretations. A cost function approach does not specify how the optimization procedure is implemented in the brain, of course, and it is possible that junctions and processes that act on them are important at this level. However, they do not appear to allow for a concise description of the computation. In particular, the results clearly cannot be predicted by looking at individual junctions, and our data are thus inconsistent with a process based just on junctions. Note that our observation of a pronounced effect at the match point in some

cases but not others demonstrates that the stimulus differences defining the junction categories are indeed sensed by the visual system. They just do not appear to matter unless they differentially affect the illusory edges in the image interpretations.

The illusory edge minimization that seems to be at work in our effects can be viewed as one example of a genericity-based computation. The notion of genericity was introduced in computer vision (Clowes, 1971; Huffman, 1971; Koenderink & van Doorn, 1976; Barrow & Tenenbaum, 1981; Binford, 1981; Witkin & Tenenbaum, 1983; Lowe & Binford, 1985; Malik, 1987; Richards, Koenderink, & Hoffman, 1987) to formalize the intuition that certain image interpretations contain accidental matches (e.g., between viewing angle and object pose), and should be rejected by the visual system as unlikely coinci-

dences. Perceptual preferences for generic interpretations have been shown to fall naturally out of a probabilistic framework for perception (Freeman, 1994), and have been well documented in human vision (Rock, 1983; Nakayama & Shimojo, 1992; Albright, 2001).

Genericity has classically been applied to viewpoint and object pose, but is equally applicable to any two variables that describe a scene. For our phenomena, the variables of interest are the albedos (grey levels) of the surfaces that generate the junction in question. When there are two surfaces that match in albedo and that therefore produce an illusory edge, the situation is nongeneric. If the surfaces have different albedos, or if there is only one surface (with a corner) forming the junction, the situation is generic as there is no accidental match, and for the same reason there is no illusory edge. Our experiments demonstrate that changing a T-junction to an L-junction alters perceived motion only when one motion interpretation is more generic than the other, by virtue of segmenting two regions of the same luminance (the two bars, or one bar and its occluders) into a single surface and thus eliminating a potential accidental match. The minimization of illusory edges can thus be viewed as an instance of a computation favoring generic image interpretations and minimizing the postulation of coincidences in the world.

Conclusions

Previous studies with barberpole, plaid, diamond, and other stimuli have demonstrated numerous form and stereo influences on motion, presumably related to occlusion and transparency (Wallach, 1935; Adelson & Movshon, 1984; Shimojo et al., 1989; Vallortigara & Bressan, 1991; Lorenceau & Shiffrar, 1992; Bressan et al., 1993; Trueswell & Hayhoe, 1993; Shiffrar et al., 1995; Shiffrar & Lorenceau, 1996; Stoner & Albright, 1996; Anderson & Sinha, 1997; Castet & Wuerger, 1997; Stoner & Albright, 1998; Liden & Mingolla, 1998; Castet, Charton, & Dufour, 1999; Anderson, 1999). In this paper, we have extended this work in the domain of form. Our experiments rule out a number of intuitively plausible models of form-motion interactions that were consistent with much previous data. For instance, the idea that motions might be discounted at junctions consistent with occlusion, embodied in the model of Nowlan and Sejnowski (1995), is clearly inconsistent with our results. Junctions by themselves do not seem to greatly affect motion interpretation. Another plausible idea, embodied in the model of Liden and Pack (1999), is that motion interpretation might be handed the output of static occlusion analysis. This too seems inconsistent with our results; static cues cannot predict the effects. Our results suggest that surfaces and cost functions may figure prominently in the computations underlying motion perception.

Acknowledgments

This work was funded by National Institutes of Health Grants EY11005-04 and EY12690-02 (EA). JM was supported by the Gatsby Charitable Foundation and a Marshall Scholarship.

Commercial relationships: None.

Corresponding author: Josh McDermott.

Email: jhm@mit.edu.

