Object correspondence: Using perceived causality to infer how the visual system knows what went where

Cathleen M. Moore[1] · Teresa Stephens[2] · Elisabeth Hein[3]

Abstract

Anne Treisman and colleagues developed an influential theoretical framework surrounding the construct of "object files" as a means of understanding the functional need for an episodic representation of objects as they move, change, disappear, and reappear from view (Kahneman, Treisman, & Gibbs, *Cognitive Psychology, 24*, 175–219, 1992; Treisman, *The Quarterly Journal of Experimental Psychology*, *40*, 201–237, 1988). Within that framework, object files are defined through the process of object correspondence, whereby stimuli are associated with and represented as later instantiations of existing object representations and are used to selectively update those representations. A central assertion of the object file framework is that object correspondence is established on the basis of spatiotemporal continuity, without regard to feature information. We tested this assertion by investigating whether feature information, separate from spatiotemporal information, can determine how object correspondence is resolved. We used the perception of causality in simple dynamic displays, which provides a means of inferring how object correspondence is resolved. We found that, contrary to the spatiotemporal dominance assertion, feature information is used to resolve object correspondence. We suggest that the object-file framework be extended to reflect the importance of both feature and spatiotemporal information in establishing and maintaining episodic object representations.

Keywords perceptual organization · object-based attention · Motion: *Other

A functional challenge for the visual system is maintaining representations of objects that move and change, and yet persist as continuous entities across time and space. A core component of this functional problem is *object correspondence*,[1] which is illustrated in Fig. 1. Consider the left side of Fig. 1a, which depicts the retinal image of a simple scene at two points in time, three discs in three positions at Time 1, and three discs

in three different positions at Time 2. The right side of the figure illustrates alternative representations of the organization of the scene that gave rise to those retinal images; in other words, representations of what objects went where over time. At Time 1, the three spatially distinct stimuli will give rise to three distinct objects being represented in the scene. The problem of object correspondence arises at Time 2. Here there are also three distinct stimulus images, but which object representation from Time 1 (if any) corresponds to which stimulus at Time 2? How this correspondence problem is resolved will determine the perceived organization of this scene. The right side of the figure illustrates two, of many, alternative possible resolutions of object correspondence and their resulting alternative perceived organizations of the scene for these particular images.

The illustration in Fig. 1a depicts an exaggerated view of the object correspondence problem in at least two different ways. First, when objects move in the physical world, they move smoothly from one location to another, creating a spatiotemporal history in the retinal image that, in turn, provides a means of indexing image-level information over time. Even when an object disappears behind another surface or object,

---

[1] The process that we are referring to as *object correspondence* has also been referred to as *temporal grouping* (Gepshtein & Kubovy, 2000) and *object individuation* (Feldman & Tremoulet, 2006). As will be discussed later it is probably similar, but not identical, to motion correspondence (Kolers, 1972; Ullman, 1979).
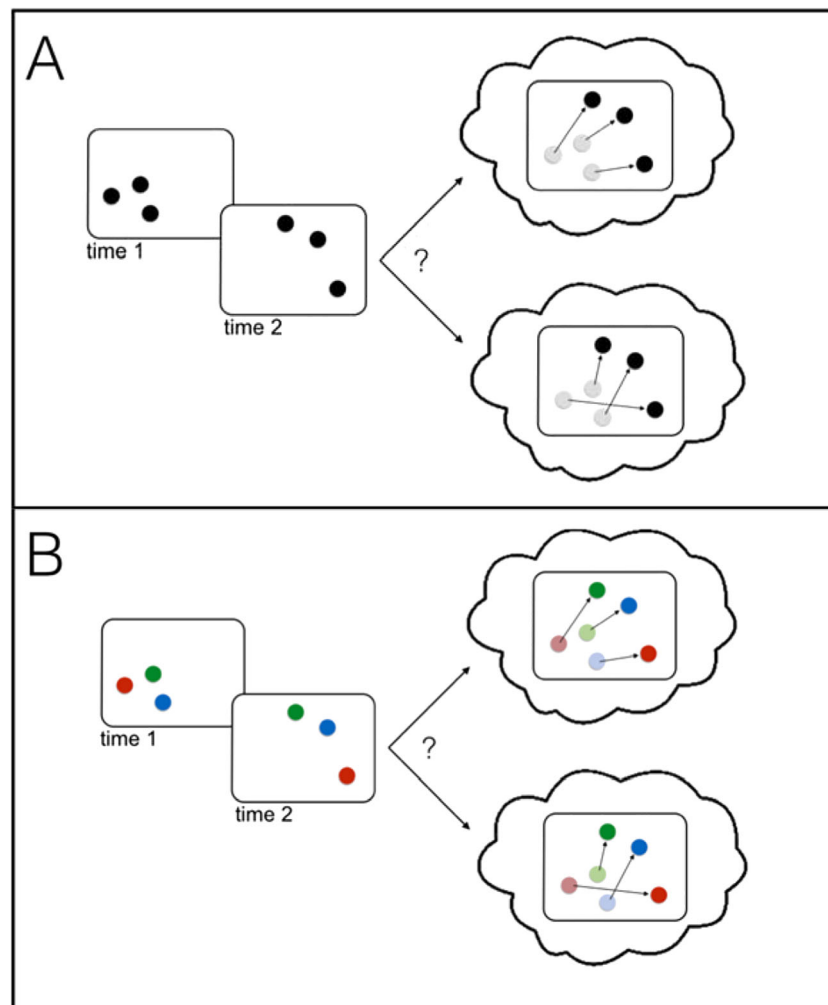
✉ Cathleen M. Moore
cathleen-moore@uiowa.edu

[1] University of Iowa, W311 Seashore Hall, Iowa City, IA 52242, USA

[2] Harford Community College, Bel Air, MD, USA

[3] University of Tübingen, Tübingen, Germany

Fig. 1 Illustration of the object correspondence problem. The left side of each panel illustrates retinal images at two different points in time. The right side illustrates alternative perceptual organizations representing the scene in terms of which objects went where, depending on how object correspondence is resolved. In **a** one might imagine that the top organization is more likely, given its simpler spatiotemporal relations. In **b**, featural continuity conflicts with spatiotemporal continuity. Neither correspondence solution reflects a physically impossible situation. How will object correspondence be resolved? (Color figure online)

the location and time of its reemergence is often predictable based on its spatiotemporal history prior to disappearing. A second way in which Fig. 1a is exaggerated is that objects are rarely identical. Figure 1b illustrates how adding a single distinguishing feature—color in this case—could, in principle, provide information regarding how to resolve object correspondence. Both factors—spatiotemporal continuity and featural continuity—would seem to play a role in resolving object correspondence.

