How Do We Know Which Object Went Where? Investigating Higher-Level Influences on Object Correspondence

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Madeleine Yvonne Stepper
aus Esslingen am Neckar

Tübingen
2021

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 29.06.2022
Dekan: Prof. Dr. Thilo Stehle
1. BerichterstatterIn: Dr. Elisabeth Hein
2. BerichterstatterIn: Prof. Dr. Markus Huff
3. BerichterstatterIn: Prof. Dr. Ingrid Scharlau

Danksagung

Danke an alle, die mich während dieses Dissertationsprojektes bis hin zum Abschluss dieser Arbeit unterstützt haben.

Mein Dank gilt ganz besonders meiner Betreuerin Dr. Elisabeth Hein für die Ermöglichung dieses Dissertationsprojektes, ihre Unterstützung und Motivation und dafür wie viel ich über die Jahre von ihr - wissenschaftlich und menschlich - lernen durfte. Auch möchte ich Prof. Dr. Bettina Rolke für ihre Beiträge und ihre Unterstützung zu diesem Dissertationsprojekt danken. Dank gilt auch allen anderen Kollegen der Arbeitsgruppe Evolutionäre Kognition, Janina Balke, Dr. Susanne Dietrich, Heidrun Fellmeth und Dr. Verena Seibold, für die großartige Arbeitsatmosphäre, eure aufbauenden Gespräche und eure Hilfsbereitschaft. Ganz besonders Dr. Verena Seibold, die mir mit ihrer (statistischen) Expertise stets zur Seite stand.

Besonderer Dank geht auch an meine Familie und meine Freunde. Meinen Eltern, Olivia und Walter, und meinem Bruder Maximilian, danke ich dafür, dass sie immer für mich da sind und mich unterstützen. Lea, danke dass ich dich als Freundin haben darf und dafür, dass ich mich immer auf dich verlassen kann. Jasmin, danke für die vielen tollen Jahre deiner Freundschaft. Svea, danke für deine Unterstützung und das du immer ein offenes Ohr für mich hast! Danke auch an alle, die hier namentlich nicht aufgeführt sind und die mich durch diese Zeit begleitet haben und die ich alle nicht missen möchte.

Das wohl größte Dank aber geht an meinen Mann Steffen. Danke für deine Unterstützung, deine Geduld, dein Verständnis und den Rückhalt, den du mir gibst in dem was ich tue. Danke dafür das ich mein Leben mit dir und unserem Sohn Ben teilen darf.

Zusammenfassung

Unser visuelles System ermöglicht eine kohärente und stabile Wahrnehmung unserer Umwelt. Diese verändert sich allerdings stetig, weswegen das visuelle System dazu das Korrespondenzproblem lösen muss, d. h. es muss bestimmen, welches Objekt sich durch Raum und Zeit wohin bewegt hat. Diese Fähigkeit des visuellen Systems ist bemerkenswert, da es sich unsere dreidimensionale Welt aufgrund zweidimensionaler Informationen, die auf unsere Netzhaut projiziert werden, erschließen muss. Diese Informationen sind mehrdeutig, denn ein und dasselbe zweidimensionale Netzhautbild kann mehreren, unterschiedlichen Bildern in der dreidimensionalen Welt entsprechen. Hinzu kommt, dass die Welt um uns herum ständig in Bewegung ist und Objekte sich oft gegenseitig verdecken. Darüber hinaus hat die Korrespondenzlösung einen starken Einfluss darauf, wie wir unsere Umwelt wahrnehmen. Sie bestimmt, ob wir Objekte über Raum und Zeit hinweg als ein und dasselbe Objekt, in Bewegung, wahrnehmen und dadurch, wie wir mit unserer Umgebung interagieren. Unsere Wahrnehmung (und letztendlich Objekterkennung) wird durch visuelle Bottom-up-Verarbeitung ermöglicht. Diese beginnt auf niedriger Ebene, wo Merkmale aus bildbasierten Informationen, wie Leuchtdichtekontrast und Kantenausrichtung, extrahiert werden. Diese werden auf mittlerer Ebene weiterverarbeitet, auf der eine Gruppierung der Merkmale stattfindet, was zu objektbasierten Informationen führt, wie der Repräsentation eines Objekts mit realen Farben (abstrahiert von der Umgebungsbeleuchtung) und kombinierten Konturen, in 3D dargestellt. Auf einer höheren Ebene erfolgt schließlich die Objekterkennung, indem die Wahrnehmung durch Top-down-Informationen, wie semantischen Informationen, ergänzt wird. Den Korrespondenzprozess betreffend ist immer noch unklar, welche Ebenen der visuellen Verarbeitung an der Lösung des Korrespondenzproblems beteiligt sind und vor allem, ob und wenn ja, wie höhere Verarbeitungsebenen dazu beitragen. Die objektbasierte Korrespondenztheorie legt nahe, dass Korrespondenz auf einer mittleren Verarbeitungsebene, unter Verwendung objektbasierter Informationen, gelöst werden könnte. Darüber hinaus wird ein aufmerksamkeitsbasierter Mechanismus vorgeschlagen, wodurch Korrespondenz zwischen Objekten anhand ihrer Identität hergestellt wird. Um die objektbasierte Korrespondenztheorie im Rahmen dieser Arbeit zu testen, wurden drei Studien, einschließlich mehrerer Experimente, durchgeführt. Dabei wurde das Ternus Display, ein mehrdeutiges Scheinbewegungsdisplay, als Maß für die Korrespondenz verwendet. Dieses besteht aus drei nebeneinander ausgerichteten Elementen, die von einem Frame zum nächsten um eine Elementposition verschoben werden. Je nachdem, wie Korrespondenz zwischen den Elementen hergestellt wird, können verschiedene Arten von Bewegung wahrgenommen werden (Element- oder Gruppenbewegung). In Studie 1 wurde untersucht, ob unser visuelles System objektbasierte Informationen zur Lösung der Korrespondenz verwendet, die auf mittlerer Ebene der visuellen Verarbeitung zur Verfügung stehen. Mit Hilfe eines (Ponzo-ähnlichen) Tiefenillusionshintergrund wurde dazu die wahrgenommene Größe des Ternus Displays manipuliert, während die Netzhautgröße, auf niedriger Verarbeitungsebene, gleichblieb. Die Ergebnisse zeigten einen Einfluss der wahrgenommenen Größe auf die Korrespondenzlösung, Evidenz dafür, dass Informationen auf mittlerer Ebene, nach Herstellung der Größenkonstanz, verwendet werden. Studie 2 untersuchte, ob objektbasierte Informationen, die vor der Präsentation des Ternus Displays, anhand unterschiedlicher Objektgeschichten, präsentiert werden, die Korrespondenz ebenfalls beeinflussen können. Dies konnte gezeigt werden, was darauf hindeutet, dass Informationen, die in Objektrepräsentationen gespeichert sind und die zum Zeitpunkt der Lösung der Korrespondenz nicht im Bild vorhanden sind, die Korrespondenz beeinflussen können. Studie 3 untersuchte, ob willentliche Aufmerksamkeit die Korrespondenzlösung beeinflussen kann, da die objektbasierte Korrespondenztheorie einen aufmerksamkeitsbasierten Mechanismus vorschlägt, mit dem Objekte über Raum und Zeit hinweg verbunden werden. Dafür wurde ein modifiziertes Ternus Display verwendet, bei dem die Elemente, die innerhalb eines Frames unterschiedliche Farben hatten, so angeordnet wurden, dass sie gleichzeitig mit Gruppen- und Elementbewegungen (Gruppen- und Element-Bias) kompatibel waren. Die Aufgabe der Probanden bestand darin, ihre Aufmerksamkeit auf ein bestimmtes Element zu lenken. Die Ergebnisse zeigten, dass die Korrespondenzlösung in Richtung der Bewegungswahrnehmung verschoben wurde, die mit dem Bias des beachteten Elements übereinstimmte, was auf einen Einfluss der Aufmerksamkeit auf die Korrespondenz, in Übereinstimmung mit der objektbasierten Theorie, hindeutet. Zusammenfassend liefert diese Arbeit neue Evidenz für den Einfluss von objektbasierten Informationen auf einer mittleren Ebene der visuellen Verarbeitung und für einen aufmerksamkeitsbasierten Korrespondenzmechanismus, in Einklang mit der objektbasierten Korrespondenztheorie. Auf der Grundlage dieser Ergebnisse und früherer Forschung wird ein zweistufiges Korrespondenzprozessmodell vorgeschlagen, das sowohl die visuelle Verarbeitung auf niedriger als auch auf höherer Ebene mit einbezieht. Abschließend werden offene Fragen und Forschungsideen in Bezug auf die durchgeführten Studien und das vorgeschlagene Korrespondenzprozessmodell diskutiert.

Abstract

Our visual system enables a coherent and stable perception of the world around us. Because our environment changes constantly, the visual system has to solve the correspondence problem, that is, determine which object went where through space and time. Our visual system's ability to solve this problem is remarkable because it has to infer our three-dimensional world from information projected on our two-dimensional retina. This information is ambiguous because the same two-dimensional retinal image can correspond to multiple different images in our three-dimensional world. In addition, the world around us is constantly in motion, and objects often occlude each other. Furthermore, the correspondence solution has a strong influence on how we perceive the world around us because it determines whether we perceive objects as one and the same object over space and time, that is, in motion, and in turn determines how we interact with our environment. To enable perception (and in the end, recognition) the visual bottom-up processing starts at a low level, where features are extracted from image-based information such as luminance constrast and orientation of edges. The information is further processed at an intermediate level, in which grouping of the features takes place, leading to object-based information, e.g., representation of an object with its true color (abstracted from the illumination of the surrounding) and combined contours represented in 3D. Finally at a high level, object recognition takes place by complementing perception with top-down information, such as semantic information. Regarding the correspondence process, it is still unclear which levels of visual processing are involved in solving the correspondence problem, notably whether and, if so, how higher levels of visual processing contribute to this process. The object-based correspondence theory suggests that correspondence could be solved at an intermediate level of visual processing using object-based information. In addition, attention is suggested to be a mechanism for establishing correspondence between objects based on their identity. To test the object-based correspondence theory within this thesis, three studies, including several experiments, were conducted. The Ternus display, an ambiguous apparent motion display, was used as a measure of correspondence. The Ternus display consists of three aligned elements, shifted by one element position from one frame to the next. Depending on how correspondence is established between the elements, different types of motion can be perceived (element or group motion). Study 1 investigated whether our viusal system uses object-based information, available at an intermediate level of visual processing, to solve correspondence. To do so, a (Ponzo-like) depth illusion background was used to manipulate the perceived size of the Ternus display, while keeping the (low-level) retinal size the same. Results showed an influence of the perceived size on the correspondence solution, which is evidence that, after size constancy is processed, mid-level information is used. Study 2 addressed the question of whether object-based information that is presented using different object histories prior the presentation of the Ternus display can also influence correspondence. Results showed that this is the case, suggesting that even information that is stored in object representations and not present in the image at the moment when correspondence is solved can influence correspondence. Study 3 investigated whether voluntary attention is able to influence how correspondence is established, as the object-based theory suggests an attention-mediated mechanism to track objects across space and time. A modified Ternus display was used, in which the elements within a frame were presented in different colors in such a way that they were compatible with group and element motion (group and element bias) at the same time. Particpants' task was to direct their attention toward a specific element. The results showed that the correspondence solution was shifted toward the motion percept matching the bias of the element that was attended, therefore suggesting an influence of attention on correspondence in line with the object-based theory. In sum, this thesis provides new evidence for the influence of object-based information at an intermediate level of visual processing and for an attention-mediated correspondence mechanism in line with the object-based correspondence theory. Based on these findings and previous research, a two-level correspondence process model is proposed, incorporating both low- and higher-level visual processing. Finally, open questions and research ideas with regard to the studies conducted and the suggested model of the correspondence process are discussed.

Table of contents

List of figures

Chapter 1 Introduction

Chapter 2 Object correspondence mechanisms

Chapter 4 Illusory size determines the perception of ambiguous apparent motion (Study 1)

Chapter 5 The role of object history in establishing object correspondence (Study 2)

Chapter 6 How voluntary spatial attention influences feature biases in object correspondence (Study 3)

Chapter 7 General discussion

Chapter 1 Introduction

“Perception is real even when it is not reality” (De Bono, 2005) can be seen as an appropriate description of how we perceive the world around us. Subjectively, our percept represents the reality we rely on. This is especially evident in the visual domain, as we are “visual creatures” that rely more on our sense of sight than on other senses (e.g., auditory, tactile). For most of us it is unimaginable not to be able to see, be it on the way to work, out shopping or watching cats play. But even if it seems that our visual perception is a natural and automatic process, it is in reality complicated; what we perceive is not simply an image, as if taken by a camera, directly reflecting our environment. Our peception is instead an interpretation by our visual system. This is in part due to the inverse projection problem (see Pizlo, 2001 for a review) which is based on the fact that our three-dimensional environment is only displayed in two dimensions on the retina. Inverse projection is the process of representing our three-dimensional world based on this two-dimensional retinal information. This poses a problem because the same retinal image can be caused by an infinite number of objects in the environment. The inverse projection is therefore under-determined and, ultimately, ambiguous. Therefore, low-level information coming into the retina is not sufficient to represent the exact “reality” of our environment, which is why our perception is simply an interpretation created by our visual system. The ambiguity of information processing by the visual system can be vividly demonstrated by visual illusions like bistable figures. One famous example is the picture of the old-young lady named “My Wife and My Mother-in-Law” (Boring, 1930). This picture can be interpreted as either a young woman or an old woman. Viewers can alternate between the two, but only one is seen at a time. As the physical stimulus itself remains constant, the percept at a given time depends solely on the interpretation of our visual system. Another famous example showing our visual system's process of interpreting ambiguous input is the Necker cube (Necker, 1832). The Necker cube (see Figure 1.1, left side; compare Figure 4.1, Stepper, Moore, et al., 2020a) shows a three dimensional line drawing of a cube. All lines are solid, so there is no visual cue on how to interpret the orientation of the cube. Based on this ambiguous information, the Necker cube can be interpreted as oriented in two different ways (see Figure 1.1, right side), with two different sides of the cube being perceived as the front. Such examples clearly illustrate that our visual perception is, to a large extent, an interpretation.

