Perception, 2001, volume 30, pages 905 – 923

DOI:10.1068/p3219

Beyond junctions: nonlocal form constraints on motion interpretation

Josh McDermott¶
Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR, UK; also Department of Brain and Cognitive Science, Massachusetts Institute of Technology, NE20-444H, Cambridge, MA 02139, USA; e-mail: josh@gatsby.ucl.ac.uk

Yair Weiss
Department of Computer Science, University of California at Berkeley, Berkeley, CA 94720, USA

Edward H Adelson
Department of Brain and Cognitive Science, Massachusetts Institute of Technology, NE20-444H, Cambridge, MA 02139, USA; e-mail: adelson@psyche.mit.edu
Received 6 June 2000, in revised form 28 January 2001

Abstract. Because of the aperture problem, local motion measurements must be combined across space. However, not all motions should be combined. Some arise from distinct objects and should be segregated, and some are due to occlusion and should be discounted because they are spurious. Humans have little difficulty ignoring spurious motions at occlusions and correctly integrating object motion, and are evidently making use of form information to do so. There is a large body of theoretical and empirical evidence supporting the importance of form processes involving junctions in the way motion is integrated. To assess the role of more complex form analysis, we manipulated nonlocal form cues that could be varied independently of local junctions. Using variants on diamond and plaid stimuli used in previous studies, we found that manipulations distant from the junctions themselves could cause large changes in motion interpretation. Nonlocal information often overrides the integration decisions that would be expected from local cues. The mechanisms implicated appear to involve surface segmentation, amodal completion, and depth ordering.

1 Introduction

An object's motion cannot be determined from a single measurement on its contour owing to the ambiguity known as the aperture problem (Wallach 1935; Marr and Ullman 1981; Adelson and Movshon 1982). Local measurements must therefore be combined across space. In figure 1a, the ambiguous motion of the contour labeled 1 can be resolved by combining it with another contour's motion, eg by intersection of constraints. Alternatively, one may utilize the unambiguous motion of a 'feature,' such as the corner labeled 2. It has long been recognized that unambiguous features can have powerful effects on motion perception (Wallach 1935; Nakayama and Silverman 1988). However, some features are spurious, such as the T-junction labeled 3 which is due to occlusion (Shimojo et al 1989). The cross-bar of the T is a contour that is 'owned' by the occluding region; the stem of the T is a contour that is occluded by that region. The stem and cross-bar lie in different planes and the T-junction in the image is artifactual, corresponding to no physical feature, and moving with no physical object. In the example of figure 1, the two squares translate to the left and right but the T-junction moves downward. Such spurious motions will distort motion estimates if they are treated as real. Occlusions and the spurious motions that occur with them are common in natural motion sequences, but humans generally discount them correctly. In the local motion domain, however, spurious features are not obviously distinguishable from nonspurious; form information is apparently needed and used in human motion interpretation.

¶ Author to whom correspondence and requests for reprints should be addressed.

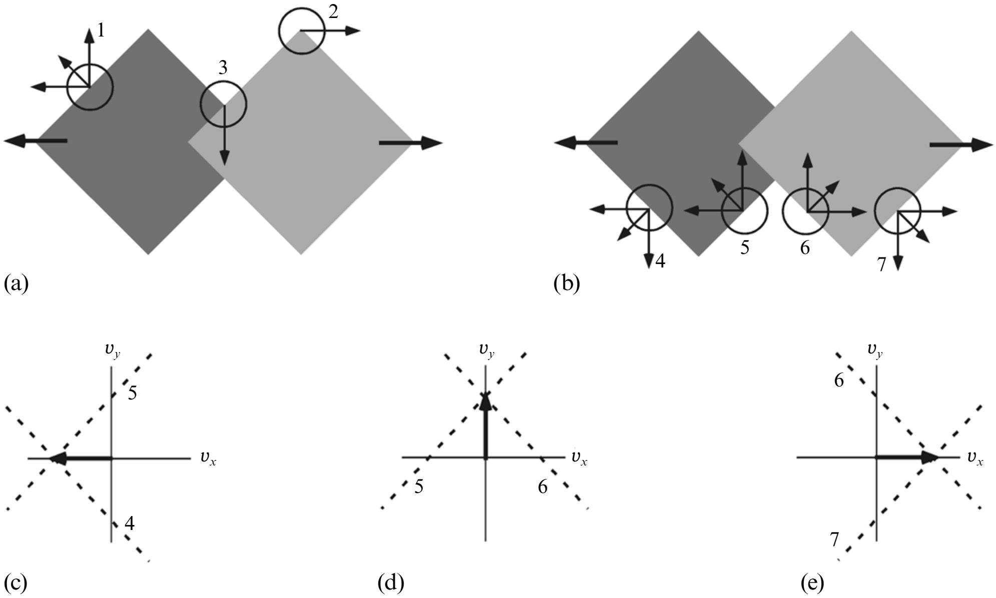
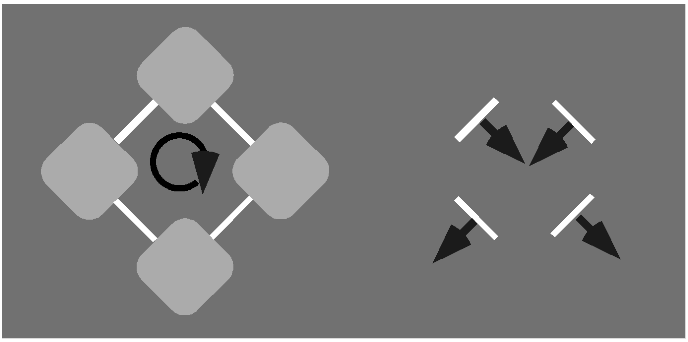


Figure 1. Example illustrating two problems that occur in motion interpretation. In (a) and (b), two squares translate horizontally. The edge motions (eg 1) are ambiguous, while the corner motions (eg 2) are unambiguous. The T-junction motions (eg 3) are also unambiguous, but their motion is spurious 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.

Even if spurious features have been detected, a further problem is imposed by the fact that some of the remaining local motions may arise from distinct objects, and must be segregated rather than combined (Braddick 1993). Consider again our two translating squares. Motion measurements made on two edges of the same object (eg at points 4 and 5, or 6 and 7 of figure 1b) can be combined to yield the correct horizontal motions, as shown in figures 1c and 1e with intersection of constraints. If, however, measurements from different objects are combined, a faulty motion estimate results, as shown in figure 1d. In this case, combining measurements 5 and 6 via intersection of constraints yields an upward motion which is completely wrong—no object in the scene is moving up. Form information may be useful in resolving this dilemma as well. In our example all three pairs of measurements are approximately equidistant, and in the motion domain alone there is no obvious reason to combine 4 and 5 but not 5 and 6. An analysis of spatial form, however, can reveal that 4 and 5 are part of the same object while 6 is not.

There are thus at least two fundamental problems in motion interpretation that appear to necessitate a concurrent analysis of spatial form: (i) some local motions are spurious and should be discounted, and (ii) some motions arise from distinct objects and should be segregated. A large body of research has demonstrated the importance of form processes in determining how motion is interpreted (eg Wallach 1935; Stoner et al 1990; Lorenceau and Shiffrar 1992; Nowlan and Sejnowski 1995). In this paper we investigate the nature of these form processes. We ask whether a simple, local junction analysis suffices, or whether more sophisticated and nonlocal processing is involved.

We began our investigations with the basic diamond stimulus of Lorenceau and Shiffrar (1992; Shiffrar and Lorenceau 1996), shown in figure 2a. An outline diamond translates with a circular trajectory, its corners hidden by occluders. Each of the four line segments, when taken alone, can be considered to move sinusoidally in the direction normal to its orientation. Indeed, the only trackable features are the spurious T-junctions moving in these directions. However, observers generally report seeing the single coherent motion of a diamond rather than the separate motions of the line segments, indicating that the Ts are ignored and the edge motions integrated. The percept is different in the condition shown in figure 2b, where the occluders are invisible due to an accidental match in color with the background. The points of occlusion become line terminators rather than Ts, and the stimulus breaks up into separate motions. The coherence seems to depend on the presence of the T-junctions, which provide a local cue to occlusion.