Despite the logical potential for both spatiotemporal and feature information to play a role in object correspondence, multiple theoretical proposals and empirical findings have led to the assertion that object correspondence is resolved almost exclusively on the basis of spatiotemporal information, without regard to feature information (e.g., Flombaum, Scholl, & Santos, 2009; Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 1989, 2001; Scholl, 2007). The general assertion

is that correspondence between stimuli that are closer in time and space will be preferred over correspondence between stimuli that are more separated in time and space, regardless of what feature matches there may or may not be. Under the strong version of this view, features such as color, luminance, texture, and shape are irrelevant to the process of establishing and maintaining object representations over space and time.

The view that only spatiotemporal variables are relevant to establishing and maintaining object correspondence is evident in two influential theoretical frameworks in the visual attention and object-perception literature. One is the theory that proposes the attentional construct of *fingers of instantiation*, or FINSTs (Pylyshyn, 1989, 2001), which are spatiotemporal pointers that define object continuity. FINSTs are indices defined by image location and time and are "blind" to other information such as color, size, and semantic identity. The large literature on multiple-object tracking that has emerged

in the wake of the FINST theory has addressed the question of what role feature and semantic identity information plays in the covert tracking of multiple moving. The findings have been variable, with some studies showing evidence that feature and identity information is used to track objects and others that it is not (e.g., Cohen, Pinto, Howe, & Horowitz, 2011; Horowitz et al., 2007; Makovski & Jiang, 2009; Sun, Zhang, Fan, & Hu, 2018).

The second theoretical framework that asserts that object correspondence depends only on spatiotemporal information is the object-file framework developed by Treisman and colleagues (Kahneman et al., 1992; Treisman, 1988). Within the object-file framework, episodic representations of objects in the world—referred to as *object files*—are established on the basis of spatiotemporal continuity. These representations are episodic in the sense that they are distinct from long-term memory of objects and refer to instantiated objects with countable identities that are within the observer's experience rather than representations of objects that have been abstracted away from specific times and places. Once established, other information, including surface features and semantic identity, can be associated with an object file, and importantly, those associations are maintained and updated in the face of change over space and time. According to this view, feature information is content that is stored "in" an object file, and it therefore cannot itself be a basis on which the carrier representation (i.e., the object file) is defined.

Early evidence concerning the object-file framework came from the object-reviewing paradigm (Kahneman et al., 1992). Figure 2 illustrates a standard version of this method. Two simple shapes (e.g., squares) are shown with, for example, letters in each one (preview stage). The letters disappear, leaving the empty squares, which then move smoothly to new locations. After stopping in their new locations, a single letter is presented in one of the two objects, and subjects name the letter as quickly as they can. The target letter can be the same letter that was presented in the original object during preview (*same-object* condition), the letter that was presented in the other object
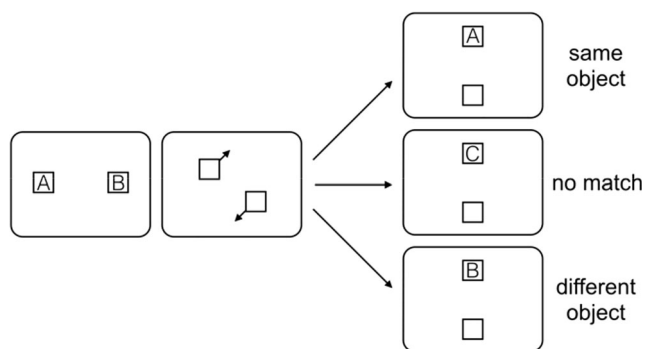
(*different-object* condition), or a new letter altogether (*no-match* condition). The standard finding is that, beyond a general priming effect whereby responses are slowest in the no-match condition, responses tend to be faster in the same-object condition than in the different-object condition (e.g., Hollingworth & Franconeri, 2009; Kahneman et al., 1992; Mitroff & Alvarez, 2007; Mitroff, Scholl, & Wynn, 2005; Moore, Stephens, & Hein, 2010). These *object-specific preview benefits*, as they are known, have been interpreted as evidence that identity information is associated with a specific object representation, and that this information travels with the representation of the object that is perceived as moving to a new location (cf. Mitroff et al., 2005).

Mitroff and Alvarez (2007) tested the assertion that object files are defined only in terms of spatiotemporal continuity using the object-reviewing paradigm. They showed that while object-specific preview benefits occurred when stimuli in the prime display were linked with stimuli in the final display through smooth motion from one location to another, as shown by Kahneman et al. (1992), they did not occur for when the stimuli in the prime display disappeared and later reappeared in the new location with no smooth motion linking them, and were linked only by having the same features (e.g., color and shape) or not.

The lack of object-specific preview benefits in the feature condition of the Mitroff and Alvarez (2007) study, however, may have been caused by a general disruption of correspondence processes due to stimulus discontinuity, rather than a failure of feature-based object correspondence in particular. Hollingworth and Franconeri (2009) placed an occluding surface along the path of motion such that objects could smoothly move behind it and reemerge in a manner that was either consistent or inconsistent with simple smooth motion, and with the same or different features (see also Bae & Flombaum, 2011). Under these discontinuity-matched conditions, object-specific benefits occurred with both spatiotemporal continuity and feature match. Related, Moore et al. (2010) showed that when surface features of objects, such as color, abruptly changed over the course of smooth motion from one location or another (without the benefit of an occluding surface to "explain" the change), object-specific benefits were eliminated, indicating that abrupt feature changes, like abrupt spatial discontinuities, can disrupt object continuity (see also Moore & Enns, 2004; Moore, Mordkoff, & Enns, 2007; Tas, Dodd, & Hollingworth, 2012b; Tas, Moore, & Hollingworth, 2012a).

The current study will test the extent to which feature information can determine object correspondence using a measure of object correspondence that does not depend on memory, as the object-reviewing paradigm does, but that nevertheless uses smooth continuous motion in the

Fig. 2 The object-reviewing paradigm, adapted from Kahneman et al. (1992). See text for details

displays. Before explaining this measure and why we think it is particularly well suited to examine the question at hand, we will first review the original basis of the spatio-temporal priority hypothesis because we believe there has been a conflation in the literature of two different corre-spondence processes that are useful to distinguish.

Object correspondence versus motion correspondence

The original basis for the assertion of spatiotemporal domi-nance in object correspondence came from an earlier literature concerning motion correspondence (Kolers, 1972; Ullman, 1979), which concluded that feature information is, at best, only minimally influential in resolving the correspondence between stimuli during the perception of apparent motion (see Green, 1986, for a review). In this section, we offer the observation that while related, object correspondence and mo-tion correspondence are separable constructs, and therefore evidence regarding the resolution of motion correspondence does not necessarily extend to the resolution of object correspondence.