Address: NE20-444, MIT, 3 Cambridge Center, Cambridge MA 02139.

Footnotes

¹This would be consistent with our observations (McDermott, Weiss, & Adelson, 1998) that small gaps between the bar endpoints and the occluders also do not appear to be resolved by motion perception.

References

- Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 339-351). Cambridge, MA: MIT Press.
- Adelson, E. H., & Movshon, J. A. (1984). Binocular disparity and the computation of two-dimensional motion. *Journal of the Optical Society of America*, *1*, 1266.
- Albright, M. K. (2001). Surface perception and the generic view principle. *Trends in Cognitive Sciences*, *5*, 197-203. [[PubMed](#)]
- Anderson, B. L., & Sinha, P. (1997). Reciprocal interactions between occlusion and motion computations. *Proceedings of the National Academy of Sciences U.S.A.*, *94*, 3477-3480. [[PubMed](#)]
- Anderson, B. L. (1999). Stereoscopic occlusion and the aperture problem for motion: A new solution. *Vision Research*, *39*, 1273-1284. [[PubMed](#)]
- Anstis, S. (1990). Imperceptible intersections: The chopsticks illusion. In A. Blake & T. Troscianko (Eds.), *AI and the eye*. New York: Wiley.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183-193 [[PubMed](#)].
- Barrow, H. G., & Tenenbaum, J. M. (1981). Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, *17*, 75-116.
- Binford, T. O. (1981). Inferring surfaces from images. *Artificial Intelligence*, *17*, 205-244.
- Bressan, P., Ganis, G., & Vallortigara, G. (1993). The role of depth stratification in the solution of the aperture problem. *Perception*, *22*, 215-228. [[PubMed](#)]

- Castet, E., Charton, V., & Dufour, A. (1999). The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. *Vision Research*, *39*, 915-932. [PubMed]
- Castet, E., & Wuerger, S. (1997). Perception of moving lines: Interactions between local perpendicular signals and 2D motion signals. *Vision Research*, *37*, 705-720. [PubMed]
- Clowes, M. B. (1971). On seeing things. *Artificial Intelligence*, *2*, 79-116.
- Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, *368*, 542-545. [PubMed]
- Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, *41*, 2521-2553. [PubMed]
- Grzywacz, N. M., & Yuille, A. L. (1991). Theories for the visual perception of local velocity and coherent motion. In J. Landy & J. Movshon (Eds.), *Computational models of visual processing*. Cambridge, MA: MIT Press.
- Guzman, A. (1969). Decomposition of a visual scene into three-dimensional bodies. In A. Grasselli (Ed.), *Automatic interpretation and classification of images* (pp. 243-276). New York: Academic Press.
- Helmholtz, H. v. (1867). *Handbuch der physiologischen Optik*. Leipzig: Voss.
- Hildreth, E. C. (1984). *The measurement of visual motion*. Cambridge, MA: MIT Press.
- Hochberg, J., & McAlister, E. (1953). A quantitative approach to figural "goodness." *Journal of Experimental Psychology*, *46*, 362-364. [PubMed]
- Huffman, D. A. (1971). Impossible objects as nonsense sentences. *Machine Intelligence*, *8*, 475-492.
- Koenderink, J., & van Doorn, A. J. (1976). The singularities of the visual mapping. *Biological Cybernetics*, *24*, 51-59. [PubMed]
- Leeuwenberg, E. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, *76*, 216-220. [PubMed]
- Liden, L., & Mingolla, E. (1998). Monocular occlusion cues alter the influence of terminator motion in the barber pole phenomenon. *Vision Research*, *38*, 3883-3898. [PubMed]
- 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. [PubMed]
- Lorenceanu, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research*, *32*, 263-273. [PubMed]
- Lowe, D. G., & Binford, T. O. (1985). The recovery of three-dimensional structure from image curves. *IEEE Transactions on PAMI*, *7*, 320-326.
- Malik, J. (1987). Interpreting line drawings of curved objects. *International Journal of Computer Vision*, *1*, 73-103.
- McDermott, J., Weiss, Y., & Adelson, E. H. (1997). Surface perception and motion integration [Abstract]. *Investigative Ophthalmology and Vision Research*, *38*(Suppl.), S237.
- McDermott, J., Weiss, Y., Adelson, E. H. (1998). What makes a good T-junction? [Abstract] *Perception*, *27*(Suppl.), 40.
- McDermott, J., Weiss, Y., & Adelson, E. H. (2001). Beyond junctions: Nonlocal form constraints on motion interpretation. *Perception*, *30*, 905-923. [PubMed]
- Mumford, D. (1995). Pattern theory: A unifying perspective. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference*. Cambridge: Cambridge University Press.
- Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357-1363. [PubMed]
- Nowlan, S., & Sejnowski, T. (1995). A selection model for motion processing in area MT of primates. *Journal of Neuroscience*, *15*, 1195-1214. [PubMed]
- O'Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research*, *34*, 1595-1604.
- Restle, F. (1979). Coding theory and the perception of motion configurations. *Psychological Review*, *86*, 1-24. [PubMed]
- Richards, W. A., Koenderink, J. J., & Hoffman, D. D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *Journal of the Optical Society of America A*, *4*, 1168-1175.
- Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- Rohaly, A. M., & Wilson, H. R. (1999). The effects of contrast on perceived depth and depth discrimination. *Vision Research*, *39*, 9-18. [PubMed]
- Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception*, *30*, 339-366. [PubMed]
- Saund, E. (1999). Perceptual organization of occluding contours of opaque surfaces. *Computer Vision and Image Understanding*, *76*, 70-82.
- Shapley, R., Gordon, J., Truong, C., & Rubin, N. (1995). Effect of contrast on perceived direction of motion in the barberpole illusion [Abstract]. *Investigative Ophthalmology & Visual Science*, *36*(Suppl.), 1845.