(a) (b)

Figure 2. The basic diamond stimulus, generated by moving a diamond in a circle behind occluders, which can either be visible (a) or invisible (b).

It is thus reasonable to ask whether these and other motion phenomena can be explained with relatively simple, local form analysis. For most published phenomena the answer is yes, and current motion models reflect this. A typical model involves (i) extracting local motions; (ii) suppressing the motions of spurious junctions; (iii) combining the remaining local motions using 'common fate', ie grouping local motions that are mutually consistent. Such mechanisms were used successfully in the model of Nowlan and Sejnowski (1995). Related models (Liden and Pack 1999; Weiss and Adelson 2000) can explain most existing data (Anstis 1990; Stoner et al 1990; Vallortigara and Bressan 1991; Lorenceau and Shiffrar 1992; Trueswell and Hayhoe 1993; Shiffrar et al 1995; Shiffrar and Lorenceau 1996; Castet and Wuerger 1997; Liden and Mingolla 1998; Stoner and Albright 1998; Rubin 2001). We will henceforth refer to this general architecture as the SJI model, for Suppress Junctions and Integrate. Note that form constraints of arbitrary complexity could conceivably be incorporated into either of the two latter stages. However, fairly minimal junction-based constraints, located at the stage where spurious motions are discounted, have seemed sufficient for prior models. In this paper we put this notion to a sharper test. We find that junction analysis alone does not suffice; nonlocal form constraints can have a dominant influence.

There is reason to suspect that the significance of a T-junction for motion perception might not be locally determined. For although many T-junctions, such as that of figure 3a, are due to occlusion, some, like that of figure 3b, are not. Both the stem and cross-bar contours of figure 3b are owned by the plane to the left of the cross-bar, and if the object moves, the T-junction moves with it. So perhaps motion integration should treat some junctions differently from others, depending on the context.

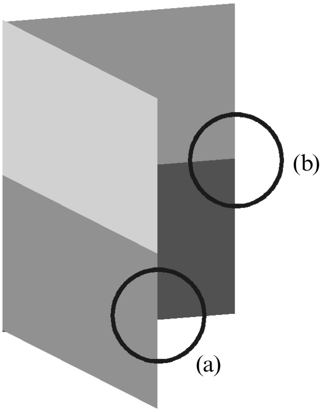


Figure 3. Not all T-junctions are due to occlusion. The T-junction at (a) is due to occlusion, but that at (b) is not. Nonlocal mechanisms evidently influence the occlusion interpretation of such junctions.

To examine this issue, our strategy was to hold the moving junctions in our displays constant while manipulating additional stimulus properties of perceptual relevance. These additional stimulus properties contain more information than do the junctions alone, but accordingly necessitate more complicated computations. We will often refer to our manipulations as 'nonlocal'; by this we simply mean that they alter aspects of the stimuli other than the junctions that abut the regions of image motion. In most of our manipulations the junctions are held constant in a small region of the visual field (eg 0.5 deg in diameter at 1.5 deg eccentricity). The justification for declaring this region to be 'the junction' is mainly empirical: previous studies pertaining to junctions have manipulated structures at a comparable scale (Stoner et al 1990; Zaidi et al 1997; Rubin 2001), and some at much smaller scales (eg Anderson 1997). Such studies have found powerful effects of junctions even at relatively small scales, so one would expect our junctions to dictate the motion of our displays if they are the main form constraint on motion interpretation.

Demonstrations of many of the moving stimuli used in this work can be viewed on the *Perception* website at http://www.perceptionweb.com/perc0801/mcderm.html

2 Experiment 1

As a first test, we constructed the stimuli shown in figures 4a and 4b. The moving junctions are identical, but the global configurations are quite different. Both stimuli contain the same four moving bars, and the T-junctions at the bar endpoints (indicated by the circles) are locally identical in terms of grey levels and geometry. However, the larger context of the bars and their junctions generates very different impressions of occlusion and depth ordering. Figure 4a is seen as a diamond with hidden corners; figure 4b is typically seen as a set of rectangles over a set of discs, with lines running down the centers of the rectangles. Thus, the diamond contours look much less occluded in figure 4b than in figure 4a. The question is whether this change to the context of the junctions will influence perceived motion.

When the diamonds are set in motion, the resulting percepts are quite different. We quantified the effect in an experiment in which subjects were presented with short clips of the stimuli and asked to judge their coherence; full details appear in the appendix. Subjects pressed buttons following each trial to indicate incoherent, partially coherent, or totally coherent motion percepts. Their responses were normalized to yield a coherence index between 0 and 1. An index of 0 corresponds to a percept of completely incoherent motion on every single trial; 1 indicates consistently coherent motion. The effects are perceptually strong, and the qualitative phenomena have been confirmed in many additional observers who have viewed our stimuli in the laboratory and at conferences.

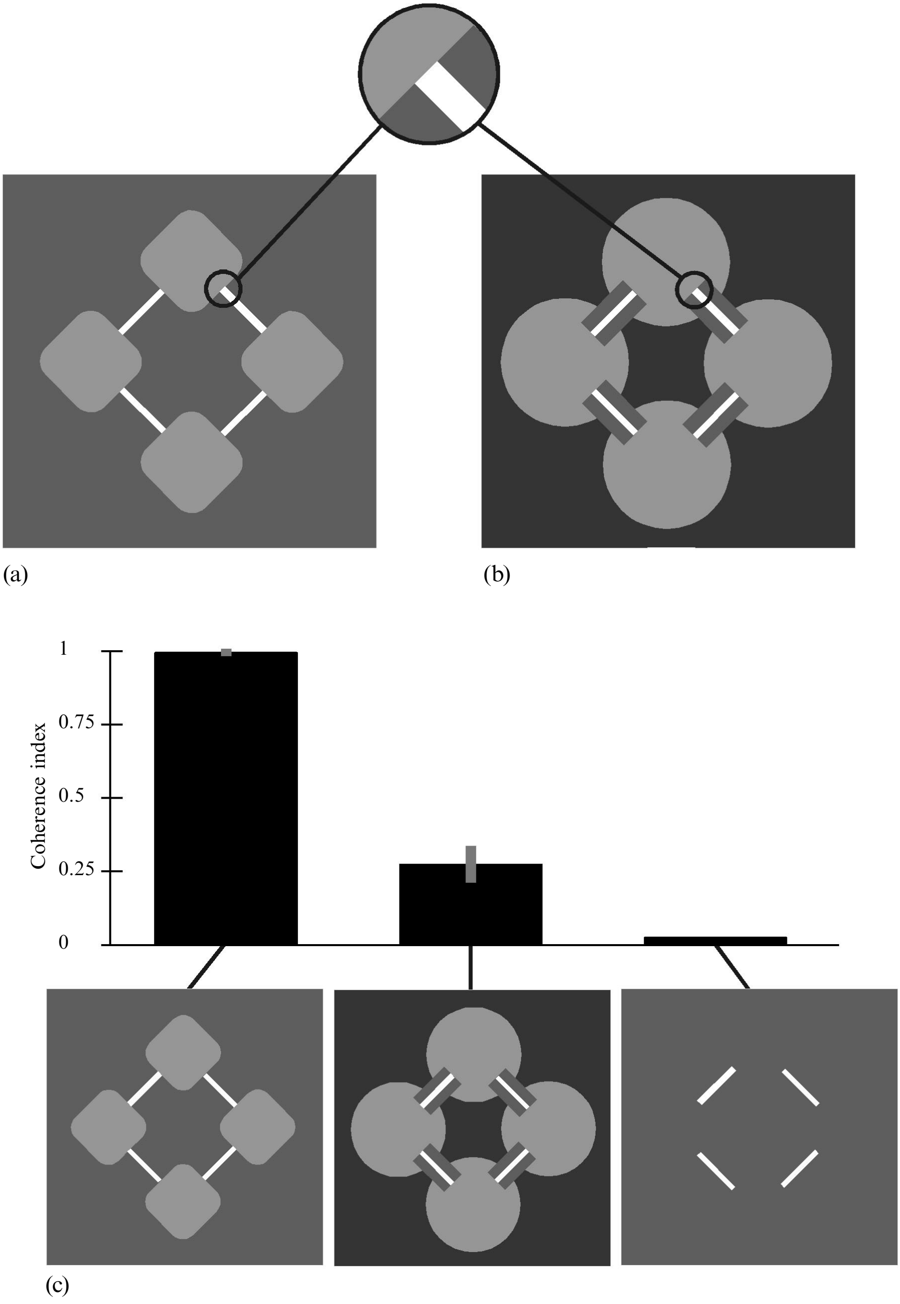


Figure 4. Stimuli and results of experiment 1. (a) and (b) Experimental stimuli that are identical in the local vicinity of the diamond contours but which differ globally in the extent to which they support occlusion. (c) Observed coherence levels for each stimulus, for six naïve subjects. Error bars in this and all other graphs denote standard errors.