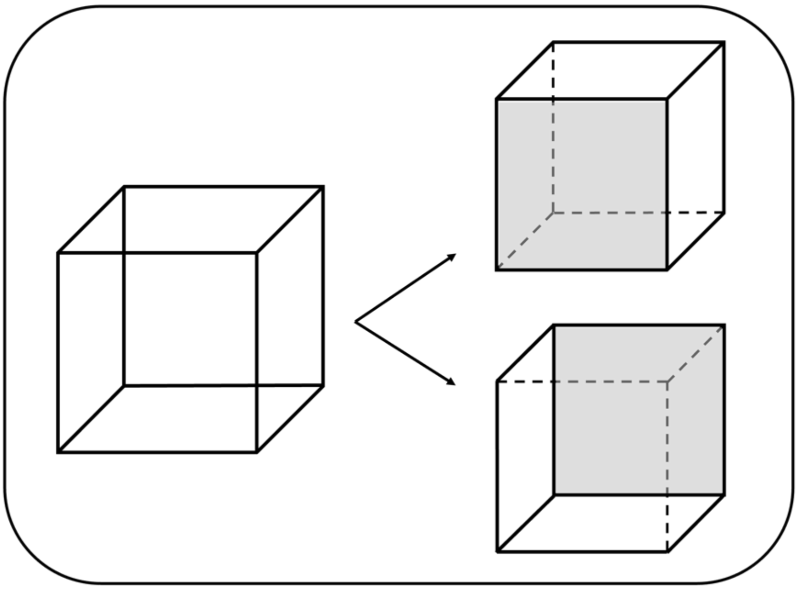

Figure 1.1: Necker Cube. Illustrated is the Necker Cube illusion on the left and disambiguated versions of its two possible percepts regarding the orientation on the right (Necker, 1832).

It is not only the inverse projection problem (see Pizlo, 2001 for a review) underlying the ambiguity of our visual system's input; there is also the object correspondence problem (Dawson, 1991; Ullmann, 1979). The world around us is dynamic – it is mostly in motion. Therefore, our visual system has the additional task of deciding which objects in a scene are the same as a moment before, meaning which objects correspond to each other across space and time. Different types of movement in dynamic scenes can lead to the correspondence problem. One type is movement of our own, such as eye movements. The challenge with regard to eye movements, or saccades, is that we move our eyes almost constantly. When eye movements are not performed to track an object in a scene, that is, holding an object fixed on the fovea, they lead to a shift in the position of that object on the retina. In addition, every eye movement leads to an interruption of the visual input, as input is suppresssed during saccades (e.g., Bridgeman, Hendry, & Stark, 1975). Our visual system therefore only receives input during fixations, times when the eye is still, leading to the input information not being continuous. In the context of eye movements, the correspondence problem specifically refers to the question of how our visual system can assign input information before and after a saccade to a single object, and through this maintain the identity of objects across eye movements (e.g., Fracasso, Caramazza, & Melcher, 2010; Richard, Luck, & Hollingworth, 2008; Tas, Moore, & Hollingworth, 2012).

Another type of movement that causes the correspondence problem (Dawson, 1991; Ullmann, 1979) is when objects in the scene around us move. A single moving object can vanish behind stationary solid objects, for example a car could drive behind a house, and if multiple moving objects are in a scene they can partly or fully occlude each other. Similar to eye movements, this situation again leads to gaps in information; input information about objects is not available at all times. In the context of objects in motion, the correspondence problem refers to the question of how our visual system determines which object went where (Dawson, 1991). No matter what kind of movement is present, whether the displacement of objects on the retina is caused by eye movements or by real movement of objects, our visual system has to resolve the correspondence problem to form a perception of our environment based on ambiguous input. The way it resolves this problem has a large impact on our perception of motion. If correspondence is established between two subsequently presented objects, these objects are perceived as one object in motion (see Figure 1.2A). In contrast, if no correspondence is established, the objects are perceived as appearing one after another (see Figure 1.2B).

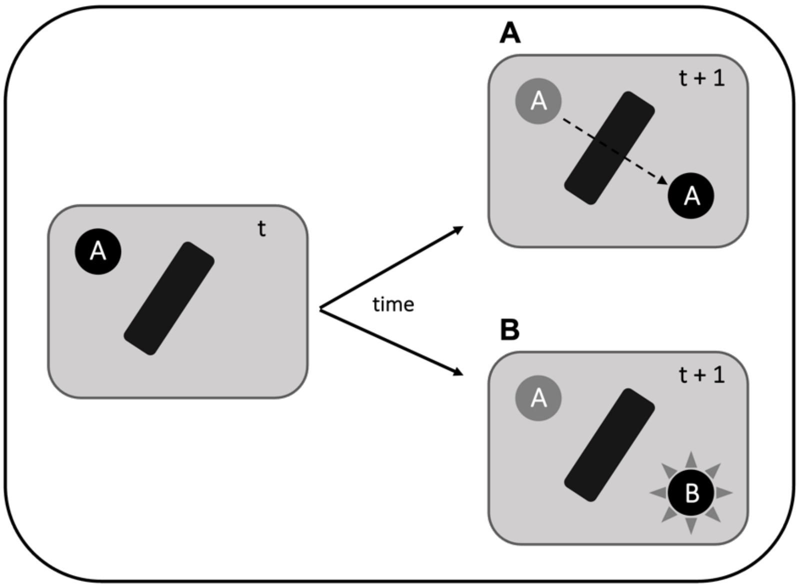

Figure 1.2: Object correspondence. Illustrated is the consequence of whether or not correspondence is established between objects. A. When correspondence is established between two successively presented objects, they are perceived as one and the same object in motion. B. When no correspondence between objects is established, they are perceived as separate objects that appear one after the other.

The focus of this thesis is on the correspondence problem (Dawson, 1991; Ullmann, 1979) in the case of object movements (Dawson, 1991). A coherent and stable representation of objects moving around us is a necessary skill for daily tasks and even for survival, with significant consequences for our decisions and actions. For example, in order to cross a road during rush hour without being hit by a car, our visual system must provide us with reliable information about which car went where so that we can find a safe way to cross the road. Therefore, establishing correspondence between objects is essential for perception of our environment, and the visual system's ability to do this based on ambiguous input is remarkable.

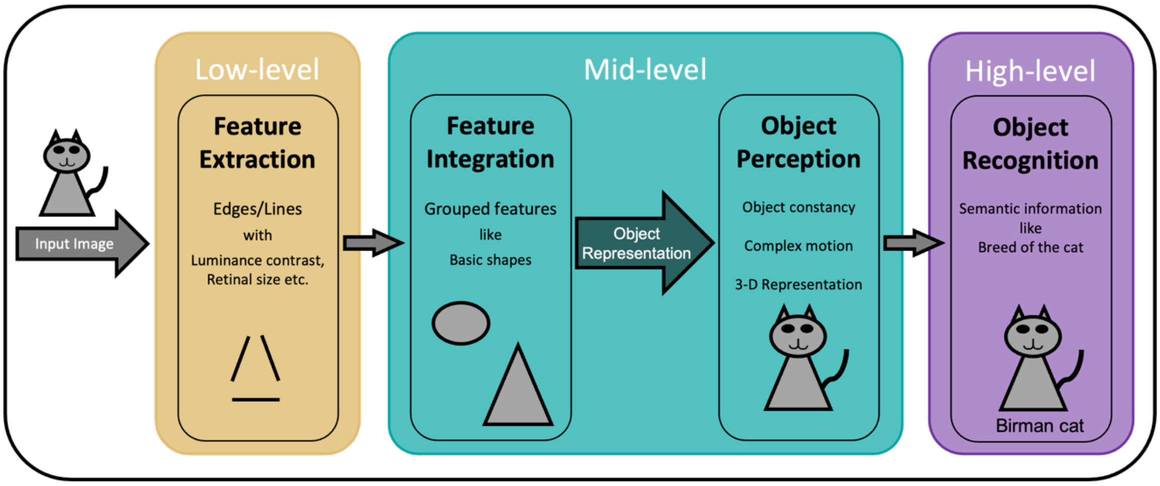

Figure 1.3: Object perception. Illustrated is the concept of the levels of visual processing required for perceiving an object, following the ventral stream (Goodale & Milner, 1992).

Research over the last few decades has investigated the means by which our visual system is able to solve the correspondence problem in the case of object movements (e.g., Casco, 1990; Green, 1986; Hein & Cavanagh, 2012; Kolers & Pomerantz, 1971; Kramer & Rudd, 1999; Navon, 1976; Petersik & Rice, 2008), but it is still unclear at which point in the visual processing stream correspondence is solved. By the time we reach our final percept (and the correspondence problem has also been solved), incoming information has been processed by our visual system at multiple cognitive levels. Image-based input at the retina is transferred via the lateral geniculae nucleaus (LGN) to the primary visual cortex (V1) in the occipital lobe (e.g., Goodale & Milner, 1992). According to the two streams hypothesis (Goodale & Milner, 1992), from there, two functionally separated paths of visual processing act in parallel. The dorsal path, leading to the parietal lobe, is suggested to guide actions or localize objects in space (often called the “where”- stream). The ventral path, leading to the temporal lobe, is suggested to support perception of faces and objects (often called the “what”-stream). As the processing level of information on these pathways gets higher (V2-V5), the available information becomes more complex, leading to so-called bottom-up processing. Object perception, through the ventral stream, can be conceptualized as follows (see Figure 1.3; see Johnson, 2018 for an overview on object perception theories): it starts with image-based information, which is relatively unprocessed low-level information extracted from the (retinal) input image, including simple features like edges and lines, their orientation, color, luminance contrast, texture, distance and retinal size. At an intermediate level of visual processing, these single features are grouped together (feature integration). This leads to basic shape information, simple object components; for example, the head of a cat is a circle. These basic shapes are then interpreted as objects by establishing object representations. At this level of object perception, perceptual completion takes place, that is, despite gaps in perception, objects are perceived as complete across space and time. The cat, for example, is perceived as a three-dimensional object that moves and is represented as a whole even if it is partly occluded (object constancy). This object-based information is then used for object recognition taking place at high levels of visual processing. In the case of the cat, it is recognized and categorized as a cat using a top-down approach and complementing the available information with higher-level semantic information. This may include knowledge about the typical shape and color of a cat, or may be more specialized such as knowledge allowing identification of the specific breed of cat. This distinction raises the question: at which level of object perception, low or higher, is correspondence established? As a thought experiment, let us imagine, as depicted in Figure 1.4, a rooster and a hen being two moving objects in our environment. One possibility is that our visual system solves correspondence between objects based on low-level information such as the spatio-temporal proximity between the objects. In this case, correspondence is established between objects which are closer together across time (see Figure 1.4A). The other possibility is that correspondence is established based on more processed, higher-level object information obtained after object representation is established. In this case, correspondence is established between each of the objects, maintaining their identity (see Figure 1.4B). Intuitively, both levels are possible. On the one hand, it makes sense that objects are perceived as corresponding when they are in spatio-temporal proximity, given that objects in our environment normally do not vanish and suddenly reappear far away. On the other hand, objects in our environment also do not suddenly change their identity; a hen does not spontaneously become a rooster.

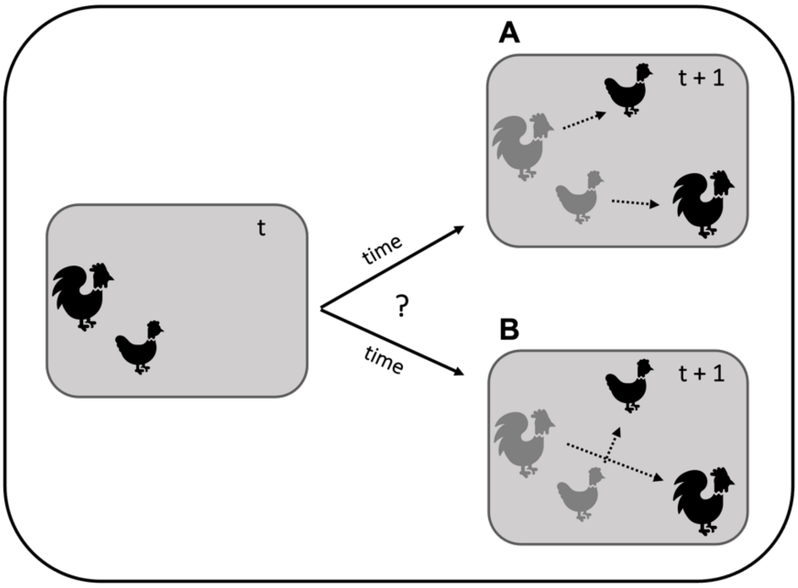

Figure 1.4: Object correspondence problem. Illustrated is the consequence of solving correspondence between multiple objects at different levels of visual processing. A. Correspondence solution based on low-level information, like the space between objects. B. Correspondence solution based on higher-level information, like the identity of objects.

The level of visual processing at which correspondence is solved remains unknown, notably to which extent higher levels, in particular an intermediate level of visual processing, are involved. For a long time, research investigating the correspondence problem (Dawson, 1991; Ullmann, 1979) focused primarily on low-level factors such as the influence of motion-based factors, like spatio-temporal continuity based on extracted features, or the influence of grouping at the level of feature integration. In comparison, research regarding the influence of higher-level factors is relatively recent and not as systematic. In additon, research regarding the influence of higher-level factors has mainly focused on the level of object recognition, primarily semantic knowledge. The aim of this thesis was therefore to systematically investigate the influence of factors at an intermediate level of visual processing, after object representations have been established (see Figure 1.3, referred to as *Mid-level*) but before semantic knowledge comes into play. Evidence about whether correspondence can be established at this level is necessary to differentiate existing theories on how object correspondence is solved, since most theories are insufficient to explain possible factors influencing correspondence at an intermediate level of visual processing. In chapter 2, several existing theories are presented, each suggesting different mechanisms of how correspondence could be solved, and each based on different levels of visual processing. Chapter 3 describes the derivation of the precise questions that

I have empirically investigated in order to better characterize mid-level influences on object correspondence. The following three chapters (4-5-6) contain the studies investigating these questions, and chapter 7 presents a discussion and summary of the results of this work.

Chapter 2 Object correspondence mechanisms

How do we know which object went where? This fundamental question in the study of how our visual system solves object correspondence (Dawson, 1991; Ullmann, 1979) has interested many researchers (e.g., Casco, 1990; Green, 1986; Hein & Cavanagh, 2012; Kolers & Pomerantz, 1971; Kramer & Rudd, 1999; Navon, 1976; Petersik & Rice, 2008). Compared to other research topics, however, such as visual attention or visual search, very few factors influencing the correspondence solution have been investigated, and only a few theories have been developed that describe possible mechanisms by which our visual system is able to solve this task. The suggested correspondence theories, which will be described in more detail below, differ in what kind of information is used for solving correspondence and at which level of visual processing it occurs (low, intermediate, or higher levels). In this chapter, I will describe and discuss these mechanisms by dividing them into three categories based on the factors influencing correspondence as well as the proposed level of visual processing (see also Hein, 2017): (1) motion-based theories, (2) feature-based theories, and (3) object-based theories.

2.1 Motion-based theories

One very influential approach to examining the establishment of object correspondence focuses mainly on spatio-temporal information (Adelson & Bergen, 1985; Flombaum & Scholl, 2006; Kahneman et al., 1992; Scholl, 2001; van Santen & Sperling, 1985; Werkhoven et al., 1994). Accordingly, correspondence is established between objects that are closer in space and time rather than between objects that are more separated in space and time (see, e.g., Figure 1.4A). This approach can be considered a low-level approach because information already available at a low level of visual processing is used to solve correspondence (see Figure 2.3, left side).