The simplest apparent motion displays consist of two iden-tical stimuli (e.g., discs) that are presented one after the other at different locations. Within a given range of temporal and spatial separations, observers perceive motion rather than two stationary stimuli (e.g., Wertheimer, 1912). When the distance in space or time is too great or too small, however, observers tend to perceive two stationary stimuli either sequentially or simultaneously (Korte, 1915; Neuhaus 1930). In other words, when the spatiotemporal parameters are "wrong," motion cor-respondence processes fail (see Kolers, 1972; Ullman, 1979, for discussions of motion correspondence and the specific spatiotemporal conditions under which it succeeds or fails).

Multiple studies addressed the question of whether feature relationships, such as whether or not the stimuli have the same color, shape, size, or luminance, for example, might also play a role in motion correspondence. These studies used different types of measures. Some asked observers to report the quality of the apparent motion (e.g., report "good" motion vs. "bad" motion) and tested whether feature similarity increased per-ceived quality of motion. Other studies used ambiguous ap-parent motion displays and tested whether feature similarity could bias the perceived direction of motion. In general, this work can be summarized as showing that the effects of feature similarity on apparent motion, when present, tended to be small in comparison with the effects of spatiotemporal param-eters (e.g., Burt & Sperling, 1981; Cavanagh, Arguin, & von Grünau, 1989; Green, 1986; Kolers & Pomerantz, 1971; Kolers & von Grünau, 1976; Navon, 1976; Nishida, Ohtani, & Ejima, 1992; Sekuler & Bennett, 1996; Shechter,

Hochstein, & Hillman, 1988; Werkhoven, Sperling, & Chubb, 1994).

Collectively, the results from the apparent motion literature were taken as evidence that spatiotemporal information is the primary determinant of motion correspondence. And because perceiving apparent motion between two stimuli seems to im-ply that they were represented as a single object, the results from the apparent motion literature were extended to the ob-ject perception literature, and were also taken as evidence that object correspondence is resolved on the basis of spatiotem-poral information without regard to feature information (e.g., Kahneman et al., 1992; Flombaum et al., 2009; Pylyshyn, 1989, 2001; Scholl, 2007).

Although there is clearly a relationship between motion correspondence and object correspondence, they are not iden-tical constructs (cf. Odic, Roth, & Flombaum, 2012). At a theoretical level, object correspondence is a broader construct than motion correspondence. It includes the continuity of rep-resentations across larger time scales than motion correspon-dence, and does not always include a direct experience of perceived motion, though movement may be implied. Imagine, for example, that you are following a friend in a car on a busy highway, and you lose sight of the car for several minutes. When you sight it again, you perceive it as the same car, and part of that perception is an inference of movement, but the motion was not experienced as such. This is an exam-ple of object correspondence without motion correspondence. Motion correspondence can also occur without object corre-spondence. Specifically, motion can be experienced in the absence of any concomitant set of first-order image features with which the perceived motion energy can be associated (Adelson & Bergen, 1985). This is analogous to feature-less correspondence in stereopsis that Julesz (1971) famously demonstrated. Thus, while motion and object correspondence are related constructs, they are double dissociable.

Using perceived causality as a domain for testing the role feature information in object correspondence

In the current study, we assessed the role of feature informa-tion in the perception of dynamic causal events that allows for an operational definition of correspondence during online per-ception, not memory, such as in the object-reviewing para-digm, and uses continuous, smooth motion. We describe the logic of this approach in the context of introducing the displays.
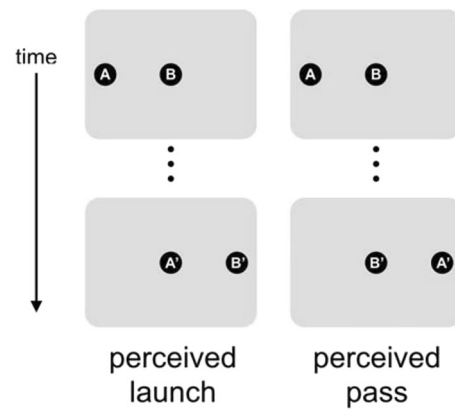
The left side of Fig. 3 (clear launch) illustrates a display that was originally described and studied by Michotte (1963). A stimulus translates smoothly toward a second, identical, but stationary, stimulus until the first is immediately adjacent to the second. The second stimulus then begins to move
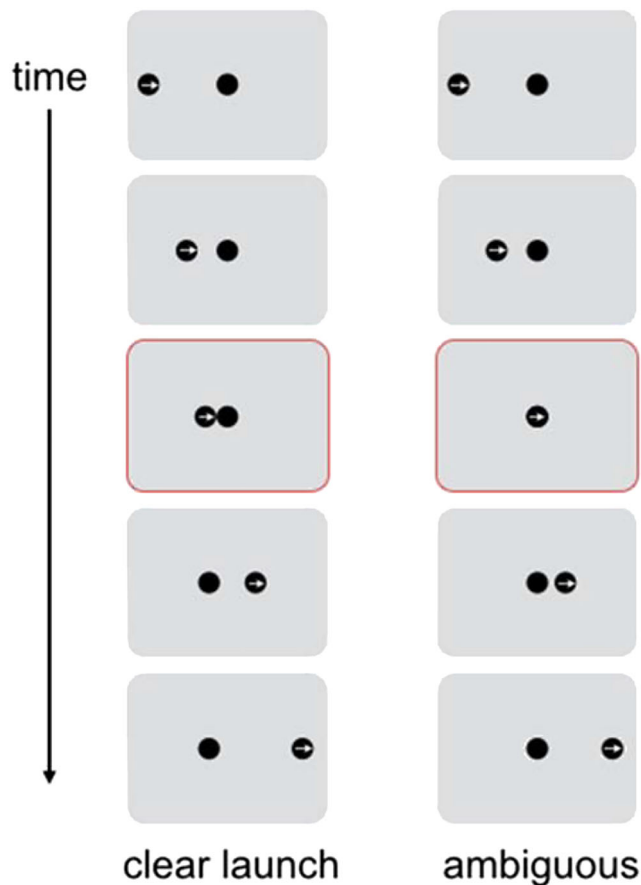
Fig. 3 Conditions under which observers tend to perceive a moving object launch or pass. On the left are displays that observers tend to perceive as clearly launching. On the right is a more ambiguous condition, though the tendency is to perceive it as passing more often than launching. The only difference between these two sequences occurs in the middle of the motion sequence (marked in the figure by the red frame). In the clear launch condition, the moving disc stops when the edges of the two objects are adjacent. In the ambiguous condition, the moving disc completely overlaps the stationary disc. Adapted from descriptions by Michotte (1963)

Fig. 4 Whether a display is perceived as launching or passing implies alternative, mutually exclusive resolutions of object correspondence. Observer reports of "launch" versus "pass," therefore, provide an indirect measure of object correspondence

disc "passing" over a stationary disc that happens to be in its path of motion. However, it can also be perceived as the first disc launching the second disc into motion similar to the display in the left side of Fig. 3 and Movie 1 (e.g., Scholl & Nakayama, 2002; see also Bertenthal, Banton, & Bradbury, 1993, for a similar display).