Figure 4c shows measurements of coherence. The occluded diamond stimulus (figure 4a) is seen as highly coherent. The modified stimulus of figure 4b produces much poorer coherence. This is consistent with the change in apparent depth ordering and border ownership in figure 4b due to the different contexts. The T-junctions in figure 4b are less likely to be due to occlusion, and the moving bars are less likely to be part of the same object, so the decrease in coherence is sensible. Note that the T-junctions in figure 4b still appear to exert some influence, as this stimulus is certainly more coherent than when no junctions are present at all (figure 4c, far right). However, the context of the junctions clearly has a large effect—motion interpretation is far from dominated by the T-junctions at the bar endpoints. Further experiments reported elsewhere indicate that the convexity of the occluding contour (eg whether the contour abutting the moving lines curves towards or away from them) is an important component of this context (McDermott et al 2000), consistent with evidence that convexity is an important cue to border ownership (Arnheim 1954).

The SJI model outlined in the introduction clearly must be modified to account for these nonlocal form influences. Perhaps most obviously, form influences could be incorporated at the stage where local motions are discounted: T-junctions such as those in figure 4b could be discounted to a lesser extent in the presence of decreased evidence for occlusion. Alternatively, the contextual influences could bias the extent to which different local motions are integrated: in this case, the four bars could be integrated to a lesser extent because there is evidence that they lie on disconnected surfaces and are therefore unlikely to be part of the same object. In either case, the junctions have less influence, and in this paper we will not attempt to distinguish between these and other possible mechanisms for this decreased influence. Instead, we focus on exploring other sorts of nonlocal form constraints.

3 Experiment 2

Given that a goal of motion integration is to combine only those motions that are due to the same object, it is conceivable that the coherence of the diamond contours might also be related to their amodal completion behind occluders (Kanizsa 1981; Michotte et al 1964; Nakayama et al 1995). Several authors (Wallach 1935; Shimojo et al 1989; Alais et al 1998) have suggested that amodal completion might be involved in motion perception. To test the role of completion in our displays, we compared the thick occluders of figure 5a with the thin ones of figure 5b, which are simply L-shaped lines. Note that in each case the diamond contours end with T-junctions. However, the two sets of occluders differ in their support of amodal completion. We added dim lines in the background to ensure that the entire background plane would be seen as a single surface, leaving no space for the diamond contours to complete behind the thin occluders (other background patterns that continue behind the occluders work as well).

Although the T-junctions are locally similar in the two cases, their effect is quite different. The stimulus of figure 5b usually broke up into oscillating line segments, almost as frequently as when no T-junctions were present at all (figure 5c). Notably, we were able to restore coherence by closing the L contours as shown in figure 5d, so that the Ls were seen as the boundary of an extended occluder, making completion possible once more. The junctions of figures 5b and 5d were the same, but their effects on coherence were dramatically different.

Once again, a local junction analysis would not predict the difference in perceived motion. Evidently motion interpretation is aided by a process that senses the presence of an extended occluding surface, perhaps via the closure of the occluding contour. Closure is not all that matters, though. As shown in figure 5e, continuing the background lines inside the occluder outlines reduced coherence, presumably because there was less evidence for a continuous occluding surface. Similarly, figure 5f shows that removing the background lines from the displays increased coherence for the thin L-shaped occluders, again presumably because without the lines they are more plausibly

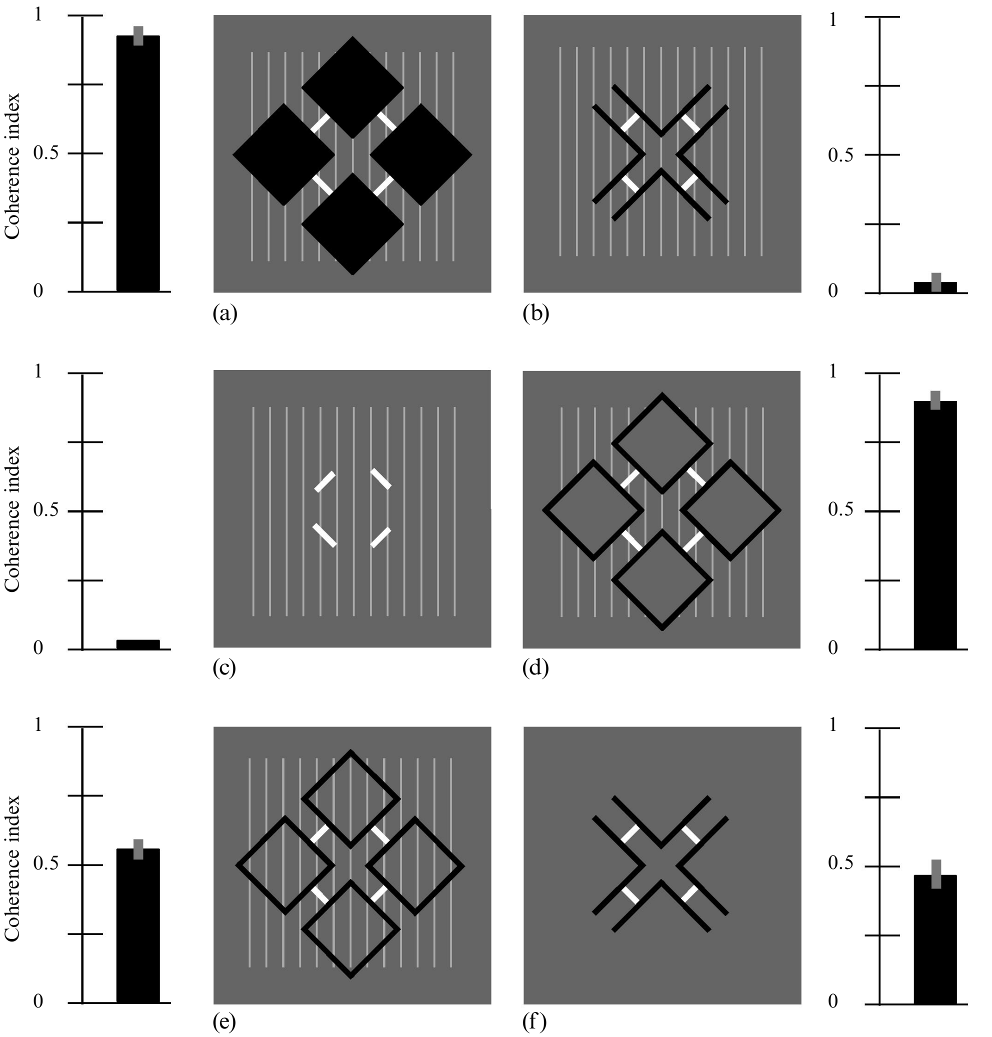


Figure 5. Stimuli and results of experiment 2. (a) Diamond with thick occluders, supporting amodal completion. (b) Diamond with thin occluders, preventing amodal completion. (c) Diamond contours without occluders or T-junctions. (d) Diamond with outline occluders, restoring amodal completion and coherence. (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 naïve subjects.

interpreted as the borders of extended surfaces. However, closure produces yet higher coherence, and gradually closing the occluding contour gradually boosts coherence, as shown in figure 6.

Motion perception again seems to be influenced by fairly complex aspects of spatial form. It appears that integration depends in part on whether there is 'room' for amodal completion, ie whether there is a visible occluder behind which the completion can occur.