Theories that suggest that object correspondence depends on spatio-temporal information are called motion-energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993). The mechanisms underlying motion-energy models are low-level motion detectors (e.g., Reichardt detectors; Reichardt, 1961; see Figure 2.3). A motion detector receives input from two specific locations of the retinal image in the form of a luminance pattern with a specific contrast. Based on spatio-temporal activation changes, that is, the contrast change at these subsequent locations over time, motion energy is calculated. Each motion detector is tuned to a specific motion direction and speed. This means that the detector will be most activated, sensing the most motion energy, by spatio-temporal activation changes which fit its particular tuning. In the end, motion is perceived in the direction the most motion energy occurs, the direction the "winning" detector is tuned for. According to this perceived motion, correspondence between the objects is established.

Empirical evidence for motion energy models comes from apparent motion displays (Wertheimer, 1912). In apparent motion displays, two stationary objects can be perceived as one object in motion if the timing and spacing between them is appropriate (Kolers, 1972; Korte, 1915; Wertheimer, 1912). An everyday example of this phenomenon is flip-books. Despite no physical motion being present, motion is perceived as long as the individual images are presented in rapid succession. This also means that as long as motion is seen, correspondence has been established between the subsequently presented objects. Such single object apparent motion displays are unambiguous with regard to the correspondence solution as either motion between the two single stimuli is seen or not (Cavanagh et al., 1989; Kolers & Pomerantz, 1971), that is, correspondence has either been established or it has not. In contrast, the correspondence problem becomes more interesting in ambiguous apparent motion displays. In these, different motion percepts are available, depending on how correspondence is solved between multiple objects presented at the same time (Burt & Sperling, 1981; Navon, 1976; Ullmann, 1979). One example of an ambiguous apparent motion display is the motion quartet (Navon, 1976; von Schiller, 1933). Two stimuli are presented at the diagonally opposite edges of a square, alternating with two stimuli presented at the other two edges of that square. Motion between the alternating stimuli can be perceived either as horizontal or vertical. The motion direction perceived has been shown to depend on the spacing and timing between the alternating stimuli. For example, reducing the horizontal spacing leads to more horizontal motion percepts (e.g., Hock, Kelso, & Schöner, 1993; von Schiller, 1933), showing the importance of spacing information for the correspondence solution. Another prominent example of an ambiguous apparent motion display, which has been used by many researchers investigating motion energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993), is the Ternus display (Pikler, 1917; Ternus, 1926). In its original version, it consists of two elements horizontally aligned next to each other, shifted about one element position from one frame to the next, with a blank screen in-between the stimulus frames (see

Figure 2.1). Depending on how correspondence is solved, the elements are perceived either as moving together (group motion) or as one element jumping across the other (element motion). The type of motion perceived shows between which objects correspondence has been established. Like for the motion quartet (Navon, 1976; von Schiller, 1933), in the Ternus display (Pikler, 1917; Ternus, 1926) the timing between the two frames has been shown to strongly influence how correspondence is solved and apparent motion is perceived. With an increasing time between the presentation of the two frames (interstimulus interval [ISI]), group motion percepts also increase (Pantle & Petersik, 1980; Petersik & Pantle, 1979). Such findings are in line with motion-energy models, suggesting the importance of spatio-temporal information (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993). In addition, the findings are in line with the idea of simple motion detectors (e.g., Reichardt detectors; Reichardt, 1961), which calculate motion direction based on only two spatio-temporal locations. This is exactly the information available from apparent motion displays, which consist of static images.

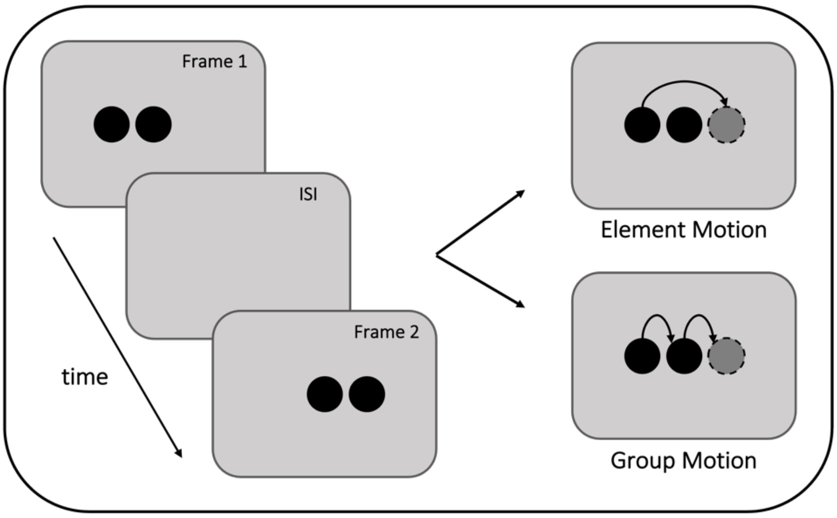

Figure 2.1: Ternus display. The left side shows the presentation sequence of the single frames. The right side shows the two possible motion percepts. Element motion, where one element is perceived as jumping across the other. Group motion, where each element is perceived to move one element position to the right, i.e., the whole group is perceived as moving together.

Another approach of motion-based theories, also emphasizing the importance of spatio-temporal information, but proposing a different mechanism for solving correspondence is the object-file framework (Kahneman et al., 1992). Whereas motion-energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al.,

1993) suggest a simple mechanism (motion detectors) that can be easily implemented at the neural level, the object-file framework (Kahneman et al., 1992) suggest that correspondence is solved at a more cognitive level of object-files: episodic (i.e., temporary, present in the here and now) representations of objects. Object-files are only indexed by spatio-temporal information, meaning that they are tracked based on spatio-temporal information to solve the question of which object went where. An object is perceived as persistent over space and time when it is recognized as an object coming from a previous position. If the spatio-temporal information is too separated, it is recognized as a new one and a new object-file is created. In addition to spatio-temporal information, feature information about the object is also assigned to the object-file, but this information does not play a role in establishing object correspondence because features can change over time and are therefore not a relevant information carrier for the correspondence solution (Kahneman et al., 1992). This theory is special in that object-files are created at a higher-level of visual processing, after object representations has been established (see Figure 2.3 in which object-files go across all processing levels), but for solving correspondence, only low-level spatio-temporal information, which is information about motion, is used. This is why I have classified this as a motion-based theory.

Empirical evidence for the object-file framework comes from the object-reviewing paradigm (Kahneman et al., 1992). In the object-reviewing paradigm, two squares are presented equidistant left and right from fixation, and each contains a preview letter. After the letters disappear the squares move continuously, clockwise or counterclockwise until they stop below and above fixation. A target letter then appears in one of the squares, which participants have to name as quickly as possible. The main finding in this paradigm is that the participants have faster naming latencies when the target letter was the same as the preview letter and, most importantly, when it appeared at the same object as in the preview (object-specific preview benefit [OSPB]). This OSPB is seen as an indicator of the establishment of object correspondence and as being inferred on the basis of the spatio-temporal continuity (Kahneman et al., 1992). Empirical evidence regarding the importance of spatio-temporal information also comes from other paradigms using continuous motion displays. One example is the tunnel effect (e.g., Burke, 1952; Michotte, Thinés, & Crabbé, 1991). In the tunnel effect, an object moves continuously and disappears behind an occluder (e.g., in a tunnel). When the object reappears from behind the occluder and it is perceived as the same object, object correspondence has been established. On the contrary, when it is perceived as a new object appearing, implying that no motion is perceived between the two objects, object correspondence has not been established. It has been shown that as long as an object disappearing behind an occluder reappears after an appropriate amount of time considering the space, it is perceived as the same object moving (e.g., Burke, 1952; Michotte, Thinés, & Crabbé, 1991). Thus, using continuous motion displays, it has been shown that spatio-temporal information plays a crucial role for solving object correspondence, as suggested by the object-file framework (Kahneman et al., 1992), as well as by motion energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993).

The influence of spatio-temporal factors is even considered by some researchers to be so important that a spatio-temporal priority (Flombaum et al., 2012; Scholl, 2007) is suggested for the correspondence process. According to this view, spatio-temporal factors are not only an important factor for solving correspondence and therefore a central component of theories, as in the object-file framework (Kahneman et al., 1992) or motion energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993), but are considered the only relevant factor. Evidence for this view comes from studies showing that other information, like features such as color or orientation, have little or no impact on how correspondence is solved. Using the object-reviewing paradigm, for example, Kahneman et al. (1992) did not find an advantage in naming latency when the letters had the same color compared to different colors. Further, Mitroff and Alvarez (2007) also did not find evidence that object features could be used for establishing object files. They modified the object reviewing paradigm by replacing spatio-temporal information (i.e., the motion period) through feature information. More precisely, the objects were distinct in their combination of features, and a blank interval was used instead of the motion period. Their results suggest that object correspondence could not be established based on object features alone. In addition, there are phenomena like the tunnel effect (e.g., Flombaum & Scholl, 2006), for which it has been shown that motion between the disappearing and appearing object can be perceived even when the object changes its appearance. Supporting these findings, there is evidence that apparent motion can be perceived between objects as long as the timing and spacing is appropriate, despite changing their shape or color (e.g., Burt & Sperling, 1981; Kolers & Pomerantz, 1971; Navon, 1976; Ullmann, 1979).

In sum, evidence for the importance of spatio-temporal information for solving object correspondence has been shown for a large variety of paradigms, using continuous motion displays (e.g., tunnel effect; e.g., Burke, 1952; Michotte, Thinés, & Crabbé, 1991; or object-reviewing paradigm; Kahneman et al., 1992) as well as apparent motion displays (e.g., ambiguous apparent motion displays, like the motion quartet, Navon, 1976; von Schiller, 1933, or the Ternus display, Pikler, 1917; Ternus, 1926). Theories like the object-file framework (Kahneman et al., 1992) or motion energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993) both highlight spatio-temporal information as the most important factor for establishing correspondence. The mechanisms underlying these theories, however, are located at different levels of visual processing. While motion energy models suggest low-level motion detectors, correspondence in the object-file framework is solved at the higher level of object representations. However, these theories can be grouped together as motion-based because their correspondence solutions are based mainly on the change of location over time.

2.2 Feature-based theories

A second approach to object correspondence, though it also emphasizes the importance of information available at lower levels of visual processing (see Figure 2.3, feature extraction and feature integration), proposes correspondence establishment by image-based feature information (e.g., color or size of an object), directly extracted from the retinal input.

Several correspondence theories highlighting the influence of image-based feature information were developed using the Ternus display (see Figure 2.1; Pikler, 1917; Ternus, 1926). One such theory is the persistence theory (Breitmeyer & Ritter, 1986a, 1986b) in which visual pattern persistence determines how correspondence is established. Visual pattern persistence describes the phenomenon that briefly presented visual stimuli, like the individual Ternus frames, remain visible for a short time after their physical offset. The persistence theory suggests that the strength of persistence, how long the first Ternus frame persists, determines how likely it is that the central Ternus elements are perceived as stationary, leading in turn to element motion perception. More precisely, when the temporal gap between the Ternus frames, especially between the spatially overlapping center elements, is bridged by visual persistence, element motion is more likely to be perceived (Breitmeyer & Ritter, 1986a, 1986b). The strength of visual persistence has been shown to depend on spatio-temporal information, being reduced with increasing presentation duration (e.g., Di Lollo, 1977), as well as on feature information, being reduced with increasing size (e.g., Breitmeyer, Levi, & Harwerth, 1981) or increasing contrast (Bowling

& Lovegrove, 1981). Therefore, this theory is assigned to low-level visual processing in Figure 2.3.

Empirical evidence for the persistence theory (Breitmeyer & Ritter, 1986a, 1986b) comes from studies in which factors affecting the strength of visual persistence in the Ternus display were manipulated. It has been shown that factors like frame duration (Petersik & Pantle, 1979), as well as low-level features like the size of the Ternus elements (Breitmeyer & Ritter, 1986a, 1986b) and the luminance contrast of the Ternus elements (Alais & Lorenceau, 2002) influence how correspondence is established. For example, Alais and Lorenceau, (2002), showed that more group motion is perceived with increasing contrast of the Ternus elements. Breitmeyer and Ritter (1986a, 1986b) showed that with increasing size of the Ternus elements (the distance between elements was held constant) more group motion was perceived. Both findings are in line with the predictions of the persistence theory, as the strength of visual persistence decreases with increasing contrast (Bowling & Lovegrove, 1981) and increasing size (Breitmeyer & Ritter, 1986a, 1986b) leading to a lower temporal bridging between the two Ternus frames making the element motion percept less likely.

Another subset of theories accounting for the influence of low-level features concerning the Ternus display includes the grouping theories (e.g., Alais & Lorenceau, 2002; He & Ooi, 1999; Kramer & Yantis, 1997) in which correspondence is based on grouping strength between the Ternus elements. The basic idea of these theories is that correspondence is more likely to be established between objects that are similar in terms of their features, including orientation, luminance, size, or color. For example, Kramer and colleagues (Kramer & Rudd, 1999; Kramer & Yantis, 1997), suggest that the correspondence solution in the Ternus display, whether group or element motion is perceived, depends on the temporal grouping across Ternus frames: The more similar the elements across frames are, especially the central elements on the positions that are present over both frames, the more element motion is perceived. Moreover, correspondence is also thought to depend on the spatial grouping within a Ternus frame: The more similar the elements within a frame are, the more group motion is perceived. Considering that, for such theories, a grouping of the available feature information takes place, correspondence could be classified as taking place at somewhat further along the visual processing stream (see Figure 2.3) than the persistence theory (Breitmeyer & Ritter, 1986a, 1986b).

Like the persistence theory (Breitmeyer & Ritter, 1986a, 1986b), evidence for grouping theories comes directly from the Ternus display. It has been shown that the similarity of elements within a frame (Alais & Lorenceau, 2002; Wallace & Scott-Samuel, 2007), as well as their similarity across frames (Casco, 1990; Dawson et al., 1994; Hein & Cavanagh, 2012; Hein & Moore, 2012; Kramer & Yantis, 1997; Petersik & Rice, 2008) can influence the correspondence solution. Alais and Lorenceau (2002), for example, used Gabor patches, oriented either horizontally or vertically, as Ternus elements (see Figure 2.2A). The authors showed that more group motion was perceived when they were oriented horizontally compared to vertically. This suggests a stronger grouping between horizontally-oriented Gabor patches due to collinearity, therefore suggesting that feature information within a frame can influence how correspondence is established. Another example comes from Hein and Moore (2012) who created Ternus displays biased toward either element or group motion by manipulating the features of the Ternus elements across frames (see Figure 2.2B). For example, by using two colors within one frame and keeping the order identical across frames, they created a group bias. In contrast, changing the order across frames so that the central elements maintain a constant color (colors of the Ternus elements from left to right: e.g., frame 1: blue, green, blue; frame 2: green, blue, blue; see Figure 2.2B bottom), they created an element bias. Hence, Hein and Moore (2012) showed that the percept in the Ternus display follows the bias introduced by different features (e.g., polarity, color, orientation, or luminance): more group motion was perceived with the group bias and more element motion was perceived with the element bias. This suggests that feature information can influence correspondence across frames.