These simple dynamic displays provide a powerful method for measuring object correspondence during online perception because the different perceptions (launching vs. passing) imply different, mutually exclusive resolutions of object correspondence: Assume that the represented identities of the two discs at the beginning of the display are represented as Object A and Object B, respectively (see Fig. 4). We can then ask which of the two stimuli in the final display corresponds to Object A and which corresponds to Object B? That is, how was object correspondence resolved? When launching is reported (left side of Fig. 4), it implies that Object A ended in the middle position (labeled A′) and Object B ended in the far right position (labeled B′). In contrast, when passing is reported (right side of Fig. 4), it implies that Object A ended in the far-right position and Object B ended in the middle position. Thus, reports of "launch" versus "pass" imply different, mutually exclusive resolutions to the object-correspondence problem, and therefore constitute an indirect measure of object correspondence. By asking what factors affect reports of launching versus passing, we can test what factors, features versus spatiotemporal, affect object correspondence.[2]

smoothly away, while the first stimulus stops. This simple dynamic scene gives a strong impression of the first stimulus as an object that "launches" a second object into motion (see Movie 1). This perception goes beyond the perception of apparent motion in that it represents distinct objects with episodic countable identities and attributes different causal functions to each. It is therefore an excellent context in which to assess the role of feature versus spatiotemporal information in resolving object correspondence, as distinct from motion correspondence.

The right side of Fig. 3 illustrates a more ambiguous situation with regard to causality. Here, the first stimulus travels until it completely overlaps the second stimulus, and then stops. One stimulus then moves smoothly away, while the other remains. Perceptually, this display is asymmetrically bistable (see Movie 2). It is most often perceived as the first

---

[2] Mitroff et al. (2005) used a similar display in which two objects are shown in smooth motion moving toward each other. As they cross paths, they are perceived as either passing by each other and continuing on in the same direction, or bouncing off one another and returning in the direction they came. Like the causality displays here, these two perceptions imply different object correspondences. They measured object-specific preview benefits and found that benefits were systematically associated with the *opposite* object from that predicted by the implied object correspondences. Because we are not using the object-reviewing paradigm here, we simply note this puzzling finding.

Overview of the current study

We used spatiotemporally ambiguous displays like those illustrated on the right side of Fig. 3 to test whether feature information can systematically influence object correspondence, or alternatively, whether only spatiotemporal variables are relevant to that process. Specifically, as illustrated in Fig. 5, we manipulated the features of individual stimuli such that in different conditions, the feature information was consistent with launching, consistent with passing, or provided no biasing information (neutral). In addition, for comparison we included a condition that, similar to Michotte's (1963) original displays, was relatively unambiguous spatiotemporally and regarding its feature information, which we refer to as the "clear launch" condition. Movies 2–4 provide demonstrations of the neutral, launch-consistent, and pass-consistent conditions, respectively, using contrast polarity as the biasing feature.

Notice that the three main conditions (launch-consistent, neutral, and pass-consistent) are identical in terms of their spatiotemporal information. The only thing that differs is the pattern of feature information over time. Therefore, if there are systematic differences in reports of perceived causality (launching vs. passing) across these conditions, then they must be attributed to the feature differences. And because perceived causality reflects object correspondence, any such differences would imply that feature information contributed to the resolution of the object correspondence process.

We used four different feature dimensions—size, contrast polarity (as illustrated in Fig. 5), color plus luminance, and isoluminant color—in order to be sure that whatever results we obtained were not idiosyncratic to any particular type of feature. By way of preview, feature information strongly biased the perception of causality, indicating that feature information, not just spatiotemporal information, contributes to object correspondence processes in these perceived causality displays.

General method

Observers

Twenty-four individuals (20–46 years, $M = 24$; four male, 20 female) who were mostly students from the University of Tübingen were recruited. They were compensated for their time by either getting course credit (10 individuals) or money (€8/hour; 14 individuals). All reported normal or corrected-to-normal visual acuity and color vision, and all were naïve as to the purpose of the experiment.

Design

A 4 (feature type: size, contrast polarity, color plus luminance, isoluminant color) × 4 (display type: clear launch, launch consistent, neutral, pass consistent) within-subjects design was used. The 16 conditions appeared equally often in a pseudo-random order within each block of trials. Data were collected from 10 blocks of 64 trials each for a total of 40 observations in each condition.
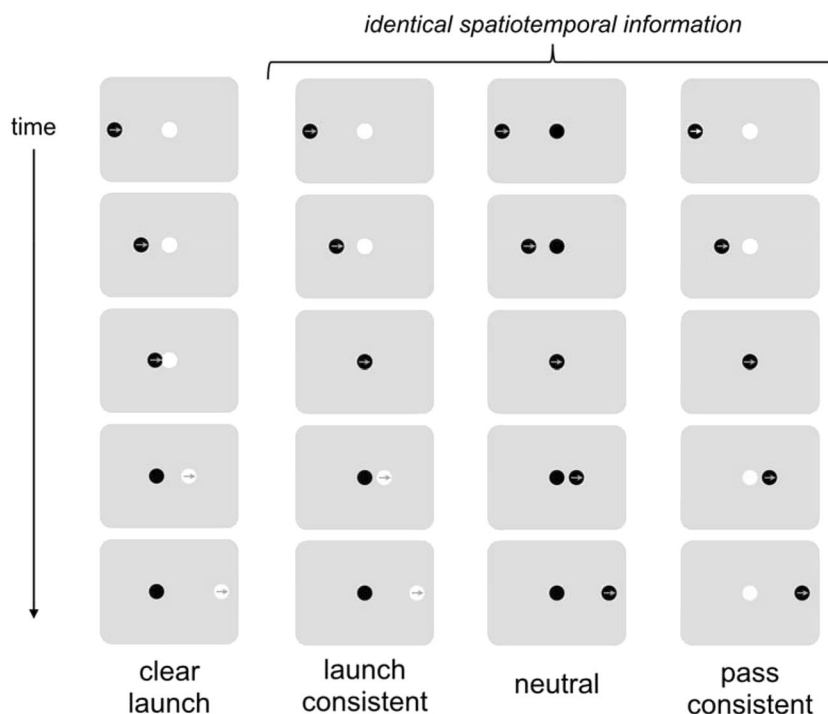
Fig. 5 The different display conditions using contrast polarity as an example feature. See text for details

Apparatus

The experiment was conducted on a PC with Windows XP driving a 17-inch CRT color monitor, set at a spatial resolution of 1,024 × 768 pixels and a refresh rate of 100 Hz. It was programmed using MATLAB software (Version R2012a, 7.14, The MathWorks Inc., Natick, MA, USA) with the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Viewing distance was fixed at 65 cm using a chin rest. Observers were tested in individual rooms with dimmed room illumination.