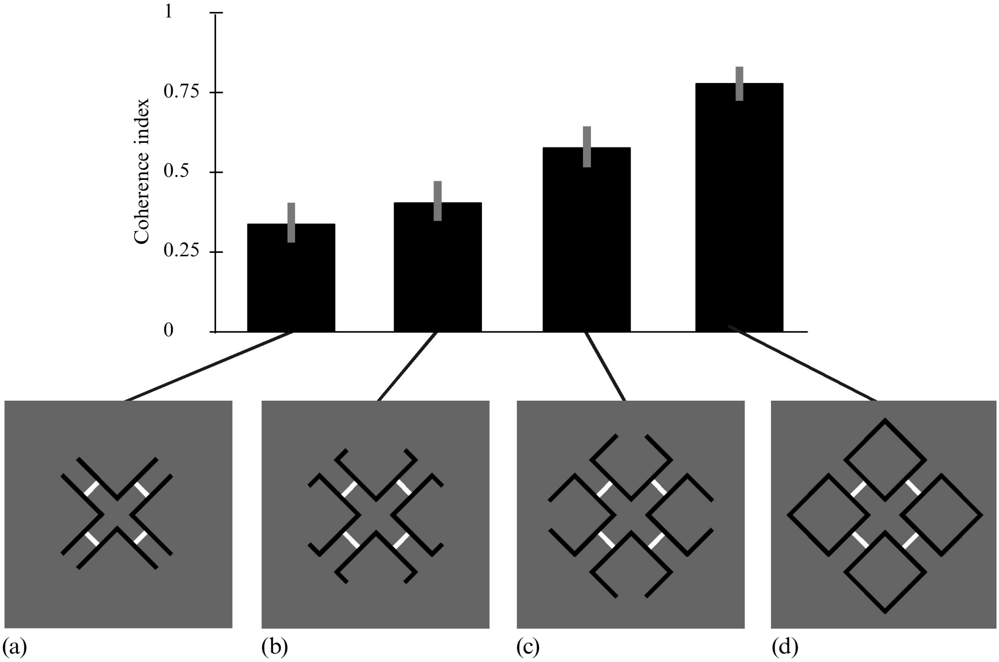


Figure 6. Stimuli and results of four separate conditions from experiment 2. (a) Diamond with L-shaped occluders, preventing amodal completion. (b) – (d) Increasing closure increases coherence. The results are for five naïve subjects.

4 Experiment 3

If amodal completion is really at issue, then the tendency to cohere should depend on the thickness of the occluders in an orderly way. We varied the thickness of the L-shaped occluders to see how coherence would be affected. Figures 7a – 7c show example stimuli with three different occluder thicknesses. (The dashed lines were not visible in the stimulus; they merely show the geometry for the amodal completion.) Figure 7d shows how coherence varied as a function of occluder thickness. The second panel, figure 7b, shows the point where the occluders were just thick enough to completely hide the virtual corners. Coherence was quite low when the occluders were thin, and increased as they became thicker. Coherence did not start to level off until after the critical point of 7b, consistent with what one might expect if amodal completion were involved.

The point of 'full coverage' shown in figure 7b can be shifted by moving the occluders further apart, so that a higher proportion of the diamond is visible. This is shown in figure 8a (compare to figure 8b). This manipulation has the effect of making the diamond more coherent overall when properly occluded, which makes it difficult to avoid ceiling effects in many subjects. However, for the five subjects with whom we were able to obtain subceiling data, we observed the expected pattern of results, shown in figure 8c (solid line): coherence was again quite low when the occluders were thin, increased as they became thicker, and leveled off after the critical point. We then repeated the experiment with these subjects using the original configuration of closely spaced occluders—shown in figure 8b at the point of critical thickness. As shown by the dashed line in figure 8c, the point of asymptote shifted back to its original location. The results are consistent with the notion that the motion system utilizes fairly precise knowledge of amodal completion in determining whether to integrate local motions. We note also that the side effect of coherence increasing as

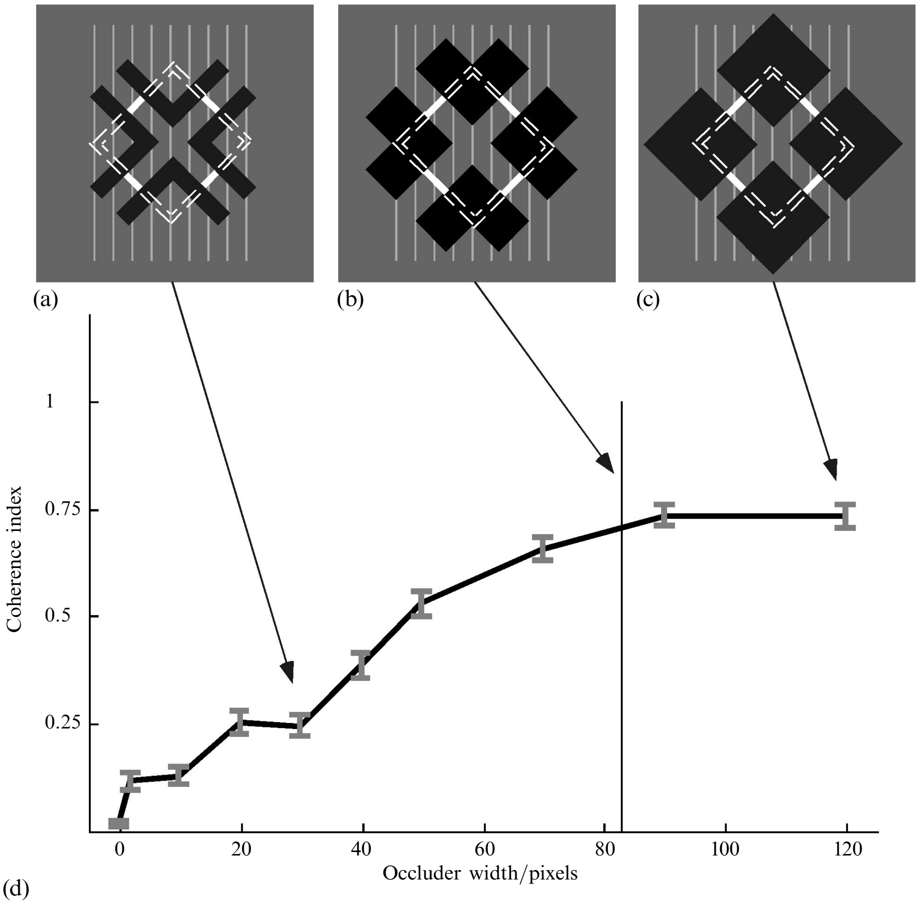


Figure 7. Stimuli and results of experiment 3, part 1. (a) – (c) Three sample stimuli with the path of contour completion schematically depicted with dashed lines (not included in the stimulus). (d) Observed coherence levels as a function of occluder width, for twelve naïve subjects. The thin vertical line indicates the point at which the occluders are large enough to cover all of the amodal contour.

more of the diamond becomes visible is also consistent with a completion-related constraint on motion integration. There is reason to think that completion should be stronger in situations where a lower proportion of the contour has to be completed (Shipley and Kellman 1992), consistent with our observation that coherence is similarly strengthened under such conditions.