In addition, evidence for the influence of low-level features on correspondence has also been found using apparent motion displays other than the Ternus display (e.g., spatial frequency: Green, 1986; Watson, 1986; object shape: Berbaum, Lenel, & Rosenbaum, 1981; Shechter, Hochsteinn, & Hillman, 1988; phase: Sekuler & Bennett, 1996; orientation and color: Green, 1986, 1989). For example Green (1986), used Gabor patches differing in their spatial frequency, orientation or phase as stimuli. One frame contained four Gabors, two of which were identical. In the first frame, these were arranged at the endpoints of a fictive cross, whereby the opposite ones were identical. Over three further frames the position of each Gabor was rotated clockwise by 45 degrees. The idea behind this display was that the distance from one Gabor to its neighbor to the left and to the right in the next frame was identical and thus there was no spatio-temporal determinant for the perceived direction of motion (clockwise or counterclockwise). However, if the feature of the Gabor determines correspondence, motion in the direction of the Gabor with the same feature (clockwise) should be perceived. This was the case for spatial frequency and orientation, which were therefore suggested to be features influencing the correspondence solution (Green, 1986). All these studies using apparent motion displays support both the grouping theories (e.g., Alais & Lorenceau, 2002; He & Ooi, 1999; Kramer & Yantis, 1997) as well as the persistence theory (Breitmeyer & Ritter, 1986a, 1986b) suggesting that correspondence is established based on the strength of grouping or visual persistence which can be influenced by low- and mid-level feature information.

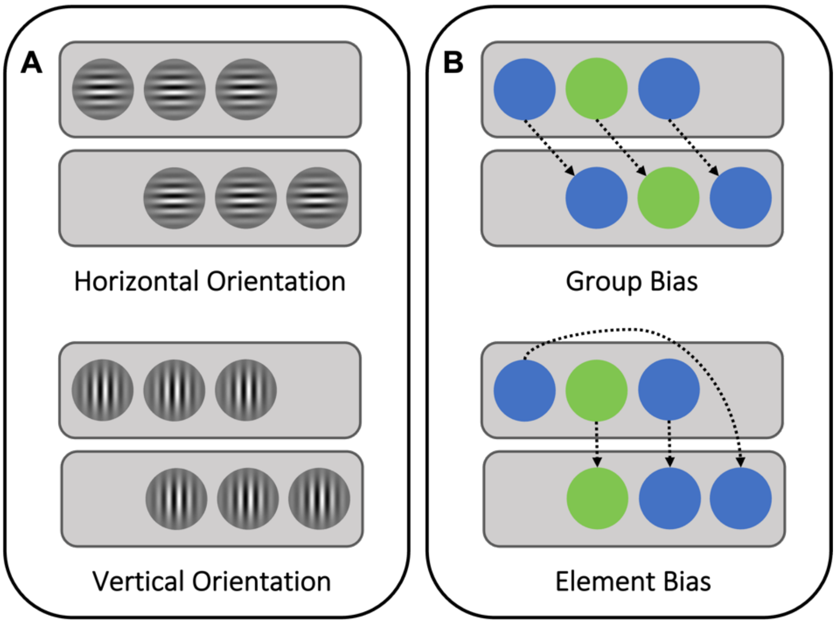

Figure 2.2: Feature influence in the Ternus display. A. The stimuli used by Alais & Lorenceau (2002). The Gabor patches were oriented either horizontally (top) or vertically (bottom). B. The stimuli used by Hein & Moore (2012). The Ternus elements had different colors ordered across frames compatible either with the group motion percept (top, group bias) or the element motion percept (bottom, element bias). For both studies, the Ternus display on the top led to more group motion percepts compared to the Ternus display on the bottom.

Evidence for the influence of features also comes from other paradigms using continuous motion, like the object-reviewing paradigm (Kahneman et al., 1992) or the perceived causality display (Michotte, 1963; Moore et al., 2020). Hollingworth and Franconeri (2009), for example, extended the object-reviewing paradigm by using an occluder at the end of the motion path of the objects. The objects vanished behind the occluder, and after the occluder disappeared and the objects were visible again, participants had to answer as fast as possible whether the shapes that appeared in the objects were the same as in the preview. Results showed the typical object-specific preview benefit (OSPB) when the shapes were presented in the same objects as in the preview, indicating the establishment of correspondence despite the occlusion. To investigate the influence of features, the color was manipulated during the occluding phase. The authors (Hollingworth & Franconeri, 2009) showed that the OSPB was affected by this feature manipulation, with a larger OSPB when the colors matched. Further evidence for the influence of features comes from Moore, Stephens, and Hein (2010), who introduced an abrupt change of the object's color during the motion path of the object-reviewing paradigm. This feature change eliminated the OSPB. So, in both studies, the OSPB was influenced by feature information. As the OSPB is seen as an indicator of the establishment of object correspondence (Kahneman et al., 1992), the authors (Hollingworth & Franconeri, 2009; Moore et al., 2010) of both studies suggest that not only spatio-temporal information but also feature information is used to establish correspondence between objects. The importance of feature information for solving correspondence is also highlighted in a recent study from Moore et al. (2020). In their study they used a causality display in which two objects were presented. The first one moved in the direction of the second one until they completely overlapped with each other. Then, one of them moved away and the other stayed at its position. This display is ambiguous and can be perceived in two different ways. Most of the time, it is perceived as the first object passing the second one, which remains stationary, but it can also be perceived as the first object stopping and launching the second into movement. To investigate the influence of features, the objects’ features were manipulated in order to bias the percept toward one or the other option. In the neutral condition, both objects were identical, and in the experimental conditions the two objects differed in one feature dimension (size, polarity, color, or isoluminant color). The first object, for example in the case of polarity presented in black, moved toward the second object until the complete overlap of both objects. To bias the percept toward launching, the second moving object was presented in a different polarity from the object that had initially moved, in the example above it would have been white. In contrast, to bias the percept toward passing, the polarity of the second moving object was in the same polarity as the initial moving object, in our example black. For different feature manipulations (size, polarity, color, or isoluminant color) the results showed that for a feature bias compatible with passing, more passing percepts occurred compared to a feature bias compatible with launching. Moore et al., (2020) therefore suggest that features influence how motion in this causality display is perceived and therefore how correspondence is established. To account for this new finding the object-file framework (Kahneman et al., 1992) could be extended: perhaps object-files are not only indexed by spatio-temporal information but also by feature information (Moore et al., 2020).

In sum, many studies have shown that our visual system uses feature information, such as size, orientation, color, luminance, or polarity, to establish correspondence between objects. This has been shown with a variety of paradigms, using ambiguous apparent motion, like the Ternus display (Alais & Lorenceau, 2002; Hein & Moore, 2012), as well as continuous motion, like the object-reviewing paradigm (Hollingworth & Franconeri, 2009; Moore et al., 2010) or the causality display (Moore et al., 2020), which underscores the importance of features for the correspondence process. These findings have been incorporated in correspondence theories that either extend existing theories, like the object-file framework regarding the influence of features in continuous motion (Moore et al., 2020), or in new theories that have been developed especially to explain the influence of feature information, like the grouping theories (e.g., Alais & Lorenceau, 2002; He & Ooi, 1999; Kramer & Yantis, 1997) or the persistence theory (Breitmeyer & Ritter, 1986a, 1986b) for the Ternus display. In the following sections, motion-based and feature-based theories will be grouped under the umbrella term "image-based theories", as both theories focus on the influence of information that comes directly from the retinal image (e.g., spatio-temporal or feature information) and that has been processed at most up to an early intermediate level of visual processing (see Figure 2.3).

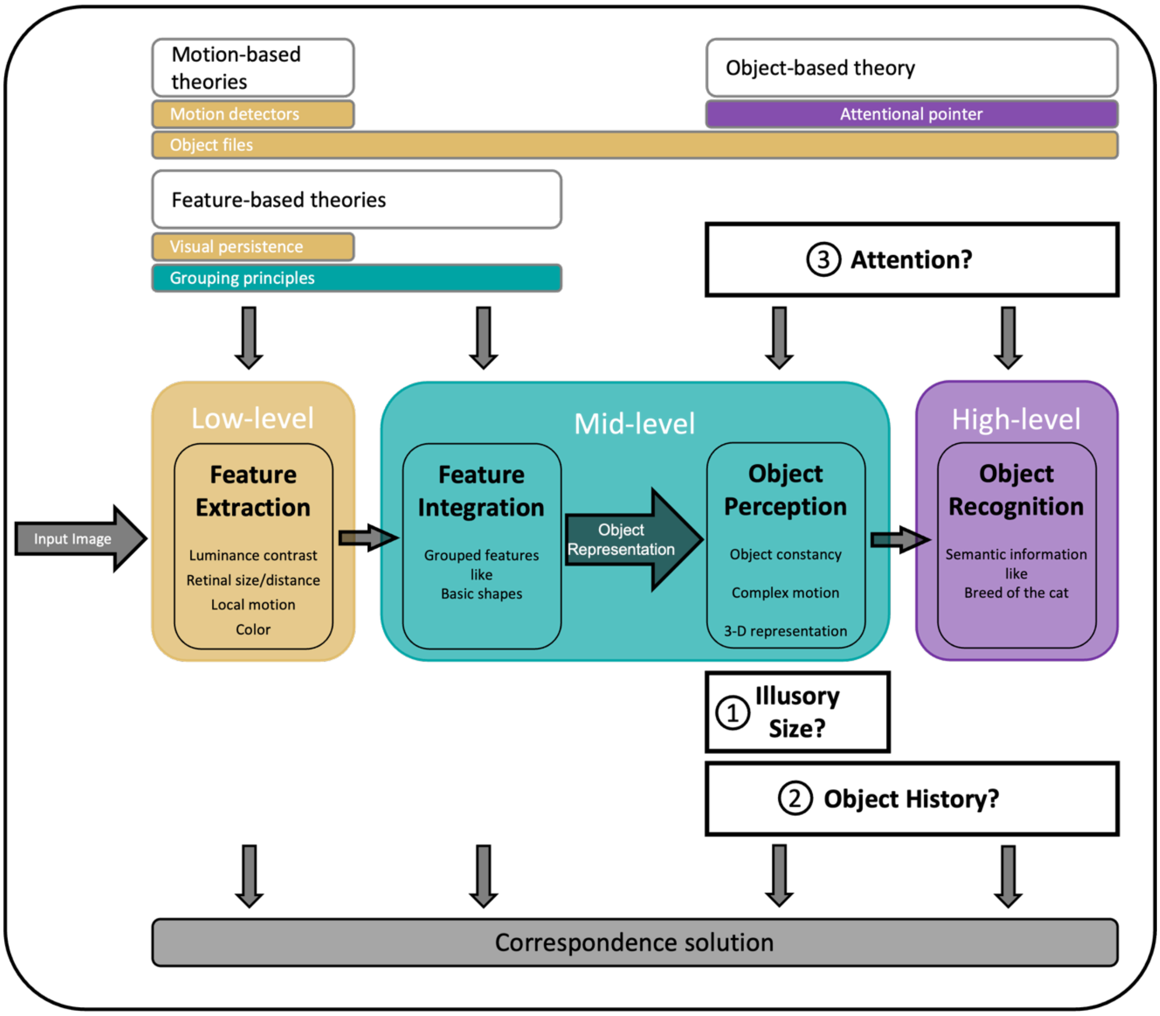

Figure 2.3. Correspondence mechanisms. Diagram showing at which processing levels of object perception (compare Figure 1.3) the different correpondence theories and their mechanisms are located. While image-based theories (motion-based and feature-based theories) use information available from the retinal image, the object-based theory uses more highly processed information at the object-level for solving correspondence. Numbered black boxes highlight open questions that still exist; these were investigated in the context of this dissertation and will be explained in Chapter 3.

2.3 Object-based theory

A third approach to understanding the correspondence process is in terms of objects (Hein & Cavanagh, 2012; Hein & Moore, 2014). The idea behind this approach is that establishing correspondence is based on object-based information, that is, between the object representations of the objects in a scene. In contrast to image-based theories, which include both motion-based theories (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992) and feature-based theories (e.g., Kramer & Yantis, 1997), correspondence is purportedly based on more highly-processed information, after the identity of an object has been established. Such information can include knowledge about typical attributes defining a specific object, like when comparing a rooster to a hen (see Figure 1.3B). Figure 2.3 illustrates this approach, in which image-based information like spatio-temporal information and image-feature information, available at the level of feature extraction and feature integration, is further processed up to the object level. The object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014) suggests that, for all objects in a scene, episodic object representations are created, and these representations store all available information about an object. Based on this information, correspondence is established between those objects that are most similar to each other over space and time. Establishing correspondence or not is then suggested to change the object representations, possibly via an object-mediated updating process (Enns et al., 2010; Lleras & Moore, 2003; Moore et al., 2007; Moore & Lleras, 2005). More precisely, when correspondence is established between an actual object and one from a moment ago, they are mapped to the same object representation, which is updated according to the newly sampled information, such as that about the actual location. In contrast, when correspondence is not established, a new object representation is created for the actual object. A mechanism, suggested by Hein and Cavanagh (2012), by which the most similar perceived objects in a scene could be tracked may be attention, more precisely, attentional pointers (Cavanagh, 1992; Cavanagh et al., 2010), which could connect image elements identified as belonging to the same object across space and time.