Stimuli

Size Stimuli consisted of black (<1 cd/m^2) discs that were either 2.3° or 1.1° in diameter, presented on a gray (25 cd/m^2) background.

Contrast polarity Stimuli consisted of white (143 cd/m^2) or black (<1 cd/m^2) discs (2.3° in diameter), presented on a gray (25 cd/m^2) background.

Color plus luminance Stimuli consisted of blue (CIE: 39, 114, 302; 17 cd/m^2) or yellow (CIE: 96, 97, 106; 127 cd/m^2) discs (1.4° in diameter), presented on a gray (CIE: 56, 13, 257; 25 cd/m^2) background.

Isoluminant color Stimuli consisted of turquoise (average CIE: 56, 32, 63; 25 cd/m^2) or orange (average CIE: 55, 56, 63; 25 cd/m^2) discs (2.3° in diameter), drawn on a gray (56, 13, 257; 25 cd/m^2) background of the same luminance as the stimuli.

The four different display types are illustrated in Fig. 5. All four conditions began with two horizontally aligned discs, with one at the center of the display and one to the left or right (equally often, pseudorandomly selected). In the neutral conditions, the two discs were identical. Specifically, they were both blue (or yellow), both turquoise (or orange), both black (or white), or both large (or small), depending on the feature condition. In the other three display conditions, the disc off to the side had one of the two features being tested in the given feature condition, while the other disc had the other feature value. Specifically, the disc off to the side was blue (or yellow), turquoise (or orange), black (or white), large (or small), while the disc at the center was yellow (or blue), orange (or turquoise), white (or black), small (or large). For the clear-launch conditions, the two discs were identical with the given feature (big or small, yellow or blue, turquoise or orange) occurring equally often, pseudorandomly selected.

The motion sequences for each of the four display conditions are illustrated in Fig. 5. In all conditions, when the discs moved, they moved smoothly at 20° per second. In the clear-launch conditions, the disc started at 16.8° (or 15.6° for the small disc) off to the side, traveled toward the center disc until its leading edge reached the near edge of the stationary disc, and then stopped. The center disc then immediately began traveling smoothly at the same speed and in the same direction as the original disc until it was 14.5° from the center and stopped. In the other three conditions (i.e., launch consistent, neutral, pass consistent), the disc started 14.5° off to the side at the beginning of the trial, traveled smoothly until it was centered on the center disc (both discs overlapping completely), and then stopped. Another disc immediately moved away from the center position at the same speed and stopped at the opposite side of the display. In the pass-consistent conditions, the disc that moved away from the center after the first disc stopped was identical to the first moving disc. In the launch-consistent condition, it had the opposite feature (like the disc that was at the center of the screen at the beginning of the trail). Whether the discs traveling toward the center moved in front or behind the disc at the center was randomly selected.

Task

The task was to report whether a display appeared as though the first moving disc "launched" the second disc into motion, or whether it appeared as though the first disc "passed over" the second disc by pressing the "j" or "f" key, respectively.

Procedure

Each observer participated in a single session that lasted approximately 1 hour. Following the informed-consent process, they received written instructions describing the task. A "launch" was described as "when it looks as though the first disc moves across the screen, makes contact with the stationary disc near the center, and causes the stationary disc to move across the screen." A "pass" was described as "when it looks as though the first disc moves across the screen, passes over a stationary disc at the center of the screen, and continues on to the other side" (translated from German).

Following instructions, observers completed a block of 10 practice trials, which were randomly selected from among the 16 possible trial types. Observers were encouraged to rest between blocks as much as they liked. They self-initiated the next block of trials by pressing the space bar.

Each trial began with the two discs in their starting positions for 300 ms, at which point the motion sequence for that condition (see above) commenced. At the end of the motion sequence, the discs were presented for another 300 ms in their end position. Then the screen went blank and 500 ms after the observer's response, the next trial began. If the participant pressed a key other than one of the two response keys, the error message "Wrong key" was shown.
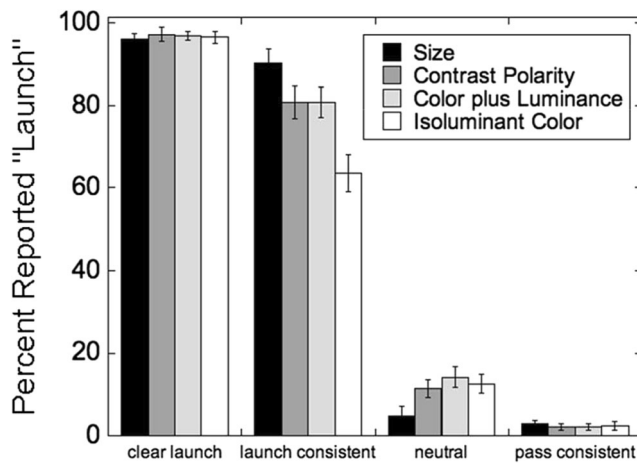
Fig. 6 Percentage of trials on which observers reported "launch" for each condition and each feature type, separately. Error bars are condition-specific within-subject standard errors (Cousineau, 2005; Morey, 2008)

Results and discussion

Fourteen trials (0.09%) of the entire data set had to be excluded from the analysis because of key presses other than the two possible response keys. Figure 6 shows the mean percentage of trials on which observers reported "launch" as a function of display type for each of the four feature conditions, respectively. We submitted participant means to a 4 (feature type) × 4 (display type) repeated-measures analysis of variance (ANOVA) on the arcsin transformation of the proportion of trials reported launching,[3] with alpha set to .05. Where appropriate, Huynh–Feldt corrections were applied to address violations of sphericity. Standardized effect sizes are reported in terms of adjusted partial eta squared (adj $\hat{\eta}_p^2$; Mordkoff, 2019).