The reader may notice that the maximum coherence level in these experiments is lower than in the previous experiment, despite what in some cases are geometrically identical displays (compare figures 7 and 8 with figure 5). This is simply due to the fact that we adjusted the contrast of the moving bars in the present experiments to avoid a ceiling effect that would have obscured the asymptotic behavior. The degree of coherence for a given display can be increased or decreased to a limited extent by lowering or raising the contrast of the bars, an effect documented elsewhere (Adelson and Movshon 1983; Lorenceau and Shiffrar 1992; McDermott et al 2001). This effect of contrast underscores the fact that the influence of amodal completion on motion interpretation does not take the form of necessary and sufficient conditions that must be satisfied for integration to occur. Substantial degrees of coherence can occur for very low-contrast moving bars even when the occluders do not provide room for

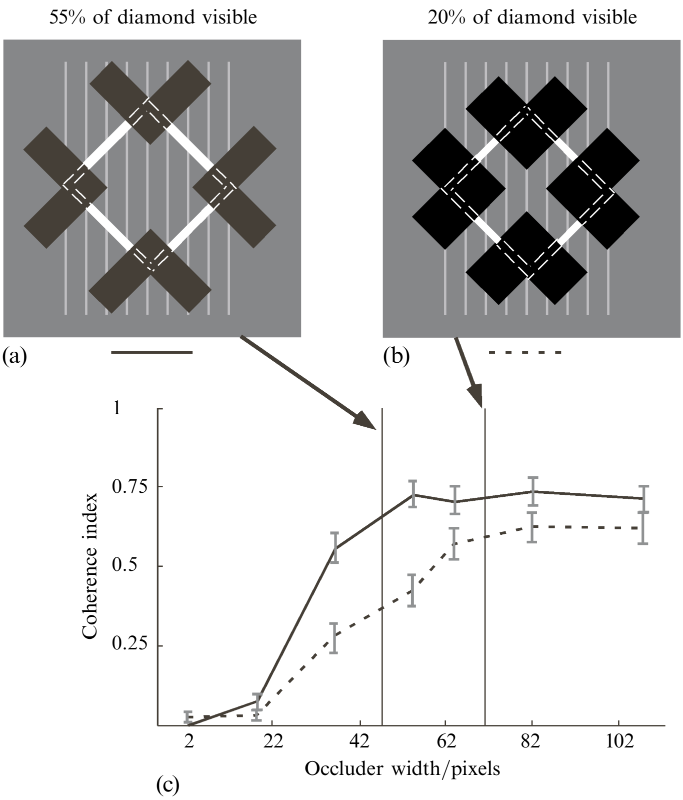


Figure 8. Stimuli and results of experiment 3, part 2. (a) – (b) Example stimuli illustrating the different critical points of the two configurations used. (c) Observed coherence levels for five naïve subjects as a function of occluder width for each configuration. The thick solid line plots data for configuration (a), the dashed line those for configuration (b). Thin solid lines indicate the critical point for each configuration at which the occluders are large enough to cover all of the amodal contour.

amodal completion. Conversely, displays that optimize conditions for completion can remain less than fully coherent if the moving bars are very high contrast. Completion, like the other form influences explored in this paper, imposes a constraint on motion interpretation, not a necessary condition.

5 Experiment 4

Another way to manipulate amodal completion is to change the 'relatability' of contours, in the sense of Kellman and Shipley (1991). Consider the stimuli of figure 9, in which line segments are seen through apertures. In figure 9a, the lines can readily be continued behind the apertures to form a single square. In figure 9b, in contrast, the vertical lines extend too far to join the horizontal ones, so a simple completion is impossible; these contours are not relatable. We found a large difference in coherence between the two cases. The relatable stimuli almost always cohered, while the nonrelatable ones almost never cohered. Note that proximity biases on motion integration (Nakayama and Silverman 1988) would suggest that figure 9b should, if anything, cohere more than figure 9a, since the moving contours are physically closer in the image. However, any proximity influence is evidently swamped by the effect of relatability, presumably owing to the importance of completion.

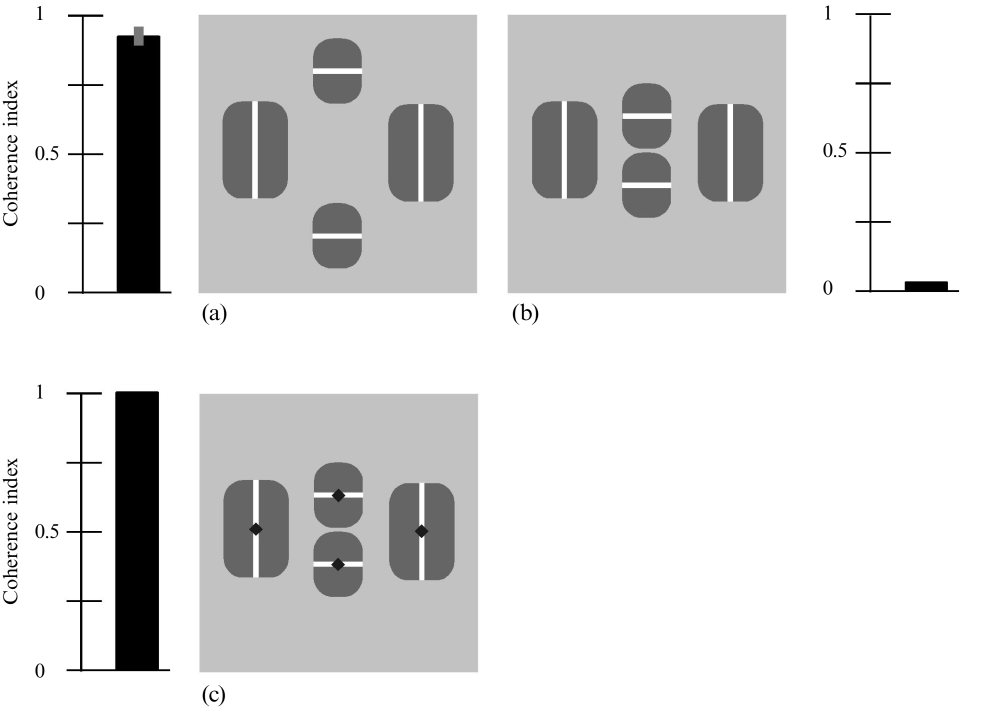


Figure 9. Stimuli and results of experiment 4. (a) Relatable configuration, which generates high coherence. (b) Nonrelatable configuration, which never coheres. (c) Nonrelatable configuration with dots superimposed on the contours. The dots move in the direction of coherent motion, and with their addition the stimulus coheres. The results are for six naïve subjects.

One might object that it is simply impossible to see the nonrelatable configuration as a single, coherently moving object. This is not the case. As a control condition, we placed dots on the nonrelatable moving contours, and moved the dots with the appropriate circular trajectory, as shown in figure 9c. Not surprisingly, each line's motion was captured by its dot, and thus each appeared to move with a circular trajectory. In addition, all four lines appeared to be part of a single coherent object, despite their nonrelatability. The nonrelatability thus does not prevent coherence per se; when coherence does not depend on integrating motion constraints across contours, as when the dots are present, it can occur perfectly well even for nonrelatable configurations. In related work, Lorenceau and Zago (1999) have described experiments on moving gratings seen through apertures. They find that integration is strongest when the apertured gratings form L-shaped configurations, and weaker when they form implicit T-shapes. We suggest this may be the result of the same completion-related process we observe in our experiments with lines.

The SJI model sketched out earlier must again be modified to account for the previous three experiments. In this case, the most obvious way to incorporate the various completion-related constraints is as biases on integration: two local motions could be integrated with higher probability if they are generated by contours that can amodally complete. This necessitates modeling the relevant form constraints on amodal completion, however, and at present this is a nontrivial task. Our experiments suggest that, at a minimum, contour relatability, closure of occluding contours, surface solidity, and the position of potentially completed contours are all factors that affect amodal completion and motion integration. It seems clear that some fairly complex form computations are involved.

6 Experiment 5

In an additional experiment on nonlocal form cues, we used two pairs of white lines, forming a minimal 'plaid' stimulus as shown in figure 10a. The intersection points were the same white color as the lines. The plaid translated left and right in an oscillatory manner. The lines could appear to move as a coherent plaid, in the horizontal direction, or they could appear as two separate line pairs, each pair sliding along the frame in a diagonal direction normal to the line orientation. As expected, in the configuration of figure 10a the lines generally appeared to move coherently, as shown by the arrows.

Next, we added a static grid of dark bars in the background, as shown in figure 10b. The plaid's motion continued to appear coherent, moving horizontally across the grid. We then moved the grid to lie in front of the lines, so that it obscured the intersection points, as shown in figure 10c. In this stimulus there were no visible intersections moving in the coherent direction. Nonetheless, the plaid cohered as before. The many line segments glimpsed through the latticework appeared to move as part of a unified object, in spite of the fact that there were no visible features to track. Finally, we used the stimulus shown in figure 10d, in which one set of lines was in front of the grid and the other was behind. Coherence broke down, and the line pairs slid independently in their normal directions.

Note that in terms of local features, figures 10c and 10d have much in common. In neither case do the lines form visible intersections. In both cases, all the visible line segments are diagonal. In both cases, a coherent (or incoherent) interpretation is entirely consistent with the image data (indeed, adding horizontally moving dots to the stimulus of figure 10d resulted in predominantly horizontal motion being seen). However, the coherent interpretation would require that the lines be linked across different depth planes, through the intermediate grating, and apparently this does not occur.