Evidence for the object-based account (Hein & Cavanagh, 2012; Hein & Moore, 2014) comes from studies showing an influence of information on the establishment of correspondence at a higher/intermediate level of visual processing, beyond simple image-based information. It has been shown, for example, that the perceived attributes of objects (He & Nakayama, 1994; Hein & Moore, 2014), in contrast to their physical attributes, or the context in which the display is embedded (He & Ooi, 1999; Ramachandran & Anstis, 1983b; Yantis, 1995) are taken into account to establish correspondence and perceive motion. In the case of perceived attributes of an object, He and Nakayama (1994), for example, have shown that amodal completion, inferring how surfaces extend behind occluding surfaces, takes place before correspondence is established. To investigate this, in their first experiment, they used a motion quartet stereogram with a black “L” and an inverted black “L” in each frame, each adjacent to a white square. The stereograms for the two experimental conditions were constructed in such a way that the black Ls were either perceived as being in front (front condition) or in the back of the white squares (back condition). In the back condition, amodal surface completion takes place, whereby the Ls are completed to black squares. Physically, at an image-level, the L stimuli are the same in both conditions, which should lead to the same motion percept (vertical or horizontal). A difference in perceived motion should only occur if correspondence takes place after amodal surface completion. Indeed, the authors (He & Nakayama, 1994) found more horizonal motion in the front condition, since horizontally the Ls had the same orientation, compared to the back condition. This suggests that the *perceived* surface, at a higher level of visual processing, is used for solving correspondence, in contrast to only image-based feature information. As another example, Hein and Moore (2014) manipulated the perceived attributes (lightness or size) of a three-element Ternus display separately from its physical attributes (luminance or retinal size). In the case of lightness, they used a filter in front of a Ternus element. This filter had a fixed position and was in front of the center element in the first Ternus frame and in front of the first element in the second Ternus frame. At a physical level, image-based luminance, such a filter should introduce a bias toward the element motion percept (compare color element bias Figure 2.2B; Hein & Moore, 2012), since the elements with the filter have the same physical luminance. In contrast, at the higher level of perceived lightness, due to color constancy, all elements should be perceived as being the same behind a filter, hence without a bias toward one of the motion percepts. In the case of size, they used an occluder in front of a Ternus element, which had a hole with a diameter smaller than the Ternus elements. Analogous to the other experiment, at a physical level, this occluder should introduce a bias toward the element motion percept since the elements behind the occluder have the same (smaller) retinal size, due to the reduced visibility of the elements behind the occluder. In contrast, at the higher level of perception, due to size constancy, all elements should be perceived as being the same without a bias toward one of the motion percepts. For both experiments, results showed motion percepts more similar to those of a Ternus display without bias, and therefore the authors suggested that perception depended on the perceived attributes rather than the physical ones (Hein & Moore, 2014). In the case of the motion context, Ramachandran and Anstis (1983), for example, showed that the context in which the motion quartet (Navon, 1976; von Schiller, 1933) was embedded influenced the motion percept. They found that if several motion quartets were presented simultaneously, the same motion direction (horizontal or vertical) was perceived in all of them. This suggests an influence of the context on how correspondence is established. Such findings indicate that information processed up to higher-level object representations is used for establishing object correspondence.

Furthermore, there are studies showing an influence on the correspondence solution of types of information that are available at even higher levels of visual processing. It has been shown that lexical (Chen & Zhou, 2011) or semantic information (Hsu et al., 2015; Yu, 2000) can influence how correspondence is established in the Ternus display, as well as in other paradigms (Ramachandran et al., 1998; Tse & Cavanagh, 2000). Hsu et al. (2015), for example, replaced the elements of a Ternus display with frog images in order to investigate whether the semantic information associated with frogs (i.e., that they jump) has an influence on the motion percept. They found more element motion percepts for the condition in which the frogs were facing in the motion direction compared to when they were facing backwards. The authors (Hsu et al., 2015) suggest that this is due to the semantic knowledge that frogs can only jump forwards and not backwards, indicating that semantic information is used for establishing correspondence. These studies suggest that object-based information available from different higher levels of visual processing is used for solving correspondence.

Besides to higher-level object information, some studies indicate that object correspondence can be influenced in a top-down fashion by even more high-level factors. Already Wertheimer (1912) suggested in his early studies on movement perception in apparent motion displays that attention could play a role in solving correspondence. He found that for ambiguous apparent motion, directing attention toward one of the possible motion percepts can influence the correspondence solution in favor of this percept. Much more recent studies also indicate that attention influences how correspondence is solved (Aydın et al., 2011; Kohler et al., 2008; Suzuki & Peterson, 2000; Xu et al., 2013). Aydın et al. (2011), for example, investigated the influence of attention on the Ternus display by manipulating the availability of attention while participants had to judge the motion percept using a dual task paradigm. In one of the two experimental conditions, the attention available for the main task, the Ternus motion judgment, was reduced by requiring participants to perform an additional task (counting) at the same time. Results showed more group motion percepts when attention was fully available to the Ternus display compared to when attention was distracted by a second task. The authors (Aydın et al., 2011) suggest that more attention is necessary to perceive group motion compared to element motion in the Ternus display. Moreover, Xu et al. (2013) tracked the spatial shift of attention using EEG and investigated how this shift is related to the motion perceived in an ambiguous apparent motion display, the motion quartet. Their results showed that when horizontal motion was perceived, there was a synchronous shift of spatial attention in the same direction. The authors (Xu et al., 2013) suggest that this is evidence for a link between attention and object correspondence. In addition to the influence of attention on correspondence, studies also indicated that memory may be involved in determining how correspondence is solved (Hein et al., 2021; Kohler et al., 2008; Scocchia et al., 2013). Scocchia et al. (2013), for example, investigated whether memory content can influence the motion percept in an ambiguous continuous motion display: a motion sphere, in which a cloud of small dots can be perceived as rotating clock- or counterclockwise. They disambiguated the perceived direction of the motion sphere by creating a version with clear clock- or counterclockwise motion and presented it prior to the ambiguous one. Their results showed a tendency to perceive the motion direction of the disambiguated sphere in the ambiguous version. The authors (Scocchia et al., 2013) therefore suggest that visual working memory content can influence how correspondence is solved and, in this case, which motion direction is perceived. Using the Ternus display, Hein et al., (2021) also showed an influence of visual working memory content on correspondence. The three-element Ternus display consisted of different colored elements within a frame. The order of these elements across frames was such that the Ternus display contained a bias toward element motion and a bias toward group motion at the same time (competitive Ternus display, Hein & Schütz, 2019). Prior to presenting the Ternus display, a square either matching one of the colors of the Ternus elements or not matching any was presented. Participants had to memorize this color and judge the Ternus motion. The results showed that more group motion was perceived when the memorized color matched the group bias compared to when the memorized color matched the element bias, suggesting that an object feature that is represented in short-term memory can influence correspondence (Hein et al., 2021). Studies regarding the influence of attention and memory therefore indicate that, in addition to higher-level object-based information, other higher-level factors not part of the object itself can influence the correspondence process in a top-down manner.

In sum, there is evidence that information at higher levels of visual processing influence how object correspondence is established. Studies have shown that further processed information, that means higher-level information such as the perceived attributes of an object (He & Nakayama, 1994; Hein & Moore, 2014) or the inclusion of motion context (He & Ooi, 1999; Ramachandran & Anstis, 1983b; Yantis, 1995) can be used for solving correspondence. In addition, lexical (e.g., Chen & Zhou, 2011) and semantic information (Hsu et al., 2015; Yu, 2000), available at an even higher level of visual processing after the identity of objects has been established, can also be used for the correspondence solution. Finally, it has been shown that even top-down information, such as voluntary attention (e.g., Aydın et al., 2011; Xu et al., 2013) or visual short-term memory (Hein et al., 2021; Kohler et al., 2008; Scocchia et al., 2013) can influence the correspondence solution. Based on such findings, Hein and colleagues (Hein & Cavanagh, 2012; Hein & Moore, 2014) suggested an object-based correspondence theory. Accordingly, correspondence could be solved based on all available information about an object and the most similar perceived objects at this object-based level are connected across space and time by an attention-based tracking mechanism (e.g. attentional pointer: Cavanagh, 1992; Cavanagh, Hunt, Afraz, & Rolfs, 2010) in order to establish correspondence.

Chapter 3 Aim of the present work

As described in Chapter 2, research on the correspondence problem (Dawson, 1991; Ullmann, 1979) has revealed different factors influencing how correspondence is solved. Based on these factors, different theories have been developed on which mechanism our visual system uses to process these factors in order to establish correspondence between objects and enable us to perceive a coherent representation of our environment. The aim of the present work is to further investigate the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). Organized by the studies conducted as part of this dissertation, this chapter discusses the open research questions on this topic and gives an overview of how these research questions were empirically investigated. For all studies, the Ternus display (Pikler, 1917; Ternus, 1926) was chosen, as it has been shown to be well-suited for investigating correspondence (for an overview see Hein, 2017; Petersik & Rice, 2006) and the object-based correspondence theory was developed based on studies using it.

3.1 Overview of the conducted studies

A few studies, mostly recent ones, have found evidence for the influence of higher-level perception, specifically object-based information on the establishment of correspondence, which is in line with the idea of an object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). In comparison, the influence of low-level image-based factors, like spatio-temporal information or image feature information, has been studied systematically over decades, resulting in ample amounts of evidence in line with image-based theories (motion-based theories: e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; feature-based theories: e.g., Kramer & Yantis, 1997). Thus, much less is known about the influence of object-based information on correspondence, and most of the studies that investigated this influence focused on the influence of top-down higher-level factors such as (voluntary) attention or visual working memory content (e.g., Aydın et al., 2011; Hein et al., 2021; Scocchia et al., 2013) or higher-level factors at the level of semantic and lexical information (e.g, Aydın et al., 2011; Chen & Zhou, 2011; Hsu et al., 2015; Yu, 2000). In contrast, only a few studies have investigated the influence of higher-level object information at a more intermediate level of visual processing, more precisely, more highly-processed input information from the retina such as perceived surface (He & Nakayama, 1994) and perceived lightness (Hein & Moore, 2014). But this exact level of processing seems to be especially important in the investigation of the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014), as it allows defining the limit point where image-based theories can no longer explain correspondence.

Therefore, the first study (Stepper, Moore, et al., 2020a, Chapter 4) conducted in the context of this dissertation started at this point and investigated the influence of another type of mid-level information, perceived size, on correspondence (see Figure 2.3 Black box 1). To investigate the influence of the perceived size of an object on correspondence, we manipulated the perceived size of a Ternus display using a Ponzo-like illusion. More precisely, the Ternus display was presented on a Ponzo-like background consisting of pictorial depth cues manipulating the perceived size of objects (Gregory, 2009; Rock, 1983). On such a background, objects perceived at the position far from the viewer's perspective are perceived as bigger compared to the objects perceived at the near position, despite the physical, retinal size of the objects being always the same. If correspondence is only solved at a low level of visual processing, the illusion should not have any effect, as the image-based information, the retinal size of the Ternus display, stays the same. In contrast, if correspondence is solved at a higher level of visual processing after the illusion is perceived, that is, after pictorial depth-cues are represented and size constancy is established, motion in the Ternus display should depend on the perceived size. This would be evidence that correspondence can be established after depth cues have been processed and therefore be further evidence that information at an intermediate level of visual processing is used for solving correspondence in line with the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014).

Even though studies showing an influence of object-based information, like perceived surface (He & Nakayama, 1994) and perceived lightness (Hein & Moore, 2014), have all taken great care to keep the image-based information in the different conditions as similar as possible, this information is not identical due to the experimental manipulation. Hein and Moore (2014), for example, showed that the perceived lightness, instead of only the physical luminance, can influence how correspondence is solved. To manipulate the perceived lightness, they introduced a filter to the display, which in turn could have introduced low-level changes in the luminance contrast. This leaves the possibility that low-level information has influenced correspondence and image-based theories (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997) could explain how correspondence was established. As long as it cannot be excluded that image-based information has also been manipulated, it is possible that the results could be influenced by it.

To answer the question of whether object-based information alone, independent from image-based information, is used for establishing correspondence, we conducted the second study (Stepper, Moore, et al., 2020b, Chapter 5), in which we manipulated mid-level object representations without changing image-based information at all (see Figure 2.3 Black box 2). To do so, we separated the manipulation of the object-based information from the presentation of the Ternus display, the task in which we measured correspondence. This was done by presenting one of two different object histories prior to the Ternus display. In the object history, the Ternus elements moved or changed their luminance either separately, appearing independent from each other (separate condition) or all together, appearing to be grouped (common condition). If image-based information alone is used for solving correspondence, object history should have no influence on the perceived motion in the Ternus display, as the information at the time of the Ternus display was identical in both conditions. If object-based information manipulated through the object history is integrated and stored into episodic object representations and can be used as basis for solving correspondence, the different histories (common vs. separate) should lead to different motion percepts in the Ternus display. This would be further evidence for the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014) and the use of episodic object representations to solve correspondence.

Within the framework of the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014) it has been suggested that attention might play a specific role for establishing correspondence. More precisely, attentional pointers (Cavanagh, 1992; Cavanagh et al., 2010) could be the means used to track the most similarly-perceived objects across space and time. Thus far, studies have shown that the overall availability of attention can influence the correspondence solution (Aydın et al., 2011) or that the deployment of attention is related to the motion percept (Xu et al., 2013). But there are no experiments specifically investigating the influence of attention as a potential mechanism for establishing correspondence between objects.

Therefore, the third study (Stepper, Rolke, et al., 2020, Chapter 6) tests this idea (see Figure 2.3 Black box 3), as attention has been proposed as a key mechanism for establishing correspondence in the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). To test this idea, we investigated whether voluntary spatial attention directed to a specific element of a biased Ternus display influences the correspondence solution. The Ternus display consisted of three differently-colored elements within a frame, arranged across frames in a way that it contained a bias toward element motion and a bias toward group motion at the same time (see Hein et al., 2021). If attention is used to establish correspondence, orienting attention to a specific element should make this element more likely to determine the correspondence solution, through being connected to the most similar element across frames. In the biased Ternus display, this should lead to more element motion if the element containing the bias toward element motion is attended compared to when the element containing the bias toward group motion is attended. Such a specific effect of attention would be evidence in line with the attentional pointer idea (Cavanagh, 1992; Cavanagh et al., 2010) and the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014), suggesting that attention could be a mechanism to establish correspondence.

3.2 Summary

The aim of this doctoral thesis is to investigate the role of object-based information available at an intermediate level of visual processing for establishing correspondence by finding further evidence for the influence of such information (Study 1: Stepper, Moore, et al., 2020a and Study 2: Stepper, Moore, et al., 2020b) and by investigating whether the information stored in the object representations alone, independent from any low-level information, can influence correspondence (Study 2: Stepper, Moore, et al., 2020b). In addition, this dissertation examines a specific prediction of the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014): that objects might be connected via attentional pointers (Study 3: Stepper, Rolke, et al., 2020). With the results of this work, further evidence can be gathered in favor of an object-based correspondence mechanism by defining the point where image-based theories, (Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997) reach their limits in explaining correspondence.

Chapter 4 Illusory size determines the perception of ambiguous apparent motion (Study 1)

Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Psychonomic Bulletin & Review, Illusory size determines the perception of ambiguous apparent motion, Stepper, M. Y., Moore, C. M., Rolke, B., & Hein, E., 2020, advance online publication, 10 August 2019 (https://doi.org/10.3758/s13423-020-01786-9)

The official citation that should be used in reference to this material is:

Stepper, M. Y., Moore, C. M., Rolke, B., & Hein, E. (2020). Illusory size determines the perception of ambiguous apparent motion. *Psychonomic Bulletin & Review, 27,* 1230-1238. https://doi.org/10.3758/s13423-020-01786-9

Psychonomic Bulletin & Review is available online at:

https://www.springer.com/journal/13423

4.1 Abstract

The visual system constructs perceptions based on ambiguous information. For motion perception, the correspondence problem arises, i.e., the question of which object went where. We asked at which level of processing correspondence is solved, lower levels based on information that is directly available in the retinal input or higher levels based on information that has been abstracted beyond the input directly available at the retina? We used a Ponzo-like illusion to manipulate the perceived size and separations of elements in an ambiguous apparent motion display. Specifically, we presented Ternus displays - for which the type of motion that is perceived depends on how correspondence is resolved - at apparently different distances from the viewer using pictorial depth cues. We found that the perception of motion depended on the apparent depth of the displays, indicating that correspondence processes utilize information that is produced at higher-level processes.