Both main effects were significant, feature type: $F(1.70, 39.12) = 12.53$, $p < .001$, adj $\hat{\eta}_p^2 = .33$, and display type: $F(2.63, 60.51) = 424.12$, $p < .001$, adj $\hat{\eta}_p^2 = .95$, as was the interaction, $F(2.79, 119.50) = 26.51$, $p < .001$, adj $\hat{\eta}_p^2 = .51$. We followed up this omnibus analysis with separate four-level (display type) one-way ANOVAS for each of the four feature types. The effect of display type was reliable for all four feature types—size: $F(1.62, 37.21) = 417.73$, $p < .001$, adj $\hat{\eta}_p^2 = .95$; contrast polarity: $F(1.72, 39.59) = 346.62$, $p < .001$, adj $\hat{\eta}_p^2 = .94$; color plus luminance: $F(1.96, 45.00) = 260.79$, $p < .001$, adj $\hat{\eta}_p^2 = .92$; and isoluminant color: $F(1.90, 43.72) = 276.71$, $p < .001$, adj $\hat{\eta}_p^2 = .92$.

Because the clear-launch condition was different spatiotemporally from the other conditions, we followed up the

initial analysis with a set of ANOVAs in which only the three spatiotemporally identical display conditions—launch consistent, neutral, and pass consistent—were included to confirm that the effect of display type holds even when the spatiotemporally distinct condition is not contributing. We began with a 4 (feature type) × 3 (display type) repeated-measures ANOVA. Again, both main effects were significant, feature type: $F(2.68, 61.65) = 20.49$, $p < .001$, adj $\hat{\eta}_p^2 = .45$, and display type: $F(1.16, 26.70) = 221.56$, $p < .001$, adj $\hat{\eta}_p^2 = .90$, as was the interaction $F(3.34, 76.82) = 37.80$, $p < .001$, adj $\hat{\eta}_p^2 = .61$. Separate three-level (display type) one-way ANOVAs confirmed that the effect of display type was reliable for each of the four feature types—size: $F(1.04, 24.02) = 322.00$, $p < .001$, adj $\hat{\eta}_p^2 = .93$; contrast polarity: $F(1.20, 27.55) = 182.60$, $p < .01$, adj $\hat{\eta}_p^2 = .88$; color plus luminance: $F(1.22, 28.15) = 131.64$, $p < .001$, adj $\hat{\eta}_p^2 = .85$; and isoluminant color: $F(1.27, 29.20) = 99.37$, $p < .001$ adj $\hat{\eta}_p^2 = .80$—confirming that the effect of display type holds even when the only differences between conditions is the feature information.

Collapsing across feature type, the pattern of the effect of display type on perceived launching was clear. Observers reported launching on 96.69% ± 2.44% of the clear-launch trials, on 77.59% ± 7.19% of the launch-consistent trials, 10.76% ± 3.56% of the neutral trials, and 2.43% ± 1.55% of the pass-consistent conditions, respectively. Given the non-overlapping confidence intervals, it is clear that first, each of these conditions yielded different levels of perceived launching from the other conditions. Second, consistent with previous findings, there was a strong tendency to perceive the neutral version of these displays (spatiotemporal information only) as passing, rather than launching. Because response rate is bounded by 0%, this means that there was relatively little opportunity for the effect of feature biasing in the pass-consistent condition to reveal itself. Despite that, launching was reported reliably less often in the pass-consistent condition than in the neutral condition, overall. Moreover, it indicates that the impact of features on correspondence did not occur simply because the spatiotemporal information was ambiguous; the launch-consistent condition yielded nearly 80% launch reports across feature conditions, despite the fact that in the absence of feature information (the neutral condition), observers reported "passing" on nearly 90% of the trials on average. Feature information supporting launching overrode a strong perception of passing based on the spatiotemporal information.

While the general pattern is consistent across feature types, as revealed in the omnibus ANOVA, feature type reliably interacted with display type in determining the specific rate of reporting "launch." We therefore conducted separate one-way ANOVAs testing the relative effects of feature type for

---

[3] Because of the bounded nature of the measure and the extreme values in some conditions (near 100% and 0%), transforming the data failed to yield normally distributed data. The violation of the assumption of normality, however, leads to increased risk of missing reliable effects, and therefore is conservative with regard to finding evidence of reliable effects.

each display condition. The following patterns were confirmed (see Fig. 6). There was no effect of feature type in either the clear-launch condition, $F(3, 69) = 0.76$, ns, adj $\hat{\eta}_p^2 = -.01$, or the pass-consistent condition, $F(3, 69) = 0.48$, ns, adj $\hat{\eta}_p^2 = -.02$. Different features affected the percentage reported "launch" to varying degrees in the launch-consistent condition, $F(2.50, 57.38) = 35.81$, $p < .001$, adj $\hat{\eta}_p^2 = .59$. Pairwise comparisons (Dunn–Šidák) confirmed that size bias resulted in the highest percentage of "launch" reports (all $ps < .01$), and isoluminant-color bias resulted in the lowest percentage of "launch" reports (all $ps < .01$), while contrast polarity and color plus luminance fell in between and were not reliably different from each other ($p > .10$). That is, when a small disc approached a large disc, and a large disc moved on, the display was especially likely to be perceived as the small disc launching the large disc into motion; the same held for the reverse relationship. Finally, feature-type reliably affected percentage of reported "launch" in the neutral condition as well. In this case, pairwise comparisons (Dunn–Šidák) confirmed that size bias resulted in fewer "launch" reports than any other feature type (all $ps < .01$), while no other feature type differed from the others (all $ps > .10$).

Why these particular relationships? Size may have been a particularly effective manipulation in the launch-consistent condition for several reasons. First, shape (of which size is a parameter) may be a particularly strong cue to correspondence. Second, there may have been a small spatiotemporal component to size differences because the edges of larger discs are in different spatial positions than those of smaller discs. The contrast polarity and color-plus-luminance conditions may have been similarly effective because they share the component of luminance differences. Related, isoluminant color, while still strongly effective, may have been the least effective feature because hue was the only distinguishing feature in that case. We offer details of the relative effectiveness of feature type in this particular experiment only for completeness of understanding this data set. The specific pattern of which features are more powerful than others with regard to influencing object correspondence in any given set of displays will no doubt depend on the specific parameters. If smaller size differences were used, for example, then contrast polarity may have been more powerful than size. Or if luminance-contrast differences were reduced in the contrast polarity condition, then its effect may have been more similar to the isoluminant color condition. Moreover, the broader context in which events occur may affect the relative effectiveness of features. Scholl and Nakayama (2002) used displays much like the launch-consistent contrast-polarity displays used in the current study and obtained much lower rates of reported launching. Theirs' was a study of the effect of including additional, spatiotemporally unambiguous (clear launch) displays in the same scene as the ambiguous displays on how the ambiguity was resolved. They found that ambiguous displays were strongly biased in the direction of less ambiguous displays that were shown at the same time. In some contexts, launch-consistent contrast polarity information may be insufficient to override the spatiotemporal bias to resolve the ambiguity as passing (see also Bae & Flombaum, 2011).