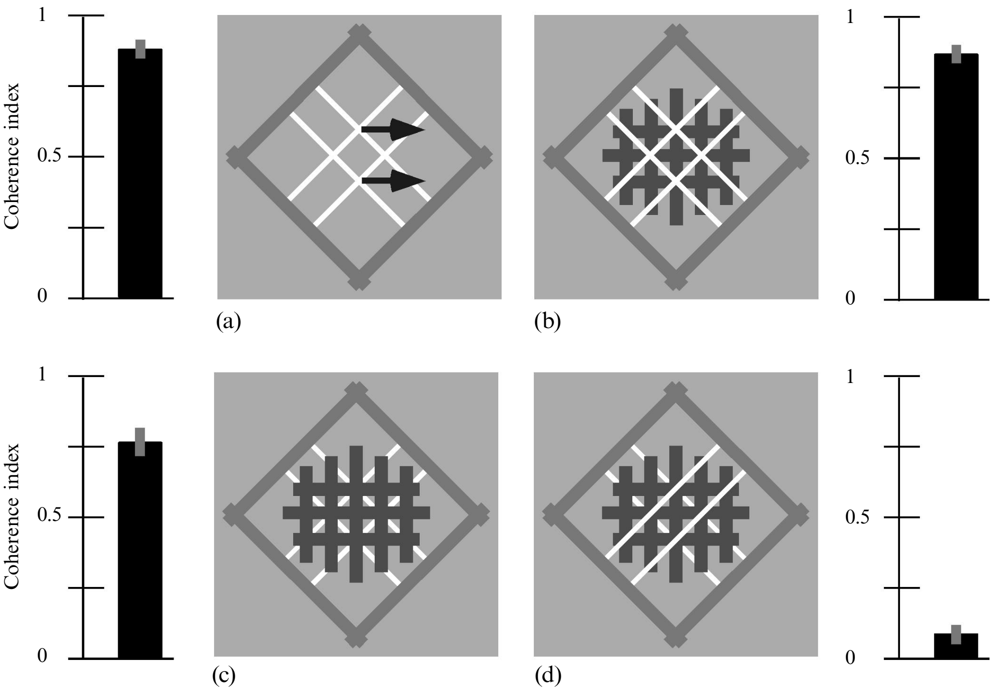


Figure 10. Stimuli and results of experiment 5. (a) Minimal plaid stimulus; the white lines move horizontally within the occluding aperture. (b) Plaid with static grating added to background. (c) Plaid with static grating added to foreground. (d) Plaid with static grating drawn between moving gratings, breaking coherence. The results are for six naïve subjects.

The sense that the lines are on opposite sides of the static grating appears to require some fairly complex form analysis; an analysis of isolated junctions does not distinguish the different depth orderings. Stoner et al (1990) and Stoner and Albright (1996, 1998), among others (eg Vallortigara and Bressan 1991; Bressan et al 1993; Trueswell and Hayhoe 1993), have extensively demonstrated the importance of the junctions at plaid intersections on plaid coherence. Junctions that are consistent with transparency or occlusion produce less coherence than those that are not. Our experiments complement these results by showing that relatively complex form constraints can also affect plaid coherence. In our stimuli, the intersections are occluded, and thus cannot be directly responsible for the effects. Low-level models (eg Nowlan and Sejnowski 1995) that explain many other plaid results (eg Stoner et al 1990) by discounting the intersection junctions would need to be considerably augmented to account for this experiment. The most obvious way to model this result is via biases at the integration stage that favor integrating the motions from contours that are connected in depth.

7 Discussion

Although there has been a great deal of research on the influence of form and stereo cues on motion perception, most previous work has emphasized the importance of relatively simple and local constraints. The work of Stoner et al (1990) and Stoner and Albright (1992, 1996, 1998) on the effects of transparency and occlusion on plaids, for instance, has demonstrated that the junctions formed at plaid intersections can often exert a strong influence on motion integration; junctions consistent with transparency or occlusion tend to reduce coherence. Furthermore, obscuring the local junctions can abolish the effects of occlusion on coherence (Stoner and Albright 1998), suggesting that the effect may be driven by a simple, local process. Other work on plaids has generally confirmed the importance of junction categories and transparency (Trueswell and Hayhoe 1993; Lindsey and Todd 1996). Additional studies concerning the influence of occlusion on plaid coherence (Vallortigara and Bressan 1991; Bressan et al 1993) have also supported the importance of local junctions, again showing that obscuring the junctions tends to eliminate the occlusion-related effects. The results of these studies are consistent with the simple theory that motions are suppressed at junctions indicative of occlusion and transparency.

Many other studies with diamond (eg Lorenceau and Shiffrar 1992), barberpole (Liden and Mingolla 1998), and assorted other stimuli (Anstis 1990; Shiffrar et al 1995; Castet and Wuerger 1997; Rubin 2001) are similarly consistent with simple and local form processes. There have also been numerous examples of occlusion-related stereo influences on motion perception (Adelson and Movshon 1984; Shimojo et al 1989; Anderson 1999; Castet et al 1999). These too are largely consistent with processes based on local cues. Indeed, the dependence of these phenomena on half-occluded regions (Anderson 1999; Castet et al 1999) underscores the importance of local, low-level processes.

We suggest, however, that such local constraints can be substantially modulated by the larger context in which they occur. The computational framework of suppressing local motions at appropriate junctions and then integrating the remaining compatible motions, which we have referred to as the SJI model (for Suppress Junctions and Integrate) (Nowlan and Sejnowski 1995; Liden and Pack 1999) would have to be significantly extended to account for our phenomena. It would appear that both stages, suppression and integration, require more sophisticated form constraints. Although there are a few relevant results in physiology (Stoner and Albright 1992; Duncan et al 2000), relatively little is known at a mechanistic level, and we will not speculate on how our form constraints might be implemented in neural circuitry.

In our first experiment, we found that the contours of the diamond stimulus were far less coherent when placed in a context that made them appear less occluded, despite the presence of identical T-junctions at the terminators. In our second and third experiments, we found that the diamond contours could be rendered incoherent by changing the shape of the occluders so as to make amodal completion unlikely. Moreover, occluders whose extent was indicated only by outlines were enough to support completion and coherence, implicating a nontrivial process to segment occluding surfaces. Our results suggest this process is sensitive to contour closure as well as to various cues that convey surface solidity. Evidently such a process determines the extent to which there is room beneath occluding surfaces for contours to complete, and this biases integration accordingly. Experiment 4 showed further that altering relatability, another nonlocal stimulus property, can have drastic effects on the coherence of the diamond, again suggestive of contour-completion processes. And in our final experiment, we found that plaid stimuli were less coherent when a static grating was placed between the component gratings of the plaid, forcing them into different depth planes. This effect cannot be due to junctions at the plaid intersections, because the intersection points were always hidden. In all, our phenomena seem to necessitate a variety of form processes that go beyond junction analysis.

As we have discussed, in most of our manipulations the junctions abutting moving contours were held constant in a small region of the visual field while other stimulus properties were varied. What makes this region 'the junction'? Justification can be found in many previous studies that have shown influences of junction-like structures at scales comparable to ours (eg Stoner et al 1990; Anderson 1997; Zaidi et al 1999; Rubin 2001). It would thus be reasonable to expect the junction regions that we held constant to dominate motion perception if they are of great importance. Contrary to this notion, our effects demonstrate that motion perception depends on a number of nontrivial stimulus properties that can be varied independently of the junctions at moving contours.

To aid discussion, we have often spoken of junctions as providing 'local' constraints and referred to the various properties manipulated outside the junctions as 'nonlocal'. These terms are admittedly ill-defined and perhaps overused in perception research, but are nonetheless useful shorthand for the two sorts of constraints being discussed. That is all they are, though—'local' and 'nonlocal' do not denote specific scales of analysis. Indeed, the locality per se of the form constraints at work is not of primary interest. The significance of the nonlocal constraints we have described is rather that they appear to necessitate more complex computations. The nonlocal mechanisms are not merely scaled up versions of the 'local' junction detectors we believe to also exist; they evidently conduct some fairly complicated form computations, regardless of how spatially extended they are. The important point is thus not so much the existence of nonlocal effects as it is the suggestion of their sophistication.