4.2 Introduction

“It’s an illusion” is how we could describe our perception of the three-dimensional world around us. This is because our perception is a constructed representation that is created by our visual system on the basis of ambiguous input information. The information that our visual system receives at the retina is a two-dimensional projection of the three-dimensional environment, which means that it is underdetermined and ambiguous. That our visual system actively resolves this ambiguity is evident in bi-stable perceptions such as our experience of the Necker cube (Necker, 1832; Fig. 4.1A), which can be perceived as oriented in two different ways, despite no change in retinal input, and in illusions like the Ponzo-like size illusion (sometimes also known as corridor illusion, Fig. 4.1B), in which identically sized stimuli are perceived as different sized objects because they are perceived as being at different distances from the viewer. Phenomena like these demonstrate that the interpretation of image-level information, which is what is directly available at the retina, depends on top-down processes that themselves utilize higher-level information (e.g., Kornmeier & Bach, 2012).

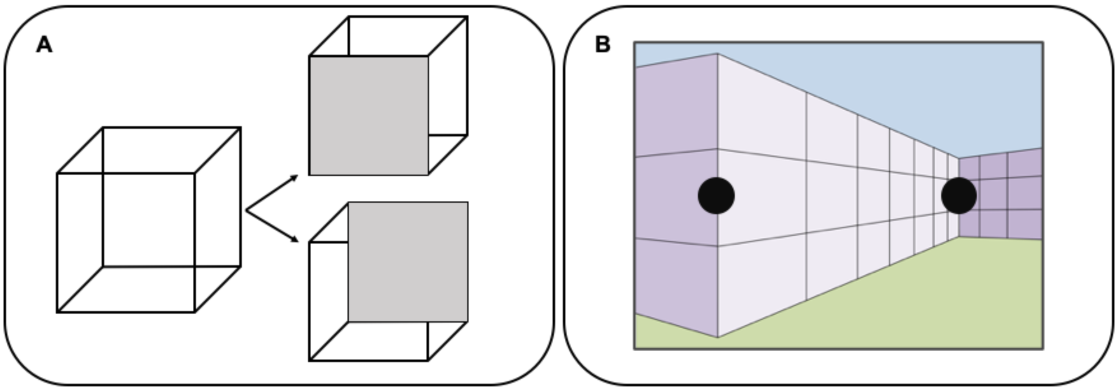

Figure 4.1: Necker cube and Ponzo-like illusion. **A**. Necker cube with its bi-stable percepts regarding the orientation on the left, disambiguated on the right side. **B.** Ponzo-like size illusion. The element on the right side is perceived as farther away and bigger in size compared to the element on the left side, although both are physically the same size.

The challenge of ambiguous input arises not only with static images like those shown in Fig. 4.1, but also with dynamic input. The identity of objects must be maintained across space and time, even as they become invisible because they are occluded by other objects due to their own or the viewer’s motion. As with the static examples, it is clear that our perception of objects over time depends on active top-down interpretation of ambiguous information. In the case of basic apparent motion, for example, successively presented static stimuli at different locations are perceived as a single object moving from one location to another if—and only if—the time and separation between them is consistent with how objects move in the world (Kolers, 1972; Korte, 1915; Wertheimer, 1912). The perception of objects over time becomes even more complex when, as is typical in natural environments, multiple stimuli are present in given static images. Fig. 4.2A illustrates the problem. Will motion be perceived based on spatial separation, retinal size, or neither? More generally, the question is how and on the basis of what information does our visual system determine which object went where? This problem, known as the *correspondence problem* (Ullmann, 1979), is a computational challenge because the image-based input is ambiguous (e.g., Dawson, 1991).

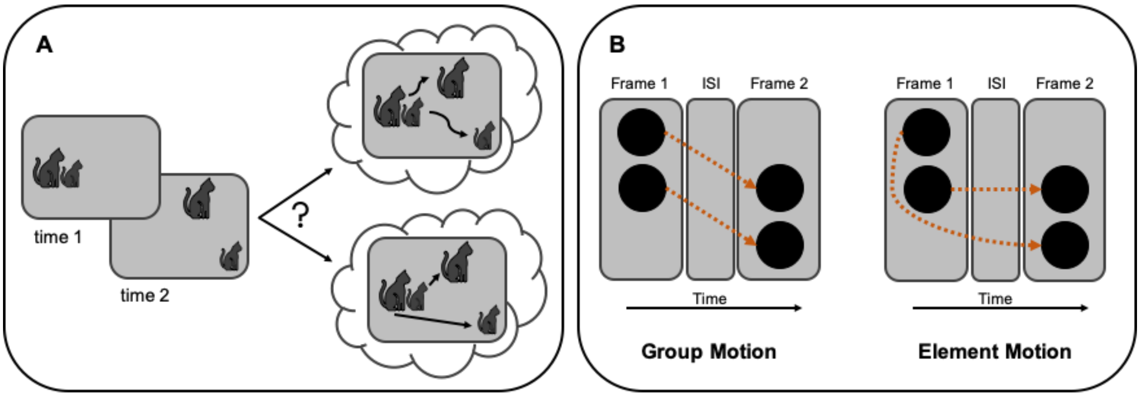

Figure 4.2: Correspondence problem and Ternus display. **A.** Illustration of the correspondence problem with the question which object went where. Are the cats connected and motion perceived based on their retinal size (upper solution) or based on their distance (lower solution)? **B.** Vertical version of the Ternus display and the two alternative motion percepts. The two successively presented Ternus frames are separated by a variable ISI. They can either be perceived as moving together (group motion) or as one element jumping across the other (element motion).

To address the question on what information correspondence is based, researchers have used ambiguous apparent motion displays, analogous to the example illustrated in Fig. 4.2A, in which depending on how correspondence is established, alternative and mutually exclusive motion percepts are experienced. An example of such a display is the Ternus display (Pikler, 1917; Ternus, 1926), which usually consists of two elements horizontally aligned next to each other, shifted by one element position from one frame to the next. For our purpose we created a vertical version of this display, in which two elements were vertically aligned, one above the other (Fig. 4.2B). Depending on the perceived correspondence between elements across frames, two alternative motion percepts are experienced. In one case, both elements appear to shift together (*group motion*), whereas in the other case, one element appears to "jump" across the other (*element motion*). Which type of motion is perceived—element or group—is therefore a way of assessing how correspondence was resolved. This is why the Ternus display is well suited for investigating the factors that determine correspondence (for an overview see Hein, 2017; Petersik & Rice, 2006). Studies using Ternus displays have shown that image-level information plays a role in determining correspondence, including the time between frames—the *interstimulus interval* (ISI)—(Navon, 1976; Petersik & Pantle, 1979) and the spatial separation of the elements (Casco, 1990; Navon, 1976; Petersik & Grassmuck, 1981). In particular, the longer the ISI and the smaller the separation of the elements the more group motion is perceived. In addition, studies have shown that feature information of the elements, such as luminance contrast, color, texture pattern (Hein & Moore, 2012; Petersik & Rice, 2008), and size (Breitmeyer & Ritter, 1986b, 1986a; Casco, 1990; Petersik & Grassmuck, 1981) influence correspondence. Breitmeyer and Ritter (1986a), for example, manipulated the size of the elements and found that larger Ternus elements lead to more group motion percepts compared to smaller ones. Thus, both spatio-temporal information and feature information determines the identity of the elements and how correspondence is resolved to give rise to one motion percept or the other.

While it is clear that feature information plays a role in determining correspondence, it is not clear whether it is feature information at the level of the retinal image or feature information at the level of the perceived object. Because correspondence reflects object identity—i.e., which object went where—it seems likely on functional grounds that it is the perceived feature information of the object that is critical, rather than the image feature. Some evidence consistent with this intuition comes from studies that have shown that the perceptual completion of objects that appear to extend behind occluding surfaces (*amodal completion*) is established before the motion percept is determined (He & Nakayama, 1994; He & Ooi, 1999; Yantis, 1995, Hein & Moore, 2014). He and Nakayama (1994), for example, showed that correspondence can be established on the basis of matching the perceived shapes of perceptually completed surfaces that were occluded by other surfaces, instead of the shapes of the physically visible parts of them. The fact that such information is used to determine correspondence suggests that it takes place at or after a level of visual processing at which amodal completion has taken place. This implies that correspondence can be determined by perceived object identity and does not necessarily have to rely only on image-level features. As this would require a significant rethinking of our understanding of the function of the correspondence process, the goal of the current study was to further investigate the influence of perceived object identity beyond the level of amodal completion processes.

In the size illusion illustrated in Fig. 4.1B, two stimuli of the same image size are perceived as different sized objects because they appear to be at different distances from the viewer within a depicted three-dimensional scene context. In particular, the stimulus that appears to be farther away from the viewer is perceived as larger than the stimulus that appears to be closer. This is consistent with the physics of three dimensions projecting onto two dimensions; an object will project a larger image onto a given projection plane when it is closer to that plane than when it is farther. Therefore, if an object that is (perceived as) farther away projects the same image size as one that is (perceived as) closer, it follows that the farther one is being projected by a larger object in the three-dimensional environment (e.g., Gregory, 2009; Rock, 1983). To investigate whether perceived size, beyond image size, determines correspondence, we presented identically sized Ternus displays on backgrounds (Fig. 4.3A & B; Illusion Ternus task; see supplementary video for an example) depicting depth such that they appeared to be either relatively near or relatively far from the viewer. If correspondence is based entirely on image size, then the perceived motion of the Ternus display should be unaffected by the apparent distance of the displays within these scenes. In contrast, if the perceived Ternus motion does vary with perceived distance, then we can infer that perceived size, which has to be abstracted from the size information that is directly available in the retinal image, contributes to the resolution of correspondence, and therefore that correspondence takes place at a higher level of visual processing than lower-level processes that extract the initial directly-available retinal information. In particular it would be one at which the representation of relative depths of objects and perceived size has been established. In separate tasks within the experiment, we additionally measured the magnitude of the size illusion (Illusion Magnitude task) and then used Ternus displays with those physical sizes and presented them on a background without implied depth differences (Fig. 4.3C; Image Ternus task). This provided a direct comparison of the correspondence solution between perceived size differences and size differences that were explicit in the image.

4.3 Method

Participants

Twenty-four observers participated in the experiment (16 female, 8 male, mean age = 24.04 years, *SD* = 3.13 years, range: 20-33). The sample size was calculated for an alpha of .05 with a power of .8 based on the effect size (partial eta square; Mordkoff, 2019) in a pilot study very similar to the Illusion Ternus task of this experiment. All observers were undergraduates from the University of Tübingen or from the surrounding community. They received 8 €/h or course credit in compensation for their time. All reported normal or corrected-to-normal visual acuity and were naïve as to the purpose of the experiment.

Apparatus

The experiment was controlled by a Windows computer (Window XP) driving a 17-inch CRT color monitor with a spatial resolution of 1024 x 768 pixel and a refresh rate of 100 Hz. MATLAB (Version R2012a, 7.14, Mathworks Inc., MA, USA) with Psychtoolbox 3 extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) was used to run the experiment. The experiment was conducted with the viewing distance fixed at 65 cm in a dimly lit individual testing room.

Stimuli

In order to create the size illusion, a full-screen background was constructed that depicts a range of depths using linear perspective cues (see Fig. 4.3A & B). The depth-generating texture, i.e., a wall, was constructed in first person perspective with two vanishing points and centered on the screen in a way that the observer perceived a nearest and a farthest point in the wall, which were equidistant from the center of the screen. Four different parts of the image were distinguished by their color: The upper part was blue (RGB: 185, 205, 229; 75 cd/m^2), the lower part green (RGB: 195, 214, 155; 74 cd/m^2), and the middle part with the wall texture was light purple (RGB: 230, 224, 236; 98 cd/m^2) and dark purple (RGB: 179, 162, 199; 46 cd/m^2), imitating the effect of an illumination source on the main part of the wall, the ends being in the shadow. Two different depth backgrounds were used (Near Left and Near Right, Fig. 4.3A & B) that were mirror versions of each other. The control background (Fig. 4.3C) was constructed to be as similar as possible to the depth background without using any perspective cues. The Ternus display (Pikler, 1917; Ternus, 1926) consisted of two frames with two elements vertically aligned with each other. By using the vertical version of the Ternus Display all elements were within the same perceived depth plane of the depth backgrounds and therefore perceived as being the same size. Each element had a diameter of 1.30° and the center-to-center separation between the elements was 1.63°. Depending on the presentation side the Ternus display was presented 8.4° to the left or to the right of the screen center, the middle Ternus element across both frames vertically centered on the screen. The color of the Ternus elements was black (RGB: 0, 0, 0; 0 cd/m^2) and the blank background between trials grey (RGB: 128, 128, 128; 20 cd/m^2).

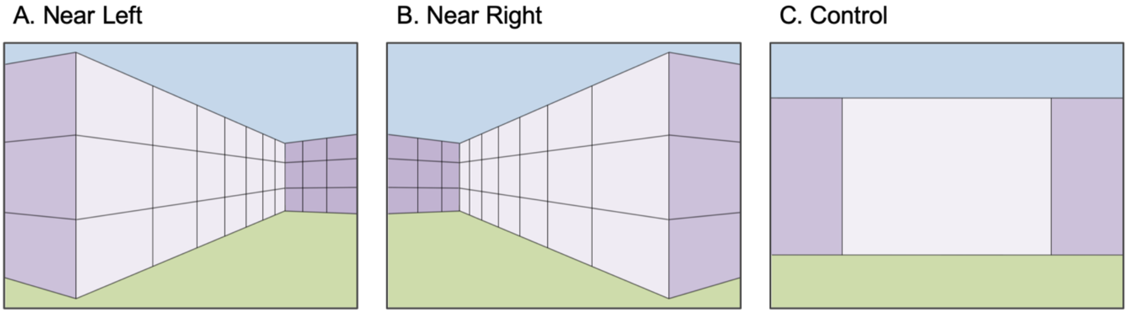

Figure 4.3: Ternus backgrounds. **A** and **B**. Depth backgrounds vertically mirrored used for the Illusion Ternus task and the Illusion Magnitude task. **C**. Control background used for Illusion Ternus task and Image Ternus task.

Procedure

Participants were first informed about the experimental procedure and completed an informed consent process according to the ethical principles of the World Medical Association (World Medical Association, 2013). The experiment lasted about 60 minutes and included three subtasks: Illusion Ternus task, Illusion Magnitude task and Image Ternus task. This order of the subtasks was the same across all participants. Each subtask started with written instructions on the screen.