The more general point that we wish to highlight is that the launch-consistent, neutral, and pass-consistent display conditions in this experiment differed from each other only in how feature information changed over the course of the display, and yet reports of perceived causality differed systematically across those display conditions, consistent with the correspondence relations indicated by feature information. Spatiotemporal information cannot, therefore, be the only determining factor in the resolution of object correspondence.

General discussion

We used perceived causality to test whether the visual system relies on feature information when resolving object correspondence during online perception of dynamic scenes or, alternatively, whether only spatiotemporal factors are relevant to that process. Perceived launching versus perceived passing in displays like those used in this study imply different solutions to object correspondence (see Fig. 4), and can therefore be used as a measure of how object correspondence was resolved. Four different types of feature information (size, contrast polarity, color plus luminance, and isoluminant color) were manipulated in displays to bias a launching percept (launch consistent), a passing percept (pass consistent), or neither (neutral). Because these conditions were identical in terms of spatiotemporal information, any differences across them would imply a role of features in perceived causality, and by logical extension, object correspondence. In fact, feature-bias conditions strongly influenced whether launching or passing was perceived for all four feature types, thus confirming that feature information is used to resolve object correspondence during online perception, even with continuous motion. This is contrary to the assertion that only spatiotemporal information is used to define episodic object representations as in the object-file (Kahneman et al., 1992) and FINST (Pylyshyn, 1989, 2001) theoretical frameworks.

The current results are analogous to a previous study in which it was shown that feature information, including color, orientation, and luminance, matches between stimuli, influenced how observers perceived ambiguous apparent motion (Hein & Moore, 2012). That study used Ternus displays (Pikler, 1917; Ternus, 1926), for which two very different motion percepts are perceived depending on how correspondence is resolved and large feature effects were observed (Hein & Moore, 2012; see also Kramer & Yantis, 1997; Petersik & Rice, 2006). These findings with Ternus

displays contrast with those of classic apparent motion which, as summarized in the introduction, has shown a strong dominance of spatiotemporal information and at best small feature effects (e.g., Burt & Sperling, 1981; Cavanagh et al., 1989; Green, 1986; Kolers & Pomerantz, 1971; Kolers & von Grünau, 1976; Navon, 1976). It is possible that Ternus displays, which are more complex than simple apparent-motion displays, engage object-correspondence processes in addition to motion-correspondence processes, and that this is why it reveals such large feature effects. This interpretation is supported by other recent studies showing that a perceived feature, specifically lightness (perceived) rather than luminance (physical), determined how Ternus displays were resolved (Hein & Moore, 2014), that the relevant spatial framework for Ternus motion is spatiotopic, rather than retinotopic (Hein & Cavanagh, 2012), that the history of the elements making up a Ternus display in terms of their object structure (grouped or ungrouped) influenced how Ternus motion was perceived (Stepper, Moore, Rolke, & Hein, 2019a), and, finally, that endogenously cued attention similarly influenced correspondence in Ternus motion (Stepper, Rolke, & Hein, 2019b).

Given the clear influence of both feature information and spatiotemporal continuity on the resolution of correspondence in these displays, the question then becomes how are different sources of information weighted? Feldman and Tremoulet (2006) suggested a Bayesian solution whereby the visual system resolves object correspondence based on the most plausible correspondence of visual information, whether it is spatiotemporal or featural, and "plausible" is determined by environmental probabilities and prior experience. In the case of no feature differences (i.e., infinite featural similarity), of course, spatiotemporal factors will dominate. In the case of feature differences, however, the greater the featural similarity across time and space, the more likely it is that it derives from the same object, and therefore the more likely it will be used to resolve correspondence.

Although a model based on prior probabilities of specific cues being associated with continuous objects is consistent with Feldman and Tremoulet's (2006) findings, Caplovitz, Shapiro, and Stroud (2011) showed that ambiguous motion displays (i.e., the bouncing-streaming display; see, e.g., Bertenthal et al., 1993) can be resolved in a way that involves the perception of the two objects switching feature values along the path of motion, which is clearly inconsistent with real-world statistics where features do not usually migrate from object to object. See Kanisza's (1979, pp. 49–54) description of a model man who appears to pass a dot from one foot to another while seeming to hop up and down for an early (and entertaining) description of this phenomenon. Caplovitz et al. (2011) initially showed this for simple features like color and texture, but it also held for more complex

features (face identity) for which switching values would be extremely unlikely in terms of prior experience (Shapiro, Caplovitz, & Dixon, 2014). Caplovitz et al. argue that the resolution of correspondence depends on image-level features that are not interpreted in terms of object properties (i.e., as belonging to one object or another). The objects that were perceived as switching some features shared a lot of other features (e.g., size, shape, orientation), and that a large set of shared features was sufficient to drive correspondence. Under this view, the perception of the switching features was a consequence of the visual system reconciling the remaining feature information, given the specific resolution of correspondence. The main point from all of these studies for current purposes is that neither spatiotemporal nor feature information is specifically prioritized over the other for resolving object correspondence, but rather that all available cues contribute to a solution that is most consistent with the entire set of cues in a specific situation.

With regard to the object-file framework, which has as a central tenet that object correspondence depends only on spatiotemporal continuity (Kahneman et al., 1992), we suggest that little is lost from the value of the theory by relaxing that assertion. It seems likely that the inclusion of the claim that object files are defined only on the basis of spatiotemporal continuity in the development of the object-file framework reflects the collective understanding at the time from the motion-correspondence literature that features are (nearly) irrelevant in resolving correspondence in apparent motion displays, rather than anything fundamental to object perception and attention. And as we noted, object-correspondence and motion-correspondence are dissociable processes. While the metaphor of a file in which information is stored does lead to the implication that content of the file cannot serve to define the file itself, that is a limit of the metaphor rather than the more general construct of an episodic carrier representation. Related, while FINSTs, as originally conceptualized, were limited to spatiotemporal definition, there is nothing that logically precludes them from including featural indexing with their spatiotemporal pointers.

In summary, this work contributes to a growing body of work confirming that feature information is not disregarded during object correspondence processes. Rather, the visual system uses all of the information that it has available to it—spatiotemporal and featural—to resolve object correspondence. This conclusion is contrary to a central assertion of the object-file (Kahneman et al., 1992) and the FINST (Pylyshyn, 1989, 2001) theoretical frameworks of attention and object perception, and future developments of those frameworks should reflect its rejection. However, the rejection of spatiotemporal priority does not undermine the value of either framework because, we suggest, they were never really dependent on it.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America a—Optics Image Science and Vision, 2*, 284–299.