Our phenomena are not the first demonstration of nonlocal form processes affecting motion perception. Most previous such findings involve the effects of illusory contours on motion perception. Wallach (1935) was the first to describe such effects, and recently several authors have replicated and extended his findings (Anderson and Sinha 1997; McDermott et al 1997; Liden and Mingolla 1998; Tommasi and Vallortigara 1999). In these phenomena, illusory contours can alter perceived motion in much the same way that a visible occluding contour can. Such phenomena may involve 'virtual' junctions formed at the intersections of real and illusory contours. In contrast, our research has yielded a number of form constraints that go well beyond junctions, whether real or virtual. In addition to their apparent sophistication, our effects are notable for typically being quite strong, suggesting that constraints such as those we have discussed play an important role in motion interpretation. As mentioned earlier, a related study has been conducted by Lorenceau and Zago (1999), who found that the spatial configuration of grating patches had a strong effect on motion integration. We again suggest that their phenomena, like ours, are a function of relatability and amodal completion. It is also worth noting that the effects of spatial frequency, duty cycle, and contrast on plaids (Adelson and Movshon 1982; Vallortigara and Bressan 1991; Bressan et al 1993; Trueswell and Hayhoe 1993; Stoner and Albright 1996, 1998) may be related to depth segregation, consistent with our findings, although in many cases the cited effects may be driven by changes to the junctions at the plaid intersections.

As noted throughout this paper, there are two conceptually distinct problems in motion interpretation that seem to necessitate form information: (i) spurious motions must be discounted, and (ii) motions generated by distinct objects must be segregated. The structure of the simple SJI model discussed throughout the paper happens to roughly parallel these two problems, and it is tempting to assume that they might correspond to distinct stages of human motion interpretation. However, this need not be the case. It might be that the two problems are not solved independently, in which case it would be inappropriate to speak of two separate stages of analysis. We therefore do not mean to advocate the simple two-stage model; it merely serves as a helpful framework in which to view these and other experiments. That said, it is useful to distinguish between the two goals of suppressing spurious moving features and restricting integration to within objects, because they remind us of what the visual system might be trying to do. The various form constraints at work in our phenomena apparently help the visual system to more accurately meet these goals and hence interpret moving images correctly. Of course, these constraints come at the cost of added computational complexity, but evidently it is a worthwhile trade-off.

8 Conclusion

By presenting ambiguous motion stimuli one can study the influence of various constraints on motion perception. A scene containing multiple overlapping objects will contain ambiguous and spurious motion signals, which pose a challenge for the visual system. Form information is critical in resolving these ambiguities, but motion processing is often thought to use rather limited information about form. Indeed, a great many motion phenomena can be explained by models in which the form processing goes little beyond junction analysis, and in which the integration is based on the grouping of consistent local motions. However, we have found evidence that the visual system employs considerably more sophisticated form constraints in interpreting motion. The effects we describe indicate the importance of configural cues to border ownership, amodal completion, and depth segregation in motion coherence. The extent to which a T-junction is discounted depends on the strength of the evidence that it is caused by occlusion, and this seems to depend on the context in which the junction appears. Even when border ownership is correctly specified, though, additional conditions involving amodal completion appear to constrain motion perception; disconnected moving parts are more likely to be seen to move coherently when there is evidence that they can belong to a single amodally completed figure. And when the depth ordering of gratings and occluders is changed, the motion coherence is changed, even when there are no visible junctions at the cross-points. All of these phenomena indicate the importance of relatively sophisticated form mechanisms in determining the way that motion information is extracted and combined.

Acknowledgements. This work was funded by NIH grants EY11005-04 and EY12690-02 to E Adelson. J McDermott was supported by the Gatsby Charitable Foundation and a Marshall Scholarship. The authors are grateful to Bart Anderson and Ken Nakayama for helpful discussions, and to the reviewers for helpful comments on the manuscript.

References

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

Adelson E H, Movshon J A, 1983 "The perception of coherent motion in two-dimensional patterns" *ACM Siggraph and Sigart Interdisciplinary Workshop on Motion: Representation and Perception* (New York: Association for Computing Machinery) pp 11 – 16

Adelson E H, Movshon J A, 1984 "Binocular disparity and the computation of two-dimensional motion" *Journal of the Optical Society of America A* **1** 1266

Alais D, Smagt M J van der, Berg A V van den, Grind W A van de, 1998 "Local and global factors affecting the coherence motion of gratings presented in multiple apertures" *Vision Research* **38** 1581 – 1591

Anderson B L, 1997 "A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions" *Perception* **26** 419 – 454

Anderson B L, 1999 "Stereoscopic occlusion and the aperture problem for motion: a new solution" *Vision Research* **39** 1273 – 1284

Anderson B L, Sinha P, 1997 "Reciprocal interactions between occlusion and motion computations" *Proceedings of the National Academy of Sciences of the USA* **94** 3477 – 3480

Anstis S, 1990 "Imperceptible intersections: The chopstick illusion", in *AI and the Eye* Eds A Blake, T Trościanko (New York: John Wiley)

Arnheim R, 1954 *Art and Visual Perception* (Berkeley, CA: University of California Press)

Braddick O, 1993 "Segmentation versus integration in visual motion processing" *Trends in Neuroscience* **16** 263 – 268

Bressan P, Ganis G, Vallortigara G, 1993 "The role of depth stratification in the solution of the aperture problem" *Perception* **22** 215 – 228

Castet E, Wuerger S, 1997 "Perception of moving lines: interactions between local perpendicular signals and 2D motion signals" *Vision Research* **37** 705 – 720

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

Duncan R O, Albright T D, Stoner G R, 2000 "Occlusion and the interpretation of visual motion: Perceptual and neuronal effects of context" *Journal of Neuroscience* **20** 5885 – 5897

Kanizsa G, 1981 *Organization in Vision: Essays on Gestalt Perception* (New York: Praeger)

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

Liden L, Mingolla E, 1998 "Monocular occlusion cues alter the influence of terminator motion in the barber pole phenomenon" *Vision Research* **38** 383 – 3898

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

Lindsey D T, Todd J T, 1996 "On the relative contributions of motion energy and transparency to the perception of moving plaids" *Vision Research* **36** 207 – 222

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, Weiss Y, Adelson E H, 1997 "Surface perception and motion integration" *Investigative Ophthalmology & Visual Science* **38**(4) S237

McDermott J, Weiss Y, Adelson E H, 2000 "The computation of occlusion for motion perception" *Investigative Ophthalmology & Visual Science* **41**(4) S544

McDermott J, Weiss Y, Adelson E H, 2001 "The role of occlusion in motion interpretation: Insights from occluded crosses" (submitted)

Marr D, Ullman S, 1981 "Directional selectivity and its use in early visual processing" *Proceedings of the Royal Society of London, Series B* **211** 151 – 180

Michotte A, Thinès G, Crabbé G, 1964 *Les Compléments Amodaux des Structures Perceptives* (Louvain: Studia Psychologica, Publications Universitaires de Louvain) [translation into English in *A Michotte's Experimental Phenomenology of Perception* Eds G Thinès, A Costall, G Butterworth (1991; Hillsdale, NJ: Lawrence Erlbaum Associates)]

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

Nakayama K. He Z, Shimojo S, 1995 "Visual surface representation: a critical link between lower-level and higher-level vision", in *Visual Cognition* Eds S M Kosslyn, D N Osherson (Cambridge, MA: MIT Press) pp 1 – 70

Nowlan S, Sejnowski T, 1995 "A selection model for motion processing in area MT of primates" *Journal of Neuroscience* **15** 1195–1214
Rubin N, 2001 "The role of junctions in surface completion and contour matching" *Perception* **30** 339–366
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 & Psychophysics* **52** 97–106
Stoner G R, Albright T D, 1992 "Neural correlates of perceptual motion coherence" *Nature* **358** 412–414
Stoner G R, Albright T D, 1996 "The interpretation of visual motion: Evidence for surface segmentation mechanisms" *Vision Research* **36** 1291–1310
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
Stoner G R, Albright T D, Ramachandran V S, 1990 "Transparency and coherence in human motion perception" *Nature* **344** 153–155
Tommasi L, Vallortigara G, 1999 "Figure–ground segregation modulates perceived direction of ambiguous moving gratings and plaids" *Vision Research* **39** 777–787
Trueswell J C, Hayhoe M M, 1993 "Surface segmentation mechanisms and motion perception" *Vision Research* **33** 313–328
Vallortigara G, Bressan P, 1991 "Occlusion and the perception of coherent motion" *Vision Research* **31** 1967–1978
Wallach H, 1935 "Über 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
Wuerger S, Shapley R, Rubin N, 1996 "On the visually perceived direction of motion by Hans Wallach: 60 years later" *Perception* **25** 1317–1367
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

APPENDIX

Naïve subjects participated in all experiments. All had corrected-to-normal vision. The 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 (i) subjects found it unnatural and difficult to maintain fixation while attending to the moving bars and (ii) free viewing more closely approximates natural viewing conditions. Informal observation by experienced psychophysical observers suggested that maintaining fixation would not have qualitatively changed any of the effects described herein.