For both the Illusion Ternus task and the Image Ternus task, participants were shown vertical versions of the Ternus display and asked to report whether they perceived element or group motion (see supplementary video for an example of the Illusion Ternus task). Following written instructions, demonstrations of clear element and clear group motion (using the most extreme ISIs of 0 and 240 ms, respectively) were presented. Participants performed a practice block of 18 trials and then completed six experimental blocks of 36 trials each. For the Illusion Ternus task in each trial (see Fig. 4.4), after a blank screen of 300 ms, one of the three different Ternus backgrounds (Near Left, Near Right, or Control; see Fig. 4.3) was presented for 800 ms, followed by the first Ternus frame superimposed on the background either in the left or right position for 200 ms. After a variable ISI of 0, 10, 20, 40, 80 or 240 ms, during which only the background was presented, the second Ternus frame was presented for 200 ms, followed by the background only for the same ISI. This cycle was repeated until participants responded.

For the Illusion Magnitude task, each trial started with a blank screen of 300 ms followed by one of the two depth backgrounds (Near Left or Near Right) for 800 ms. Then, Ternus elements were presented in each of the three positions of a Ternus display, in both the left and right position of the background. Participants adjusted the size of this three-element Ternus display on either the left or right side, which corresponded to the perceived near or far distance, depending on the depth background (Near Left or Near Right), until the displays on both sides were perceived as the same size. Which side was adjustable was chosen randomly. The non-adjustable three-element Ternus display (standard display) always had the size of 4.56° from edge to edge (element diameter: 1.30°; center to center separation between elements: 1.63°). Because both, the individual elements as well as the space between them are perceived as changing in size in the Illusion Ternus task, each adjustment affected both, the diameter of the elements and the space between the elements maintaining the proportions of the Ternus display. The start size of the adjustable three-element Ternus display was randomly either 1.47° larger or smaller than the size of the standard display (or 32.24 % of the standard three-element Ternus display). Adjustment steps were around 0.06° for the entire three-element Ternus display.

The Image Ternus task used the values estimated from the Illusion Magnitude task to set display sizes. Procedurally, it was the same as the Illusion Ternus task, with the exception that only the control background was presented and not the depth backgrounds, and the Ternus display for a given trial was one of three different sizes (standard, small, and large), randomly selected for each trial. The standard display size was identical to that used in the Illusion Ternus task. The small display size corresponded to the individual estimated size of stimuli presented in the near distance in the Illusion Magnitude task. The large display size corresponded to the individual estimated size of stimuli presented in the far distance in the Illusion Magnitude task.

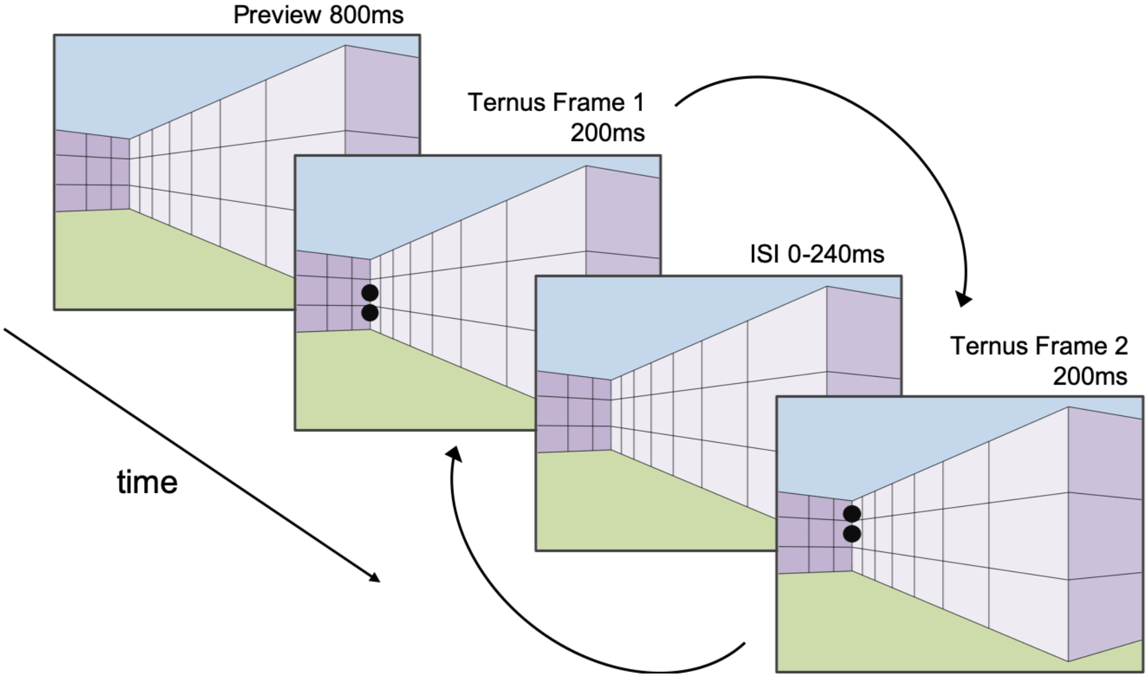

Figure 4.4: Ternus task trial. Illustration of the time course of a single Illusion Ternus task trial (shown here is the Near Right background with the Ternus presented on the left side, which corresponds to the perceived far distance).

Task

For both the Illusion Ternus task and the Image Ternus task participants reported whether the Ternus elements in the Ternus display appeared to be moving together (group motion) or as one element moving separately across the other element (element motion) by pressing the “j” or “f” key, respectively. In the Illusion Magnitude task participants adjusted the size of one of the two stationary columns of three-element display presented on the right and the left side until they perceived both as being the same size (method of adjustment; e.g., Coren & Girgus, 1972). The adjustments were made with the “j” (smaller) and “f” (bigger) key until the participants were satisfied with their result and confirmed with the space bar.

Design

For the Illusion Ternus task a 6 (ISI: 0, 10, 20, 40, 80, 240 ms) x 3 (background: Near Left, Near Right, Control) x 2 (Ternus position: left, right) within-subject design was used. All factors were counterbalanced and randomly mixed within all trials. Each participant completed 216 trials, resulting in 6 observations per condition. For the Illusion Magnitude task, a 2 (background: Near Left, Near Right) x 2 (display adjustment side: left, right) within-subject design was used. All factors were counterbalanced and randomly mixed within all trials. Each participant completed 4 observations per condition. For the Image Ternus task a 6 (ISI: 0, 10, 20, 40, 80, 240 ms) x 3 (Ternus size: small, standard, big) x 2 (Ternus position: left, right) within-subject design was used. Again, all factors were counterbalanced and randomly mixed within all trials. Each participant completed 6 observation per condition.

4.4 Results

Effect sizes are reported in terms of *adjusted partial eta-squared* ($adj\ \hat{\eta}_p^2$) which is an estimate of partial eta-squared that adjusts for the positive bias of the classic partial eta-squared that overestimates the population effect size (Mordkoff, 2019).

Illusion Magnitude task

To analyze the effect of the depth background on the perceived size of the Ternus elements we calculated the difference between the size of the standard Ternus elements and the size of the Ternus elements, which were adjusted by the participant. Negative values mean that the size of the adjusted elements was set to be larger than the standard size, and thus that the elements were perceived as smaller, while positive values mean that the size was set to be smaller than the standard size, and thus the elements were perceived as larger. As the two depth backgrounds were mirror versions of each other we combined the results from the adjusted elements based on their perceived distance (near or far). Figure 4.5A shows the mean perceived illusion size (in pixels; 1 pixel ≈ 0.02°) as a function of the perceived distance (near versus far). Participants perceived the Ternus elements at the perceived near distance as significantly smaller (*mean element diameter* = 51.22 pixel; *SD* = 9.62 pixel) than elements at the perceived far distance (*mean element diameter* = 70.12 pixel; *SD* = 5.79 pixel), $t(23) = -6.22, p < .001$, $adj\ \hat{\eta}_p^2 = .61$. In addition, one sample *t*-tests revealed that both size percepts differed significantly from zero, with the Ternus elements in the perceived near distance perceived as smaller than the standard element, $t(23) = -5.49$, $p_{\text{holm}} < .001$, $adj\ \hat{\eta}_p^2 = .55$, and the Ternus elements in the perceived far distance perceived as larger than the standard element, $t(23) = 6.87$, $p_{\text{holm}} < .001$, $adj\ \hat{\eta}_p^2 = .66$.

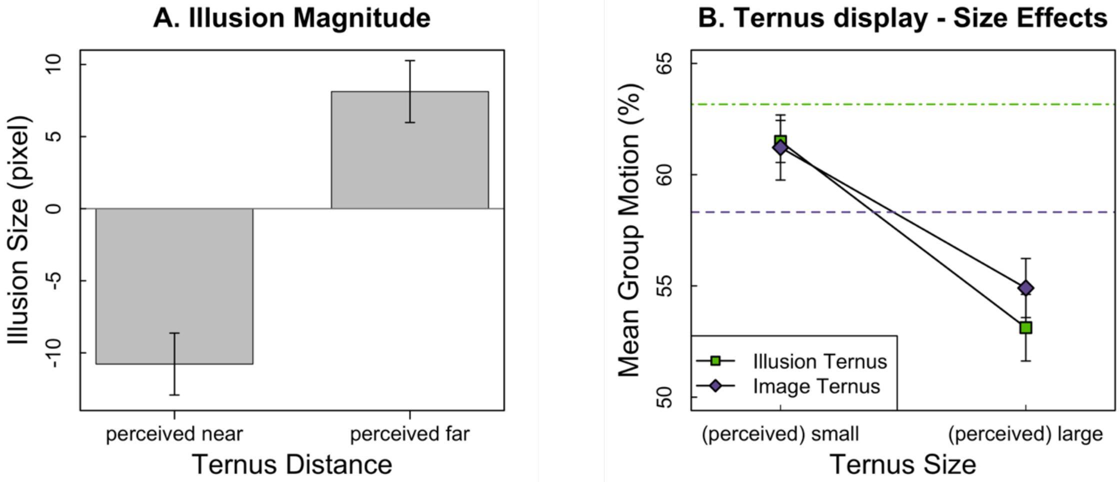

Figure 4.5: Results for all three experiments. **A**. Illusion Magnitude. Perceived illusion size as a function of illusion condition (elements are perceived as smaller in the perceived near distance and as larger in the perceived far distance compared to the standard elements). **B**. Effects of perceived and physical size in the Ternus display. Mean group motion responses as a function of the (perceived) small and large Ternus sizes for the Illusion and the Image Ternus task. The group motion responses for the baseline conditions (Image Ternus: standard Ternus size; Illusion Ternus: control background) are depicted with dotted and dot-dashed lines for the Illusion and the Image Ternus, respectively.

Illusion Ternus task

Next, we investigated whether the perceived size of the Ternus display can affect how correspondence is solved. Trials with responses other than the two response keys were excluded (0.96 %) as well as trials with RTs longer than 8,000 ms (1.17 %; mean RT: 1799 ms). On the remaining trials we calculated the percent of group motion responses. Again, as the two depth backgrounds were mirror versions of each other, we combined the results from the left and right Ternus position based on their perceived distance. We also collapsed the results from the left and right Ternus position for the control background. This way, we created the new factor Ternus distance, with the levels perceived near, perceived far and control distance. We conducted a 6 (ISI) x 3 (Ternus distance) repeated-measures ANOVA on the participants' means of group motion percepts. The analysis revealed an effect of ISI, $F(5,115) = 115.68, p < .001, adj\ \hat{\eta}_p^2 = .83$, as group motion percepts were increasing with increasing ISI (from 6.11 % at ISI 0 ms to 96.19 % at ISI 240 ms), which is the effect typically observed in the Ternus display (e.g., Breitmeyer & Ritter, 1986a; Petersik & Pantle, 1979). Most importantly, there was a strong effect of Ternus distance, $F(2,46) = 19.60$, $p < .001$, $adj\ \hat{\eta}_p^2 = .44$ (Fig. 4.5B). Holms corrected post-hoc comparisons revealed that significantly more group motion percepts were reported in the perceived near (M=61.68 %) compared to the perceived far condition (M=53.50 %), t(23)=4.26, p_{holm}<.001, $adj\ \hat{\eta}_p^2$=.42, as well as between the control (M=63.24 %) and the perceived far condition, t(23)=-5.02, p_{holm}<.001, $adj\ \hat{\eta}_p^2$=.50, but no significant difference was found between the perceived near and the control condition, t(23)=-1.62, p_{holm}=.120, $adj\ \hat{\eta}_p^2$=.06. In addition, we found a significant interaction between ISI and Ternus distance, F(10,230)=5.42, p<.001, $adj\ \hat{\eta}_p^2$=.16. To investigate this interaction more closely, we conducted post-hoc tests for each ISI between the two Ternus distance conditions that differed from each other (perceived far and near distance). They revealed significant and marginally significant differences for the ISI with the most ambiguous percept, i.e., the 10 and the 40 ms ISI, 3.94 <= t(23)<=4.89, p_{holm}<=.003, $adj\ \hat{\eta}_p^2$<=.49, and the ISI of 20 ms, $t(23)$=2.48, p_{holm}=.083, $adj\ \hat{\eta}_p^2$<=.18, but no significant differences for the less ambiguous ISI conditions, i.e., the 0, the 80, and the 240 ms ISI, 0.35<=t(23)<=0.96, p_{holm}=1, $adj\ \hat{\eta}_p^2$<=-.003.

Image Ternus task

Finally, we examined whether a physical size difference comparable to the individually perceived size difference obtained in the Illusion Magnitude task had a similar effect on the motion percept in the Ternus display. Trials with responses other than the two response keys (0.95 %) and trials with RTs longer than 8,000 ms were excluded (0.78 %, mean RT: 1,659 ms). Again, we combined the results from the left and right Ternus position. Therefore, we performed a 6 (ISI) x 3 (Ternus size) repeated-measures ANOVA on the mean percent of group motion percepts. There was again the typical ISI effect, F(5,115)=74.62, p<.001, $adj\ \hat{\eta}_p^2$=.75, as mean group motion responses increased with increasing ISI (from 6.67 % at ISI 0 ms to 94.08 % at ISI 240 ms). Most importantly, Ternus size influenced motion perception, F(2,46)=6.44, p=.007, $adj\ \hat{\eta}_p^2$=.18 (Fig. 4.5B). Holm's corrected t-tests for each Ternus size condition showed significantly less group motion percepts for the large (M=54.91 %) compared to the standard Ternus size (M=58.32 %), t(23)=-2.46, p_{holm}=.044, $adj\ \hat{\eta}_p^2$=.18, and compared to the small Ternus size (M=61.22 %), t(23)=-2.92, p_{holm}=.023, $adj\ \hat{\eta}_p^2$=.24. There was no significant difference between the small and the standard Ternus size, t(23)=1.77, p_{holm}=.090, $adj\ \hat{\eta}_p^2$=.08. In addition, there was a trend for an interaction between ISI and Ternus size, F(10,230)=2.12, p=.058, $adj\ \hat{\eta}_p^2$=.04. To investigate this trend more closely, we conducted post-hoc tests for each ISI between the two significantly different Ternus size conditions (small and large Ternus size). They revealed significant differences for the most ambiguous ISI condition of 20 and 40 ms, $3.21 <= t(23) <= 3.34$, $p_{\text{holm}} <= .020$, $adj\ \hat{\eta}_p^2 <= .30$, but no significant differences for the other ISI, $0.12 <= t(23) <= 2.28$, $p_{\text{holm}} >= .129$, $adj\ \hat{\eta}_p^2 <= .15$.