- Shiffrar, M., Li, X., & Lorenceau, J. (1995). Motion integration across differing image features. *Vision Research*, 35, 2137-2146. [[PubMed](#)]
- Shiffrar, M., & Lorenceau, J. (1996). Increased motion linking across edges with decreased luminance contrast, edge width and duration. *Vision Research*, 36, 2061-2067. [[PubMed](#)]
- Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29, 619-626. [[PubMed](#)]
- Stoner, G. R., & Albright, T. D. (1996). The interpretation of visual motion: Evidence for surface segmentation mechanisms. *Vision Research*, 36, 1291-1310. [[PubMed](#)]
- 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. [[PubMed](#)]
- Trueswell, J. C., & Hayhoe, M. M. (1993). Surface segmentation mechanisms and motion perception. *Vision Research*, 33, 313-328. [[PubMed](#)]
- Vallortigara, G., & Bressan, P. (1991). Occlusion and the perception of coherent motion. *Vision Research*, 31, 1967-1978. [[PubMed](#)]
- Wallach, H. (1935). Ueber visuell wahrgenommene bewegungsrichtung. *Psychologische Forschung*, 20, 325-380.
- Watanabe, T. (1997). Velocity decomposition and surface decomposition-reciprocal interactions between motion and form processing. *Vision Research*, 37, 2879-2889. [[PubMed](#)]
- Weiss, Y., Simoncelli, E. P., & Adelson E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5, 598-604. [[PubMed](#)]
- Wuerger, S., Shapley, R., & Rubin, N. (1996). On the visually perceived direction of motion by Hans Wallach: 60 years later. *Perception*, 25, 1317-1367.
- Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 481-543). New York: Academic Press.
- Zaidi, Q., Spehar, B., & Shy, M. (1997). Induced effects of backgrounds and foregrounds in three-dimensional configurations: The role of T junctions. *Perception*, 26, 395-408. [[PubMed](#)]