Subjects used the number pad on the keyboard to enter their responses, pressing either 1, 2, or 3 following each trial to indicate, respectively, completely incoherent (bars moving separately), partially coherent, or completely coherent motion percepts. Prior to the experiments, subjects were shown demonstrations of fully incoherent stimuli (eg high-contrast moving bars without any occluders) and fully coherent stimuli (eg the bars with dots superimposed on them moving in the direction of coherent motion, similar to the control stimulus of experiment 4, shown in figure 9c) in which the bars appear to move as a completely rigid object. These demonstrations helped to 'anchor' subjects' responses. Subjects generally found the task easy and were often happy to be binary with their responses. Coherence judgments were used instead of the more objective direction of rotation judgments used in several previous studies (eg Lorenceau and Shiffrar 1992) because pilot experiments revealed that some subjects could learn to perform the rotation judgments even for conditions which appeared entirely incoherent (these subjects were presumably learning to discriminate the phase relationships between the bars rather than integrating the bar motions, and the judgments of rotation direction were therefore not a suitable measure of motion integration). Subjective coherence judgments have been used in many previous studies with plaids (Adelson and Movshon 1982) as well as other stimuli (eg Castet and Wuerger 1997). Subjects' responses were averaged and normalized to yield a coherence index ranging from 0 to 1. A coherence index of 0 corresponds to a percept of completely incoherent motion on every single trial, while 1 indicates consistently coherent motion. The order of stimulus presentation was randomized across trials. Subjects completed several practice trials before beginning the experimental trials.

In all experiments, we plot data averaged across subjects, for the sake of clarity. Data from individual subjects were qualitatively similar and the qualitative patterns of results that we report have been confirmed in many observers during formal and informal presentations. All the effects described are perceptually strong and are readily seen by most observers in demonstrations.

It is an interesting property of these phenomena that the overall level of coherence for a particular stimulus can vary substantially from subject to subject, which can occasionally result in ceiling and floor effects if the stimuli are not adjusted. This occurs even when coherence criteria are anchored via the inclusion of conditions which are always seen as coherent or incoherent by all observers. In experiments 1 – 4, the luminance contrast of the moving bars was adjusted for each observer to prevent such effects, which would otherwise obscure potential differences between conditions. For stimuli such as the diamond, which consist of multiple moving bars, coherence decreases with increasing bar contrast (Adelson and Movshon 1983; Lorenceau and Shiffrar 1992; McDermott et al 2001). Elsewhere we have verified that the effect of contrast does not interact with the form manipulations employed in this paper (McDermott, Weiss, and Adelson, unpublished data), suggesting that contrast is a reasonable independent variable for us to manipulate to avoid ceiling effects. This was critical in experiment 3 in which the presence of an asymptote was of interest. It is not clear why superthreshold contrast affects coherence, but it may be related to its effect on the strength of the terminator motion signals.

A.1 *Experiment 1*

Stimuli were as shown in figure 4. The background luminance of the stimulus of figure 4a was 9.4 cd m^{-2} (this was also the luminance of the rectangles in figure 4b, so as to keep the junctions identical); the luminance of the occluders of figure 4a (and of the circles of figure 4b) was 30.1 cd m^{-2}; the background luminance of figure 4b was 2.4 cd m^{-2}. The Michelson contrast of the bars was set individually for each subject to help avoid ceiling and floor effects, but was always between 0.5 and 0.75. The speed of the square was 1.67 deg s^{-1}, the range of motion was 0.25 deg, and the stimulus was displayed for 2 s on each trial. The length of the moving bars was 38 pixels (0.6 deg). The width of the T-junction that was held constant across stimuli was 25 pixels (0.4 deg). Note that only the bars moved in the stimuli; the rest of each stimulus was static. The bars approached the borders of the rectangles at the extremes of their trajectories, but never touched. This proximity was not critical to the effect; larger rectangles produced qualitatively similar results. All six subjects, who were naïve as to the purposes of the experiment, completed 15 trials per condition in a single block. This block included five additional conditions, the results of which are not discussed in this paper.

A.2 *Experiment 2*

Stimuli were as shown in figures 5 and 6. The background luminance was 12.0 cd m^{-2}; the luminance of the occluders was 3.0 cd m^{-2}. The contrast of the bars was set individually for each subject but was always between 0.5 and 0.7. The length of the grey background lines was 360 pixels (5.8 deg). The speed of the diamond was 2.3 deg s^{-1}, and the range of its motion was 50 pixels (0.8 degrees). Each trial was 3 s in duration, which allowed for approximately three full rotations of the diamond. Data for the conditions of both figures were collected from five subjects; three additional subjects were run in a separate experiment that contained the conditions of figure 5 along with other conditions not presented here. The data from these three subjects were combined with that from the other five to produce the graph of figure 5. Each subject completed fifteen trials per condition in a single block.

A.3 *Experiment 3*

Stimuli were as shown in figures 7 and 8.

Part 1: Each of twelve naïve subjects completed 20 trials per condition, for a total of 200 trials which were completed in a single block. All other stimulus parameters were as in experiment 2. All subjects were below ceiling in every condition (the data for several subjects who were at ceiling were thrown out). Note that because the diamond is executing a circular trajectory, 'full-occlusion' actually requires slightly larger occluders than are depicted in figure 3b. The label on the graph of figure 3 takes the diamond's motion into account, labeling the point at which the amodal contours of the diamond are entirely covered over its full trajectory.

Part 2: Stimulus parameters were the same as for Part 1, except that data were collected in separate blocks for two configurations: one in which 55% of the diamond's contour was visible, and one in which 20% was visible. This latter configuration was identical to that used in Part 1. Subjects completed 15 trials per condition. Five subjects were below ceiling in all conditions for both configurations; the data presented are theirs.

A.4 *Experiment 4*

Stimuli were as shown in figure 9. The background luminance was 30.1 cd m^{-2}; the luminance within the apertures was 9.8 cd m^{-2}. The contrast of the bars was set individually for each subject but was always between 0.3 and 0.6. The luminance of the dots of figure 4c was 0.49 cd m^{-2}. In order to hold other factors (eg the shape of the occluding contours) constant while altering relatability, the diamond contours were displayed through apertures. Two of the contours were displayed through short apertures, and two through long, so that the contours could be translated without overlapping in the image. The length of the short apertures was 30 pixels (0.5 deg), and the long apertures were twice as long. Each trial was 3 s in duration. Six naïve subjects completed 15 trials per condition in a single block that included six other conditions with intermediate degrees of relatability. These conditions generated the expected intermediate levels of coherence and are omitted for brevity's sake.

A.5 *Experiment 5*

Stimuli were as shown in figure 10. The background luminance was 28.8 cd m^{-2}; the contrast of the moving lines with the background was 0.45, the contrast of the static grating with the background was 0.8, and the contrast of the outer frame was 0.4 (both the static grating and the outer frame were darker than the background). The length of the moving lines was 300 pixels (4.8 deg). The speed of the moving lines changed over time owing to their sinusoidal motion, but the maximum speed during an oscillation was 2.75 deg s^{-1}. The range of motion was 60 pixels (0.97 deg), and each trial lasted 3 s. Six naïve subjects completed 20 trials per condition in a single block.