Bae, G. Y., & Flombaum, J. I. (2011). Amodal causal capture in the tunnel effect. *Perception, 40,* 74–90.

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

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

Caplovitz, G. P., Shapiro, A. G., & Stroud, S. (2011). The maintenance and disambiguation of object representations depend upon feature contrast within and between objects. *Journal of Vision, 11*(14), 1, 1–14. https://doi.org/10.1167/11.14.1

Cavanagh, P., Arguin, M., & von Grünau, M. (1989). Interattribute apparent motion. *Vision Research*, *29*, 1197–1204.

Cohen, M. A., Pinto, Y., Howe, P. D. L., & Horowitz, T. S. (2011). The what–where trade-off in multiple-identity tracking. *Attention, Perception, & Psychophysics, 73,* 1422–1434.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45.

Feldman, J., & Tremoulet, P. D. (2006). Individuation of visual objects over time. *Cognition*, *99*, 131–165.

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 135–164). Oxford: Oxford University Press.

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

Green, M. (1986). What determines correspondence strength in apparent motion? *Vision Research*, *26*, 599–607.

Hein, E., & Cavanagh, P. (2012). Motion correspondence in the Ternus display shows feature bias in spatiotopic coordinates. *Journal of Vision*, *12*, (7), 16, 1–14. https://doi.org/10.1167/12.7.16

Hein, E., & Moore, C. M. (2012). Spatio-temporal priority revisited: The role of feature identity and similarity for object-correspondence in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 975–988.

Hein, E., & Moore, C. M. (2014). Evidence of scene-based motion correspondence, *Attention, Perception, & Psychophysics*, *78*, 793–804.

Hollingworth, A., & Franconeri, S. L. (2009). The role of surface features in establishing object correspondence. *Cognition*, *113*, 150–166.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics, 69,* 172–184.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: University of Chicago Press.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt Perception.* New York, NY: Praeger.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(1), 1–16. https://doi.org/10.1068/v070821

Kolers, P. A. (1972). *Aspects of motion perception*. Oxford, UK: Pergamon Press.

Kolers, P. A., & Pomerantz, J. R. (1971). Figural change in apparent motion. *Journal of Experimental Psychology, 87*(1), 99–108.

Kolers, P., & von Grünau, M. (1976). Shape and color in apparent motion. *Vision Research*, *16*(4), 329–335.

Korte, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, *72*, 194-296.

Kramer, P., & Yantis, S. (1997). Perceptual grouping in space and time: evidence from the Ternus display. Perception & Psychophysics, 59, 87-99.

Makovski, T., & Jiang, Y. V. (2009). Feature binding in attentive tracking of distinct objects. *Visual Cognition*, *17*, 180-194.

Michotte, A. (1963). *The perception of causality.* [T.R. Miles & E. Miles, Trans.]. London, UK: Methuen. (Original work published 1932)

Mitroff, S.R., & Alvarez, G.A. (2007). Space and time, not surface features, underlie object persistence. *Psychonomic Bulletin & Review, 14*, 1199–1204.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition*, *96*, 67–92.

Moore, C. M., & Enns, J. T. (2004). Substitution masking and the flash-lag effect. *Psychological Science*, *15*, 866–871.

Moore, C. M., Mordkoff, J. T., & Enns, J. T. (2007). Path of least persistence: Object status mediates visual updating. *Vision Research*, *47*, 1624–1630.

Moore, C. M., Stephens, T., & Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin & Review*, *17*, 731–736.

Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared ($adj\ \hat{\eta}_p^2$). *Advances in Methods and Practices in Psychological Science*.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.

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

Neuhaus, W. (1930). Experimentelle Untersuchung der Scheinbewegung [Experimental investigations of apparent motion]. *Archiv für die Gesamte Psychologie*, *75*, 315–458.

Nishida, S., Ohtani, Y., & Ejima, Y. (1992). Inhibitory interaction in a split/fusion apparent motion: Lack of spatial frequency selectivity. *Vision Research*, *32*, 1523–1534.

Odic, D., Roth, O., & Flombaum, J. (2012). The relationship between apparent motion and object files. *Visual Cognition*, *20*, 1052–1081.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Petersik, J., & Rice, C. (2006). The evolution of explanations of a perceptual phenomenon: A case history using the Ternus effect. *Perception, 35,* 807–821.

Pikler, J. (1917). Sinnesphysiologische Untersuchungen [Physiological investigations of the senses]. Leipzig, Germany: J. A. Barth.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32*, 65–97.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind & Language, 22*(5), 563–591.

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science, 13*(6), 493–498.

Sekuler, A. B., & Bennett, P. J. (1996). Spatial phase differences can drive apparent motion. *Perception & Psychophysics, 58*, 174–190.

Shapiro, A. G., Caplovitz, G. P., & Dixon, E. L. (2014). Feature- and face-exchange illusions: new insights and applications for the study of the binding problem. *Frontiers in Human Neuroscience*, 1–6. https://doi.org/10.3389/fnhum.2014.00804

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

Stepper, M. Y., Moore, C. M., Rolke, B., & Hein, E. (2019a). *The role of object history in establishing object correspondence.* Manuscript submitted for publication.

Stepper, M. Y., Rolke, B., & Hein, E. (2019b). *How voluntary spatial attention influences feature biases in object correspondence.* Manuscript submitted for publication.

Sun, M. D., Zhang, X. M., Fan, L. X., & Hu, L. M. (2018). Hue distinctiveness overrides category in determining performance in multiple object tracking. *Attention, Perception, & Psychophysics, 80*, 374–386.

Tas, A. C., Dodd, M. D., & Hollingworth, A. (2012b). The role of surface feature continuity in object-based inhibition of return. *Visual Cognition, 20*, 29–47.

Tas, A. C., Moore, C. M., & Hollingworth, A. (2012a). An object-mediated updating account of insensitivity to transsaccadic change. *Journal of Vision, 12*(11), 18, 1–13.

Ternus, J. (1926). Experimentelle Untersuchungen über phänomenale Identität [Experimental investigations of phenomenal identity]. *Psychologische Forschung, 7*, 81–136. Translated to English in W. D. Ellis (Ed.), (1950). *A sourcebook of gestalt psychology.* New York, NY: Humanities Press.

Treisman, A. (1988). Features and object: The Fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology, 40*, 201–237.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Werkhoven, P., Sperling, G., & Chubb, C. (1994). Perception of apparent motion between dissimilar gratings: Spatiotemporal properties. *Vision Research, 34*, 2741–2759.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung [Experimental studies of the perception of motion]. *Zeitschrift für Psychologie, 61*, 161–265. Translated in part in T. Shipley (Ed.). (1961). *Classics in Psychology.* New York, NY: Philosophical Library.