A notable aspect of the results is that the impact of size on perceived Ternus motion appears to be nearly the same whether the size differences are illusory (Illusion Ternus task) or physical (Image Ternus task). To assess this, we conducted an additional post-hoc 2 Task x 2 Ternus Size x 6 ISI repeated-measures ANOVA. Results showed no main effect of the factor task, $F(1,23) = 0.04$, $p < .849$, $adj\ \hat{\eta}_p^2 = -.04$. We also found no interaction between the factors task and size, $F(1,23) = 0.41$, $p < .529$, $adj\ \hat{\eta}_p^2 = -.03$, mean group motion responses being similar for the perceived small (61.68 %,) and the physically small condition (61.22 %), as well as for the perceived large (53.50 %) and the physically large condition (54.91 %). In addition, no other interactions with the factor task were significant, $0.87 <= F <= 2.13$, $p <= .091$, $adj\ \hat{\eta}_p^2 <= .05$.

4.5 Discussion

A critical function of vision is to establish and maintain representations of objects that have continuous identities over space and time, even as retinal input changes or disappears. An important aspect of achieving that function is determining which stimuli across time and space correspond to the same or different objects, a problem known as the correspondence problem (Ullmann, 1979). Previous work has shown that feature information at the level of the retinal image plays an important role in how the correspondence problem is solved by the visual system (e.g., Breitmeyer & Ritter, 1986a, 1986b). In this study, we investigated whether it is feature information at the level of the retinal image or at the level of the perceived object that is critical for the correspondence process. We found that the perception of Ternus motion varied with feature information at the level of the perceived object, more precisely with the perceived size of the stimuli evoked by different illusory depth backgrounds. In a separate task, we measured the magnitude of the size illusion and confirmed that the elements in the Ternus displays that appeared to be at the farther distance in the illusory depth scene were perceived as larger than those that appeared to be at the nearer distance. Finally, we used those measured magnitudes to create Ternus displays with corresponding *physical* size differences (all presented at the same apparent distance from the viewer), and found that the differences in perceived Ternus motion for the physically different stimuli matched the differences for the perceptually different, but physically identical, stimuli. Together, these results provide strong evidence that the correspondence process is resolved on the basis of higher-level properties of represented objects, rather than on lower-level properties of the image input.

The finding that larger Ternus displays, whether physically larger or illusorily larger, lead to less group motion percepts may appear contrary to studies that have found *more* group motion reports for physically larger Ternus elements (Breitmeyer & Ritter, 1986b, 1986a; Casco, 1990, Exp. 4). However, in those studies, the size of the individual elements was manipulated without changing the center-to-center separation between elements. A consequence of this is that the edge-to-edge separation between elements decreased with increasing element size and increased with decreasing element size. Larger edge-to-edge separation, however, is known to yield less group motion percepts (Pantle & Petersik, 1980). Furthermore, when element size and element separation were manipulated factorially, element separation was found to be the more important factor for determining correspondence (Casco, 1990; Petersik & Grassmuck, 1981). In the current study, size manipulations were designed to mimic metric size changes in image projections. Therefore, the ratio between the size of the elements and the distance was held constant across separation conditions, i.e., a change in element size included a corresponding change of the separation between elements. The pattern of our results, therefore, do not conflict with those of previous studies, but rather fit well with them.

Because perceived size depends on perceived depth, the fact that Ternus motion depended on perceived size further indicates that correspondence is resolved after depth information is encoded. This follows because the size illusion depends on the Ternus displays being perceived as being at different distances from the viewer as supported by the pictorial depth cues in the background displays (Gregory, 2009; Rock, 1983). The displays used in this study were inspired by the standard Ponzo illusion, which includes only two converging lines. Early considerations of that simpler illusion included the possibility that it was driven by lower-level image characteristics, rather than higher-level implications of depth relations (see Prinzmetal, Shimamura, & Mikolinski, 2001). But there is recent evidence that even that simpler version depends on higher-level information. Brown, Breitmeyer, Hale, & Plummer (2018) measured the contrast response function (CRF) for the traditional simple Ponzo illusion, i.e., how the magnitude of the illusion changes as a function of the contrast of inducing stimuli (i.e., the converging lines). They found non-linear changes in the CRF for the Ponzo illusion, indicating a dependence on higher-level perceptual coding (e.g., perceived size and distance). The authors therefore assumed that the Ponzo illusion involves higher-level information, dependent on representations in cortical regions, like V4, LOC, and inferotemporal cortex. The current results showing that Ternus motion depended on perceived depth and size, therefore suggesting that correspondence can happen at least at these levels of processing.

Finally, the current results are also consistent with previous studies showing that amodally completed stimuli play a role in how apparent motion is perceived (e.g., Hein & Moore, 2014). That work emphasized the conclusion that correspondence in motion perception depends not only on low-level motion energy (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993), but also on higher-level perceptual representation in which information about the structure and the content of the environment has been abstracted from the initial image-level input. The current results reinforce this conclusion by showing that the influence of object-based information occurs at or beyond the level of amodal completion, at which image-based information was further processed taking into account context information. Therefore, this study offers further support for an object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014; Stepper, Moore, et al., 2020b; Stepper, Rolke, et al., 2020), which states that all available information about an object, low-level and high-level, is taken into account for establishing correspondence, based on the (perceived) similarity between the individual element across frames.

In summary, using Ternus displays in the context of a depth-based size illusion, we found that the perceived size of objects, not simply image size, determines how correspondence is established. This indicates that the correspondence process takes place after pictorial cues are used by the visual system to establish representations of depth relations and structure.

4.6 Acknowledgements

We thank Ivo Adam for help with data collection. We also thank the action editor Jeanine Stefanucci and three anonymous reviewers for their helpful comments that improved this article substantially. This research was supported by DFG project HE 7543/1-1.

4.7 Open Practices Statement

Neither of the experiments reported in this article was formally preregistered. Data will be available on the Open Source Framework (OSF). Materials have not been made available on a permanent third-party archive; requests for materials can be sent via email to the lead author at [madeleine.stepper@uni-tuebingen.de].

Chapter 5 The role of object history in establishing object correspondence (Study 2)

Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Attention, Perception, & Psychophysics, The role of object history in establishing correspondence, Stepper, M. Y., Moore, C. M., Rolke, B., & Hein, E., 2020, advance online publication, 26 November 2019 (https://doi.org/10.3758/s13414-019-01923-0)

The official citation that should be used in reference to this material is:

Stepper, M. Y., Moore, C. M., Rolke, B., & Hein, E. (2020). The role of object history in establishing object correspondence. *Attention, Perception, & Psychophysics, 82,* 1038-1050. https://doi.org/10.3758/s13414-019-01923-0

Attention, Perception, & Psychophysics is available online at:

https://www.springer.com/journal/13414

5.1 Abstract

Our visual system establishes correspondence between objects and thus enables us to perceive an object, like a car on the road, as moving continuously. A central question regarding correspondence is whether our visual system uses relatively unprocessed image-based information or further processed object-based information to establish correspondence. While it has been shown that some object-based attributes, such as perceived lightness, can influence correspondence, manipulating object-based information typically involves at least minimal changes of image-based information as well, making it difficult to clearly distinguish between the two levels. To avoid this confound, we manipulated object-based information prior to the task in which we measured correspondence. We used 3-element Ternus displays to assess correspondence. These are ambiguous apparent-motion displays that depending on how correspondence is solved, are perceived as either one element jumping across the others or as all three elements moving together as a group. We manipulated object-based information, by presenting one of two object histories prior to the Ternus display. In one, they moved or changed luminance independently, and thus appeared independent from each other. In the other, the elements

moved or changed their luminance all together and thus appeared grouped with each other. We found that the object history did influence how the Ternus displays were perceived, thereby confirming that object-based information alone can be used as a basis for establishing correspondence in line with object-based theories of correspondence.

5.2 Introduction

We perceive a moving object, like a car, as a distinct entity with a continuous history based on visual information that may itself be discontinuous. To do this our visual system must establish correspondence between objects across time, i.e., represent whether an object at a current time is the same as one seen previously. If correspondence is established, then a single object is perceived. In contrast, if correspondence is not established, then separate individual objects are perceived (e.g., Dawson, 1991; Wertheimer, 1912). How the visual system establishes object correspondence, therefore, has an important impact on how the world is perceived. It could even determine whether one or two objects are represented. Beyond the direct function of determining the number and continuity of objects perceived, object correspondence is thought to support other cognitive functions such as change detection (Flombaum & Scholl, 2006), perceptual stability across eye movements (Tas, Moore, & Hollingworth, 2012), and object recognition across eye movements (Poth, 2015; Poth & Schneider, 2016). Object correspondence is especially important when the scene is complex. For example, when multiple moving objects are present, like many cars at a crossroad, they may temporarily occlude each other. In such situations, which objects belong together is ambiguous because any current object could, in principle, correspond to any object from a moment ago (e.g., ambiguous apparent motion; Pikler, 1917; Ternus, 1926). This kind of ambiguity is known as the *correspondence problem* (Dawson, 1991; Ullmann, 1979).

A central question regarding the correspondence problem concerns at which level of processing correspondence is established. One possibility is that correspondence is image-based, which means that correspondence is established on the basis of relatively unprocessed visual information that makes up the retinal image, e.g., luminance, spectral content, textures (Adelson & Bergen, 1985; Breitmeyer & Ritter, 1986a, 1986b; van Santen & Sperling, 1985; Werkhoven et al., 1993). A set of features in one part of the retina at one time is perceived as corresponding to a similar set of features at another part of the retina at a later time, and is thereby perceived as a single object such as a car. Under this view, information about the car as an object (e.g., that it was recently occluded by another car) cannot be used for establishing correspondence because it is unavailable at an image-level of processing. An alternative possibility is that correspondence is object-based (Hein & Cavanagh, 2012; Hein & Moore, 2014), which means that correspondence is established on the basis of associating episodic representations of objects. Such object representations include information about the history of the object, including for example changes in appearance or position, from the time the object representation was initially established. For the moving car example this would mean that where it was previously, and how it interacted with other perceived objects in the scene, would be used in the correspondence process. If the car becomes occluded, for example, the occluded information would be maintained as part of the representation and could influence later correspondence processes even though it is no longer present in the image. The focus of this study is to differentiate between these two levels of visual processing.

There are proponents of both image-based and object-based theories of object correspondence. Motion energy models are examples of image-based theories (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993). In these models, simple low-level motion detectors (i.e., Reichardt detectors; Reichardt, 1961) compute the direction in which the most motion energy occurs, i.e., the greatest change of features at subsequent locations in the image is detected. Based on this outcome, motion and direction are perceived, and correspondence is determined. Evidence of such image-based mechanisms comes in part from studies using apparent motion displays (Wertheimer, 1912). In these displays, movement can be perceived between two stimuli that are presented successively and at two different positions. The perception of motion in these displays is referred to as apparent motion because the display consists of two stationary stimuli, but through correspondence processes they are together perceived as a single object moving from one location to another. According to image-based theories, the perception of apparent motion emerges from the establishment of correspondence between contrast energy at one location and time with contrast energy at another location and time. Studies that show that apparent motion between two elements is dependent on the spatio-temporal gap, i.e., the time and distance between them, are in line with these theories (e.g., Kolers, 1972; Korte, 1915).

In contrast to image-based theories of correspondence, object-based theories assert that correspondence can be established between higher-order representations. An example is the use of attentional pointers to associate image elements across time that are identified as corresponding to a single object within the scene (e.g., Hein & Cavanagh, 2012; Hein & Moore, 2014). The object representation is then updated accordingly (e.g., via an object-mediated updating process; Enns et al., 2010; Lleras & Moore, 2003; Moore & Lleras, 2005; Moore, Mordkoff, & Enns, 2007). Evidence of object-based correspondence comes in part from ambiguous apparent-motion displays such as the Ternus display (Pikler, 1917; Ternus, 1926). In the classic Ternus display (see Figure 5.1), three identical stimuli (e.g., black discs) are presented next to each other with uniform separation. From the first frame to the second, the three stimuli are presented in locations that are shifted by one element position. Depending on how correspondence is solved between these two stimulus frames, motion in the Ternus display is perceived either as one element jumping across the other two elements (element motion) or as all elements moving together (group motion). These different percepts reflect different, mutually exclusive, object correspondences. And consistent with object-based theories of correspondence, higher-order information - as for example the motion context or perceived lightness, but also semantic and lexical information - can influence which type of motion, element or group, is perceived (Aydın et al., 2011; Chen & Zhou, 2011; He & Nakayama, 1994; He & Ooi, 1999; Hein & Moore, 2014; Hsu et al., 2015; Ramachandran & Anstis, 1983b; Yu, 2000). Hein and Moore (2014), for example, showed that the match between the perceived lightness of the stimuli in the Ternus display, rather than their physical luminance, determined whether element or group motion was perceived, indicating that correspondence was based on the higher-order attribute of lightness not the image-level attribute of luminance.

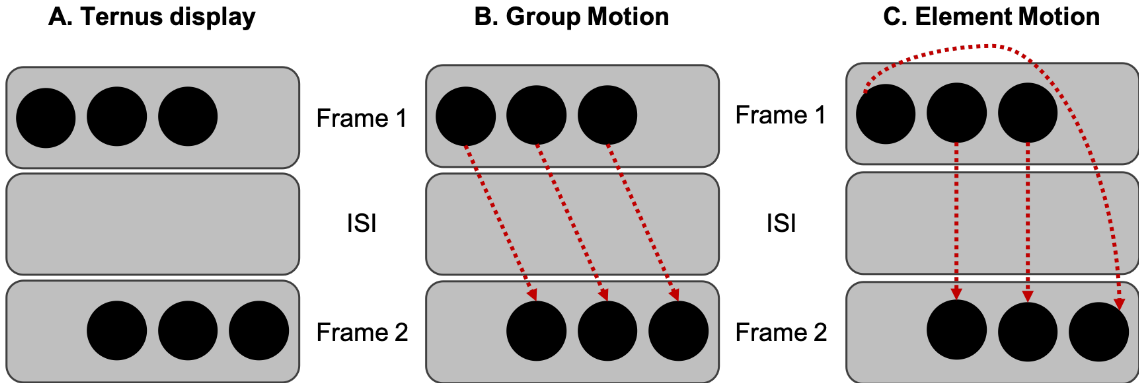

Figure 5.1: Ternus display and the two alternative motion percepts. A. Ternus display: The two successively presented Ternus frames are separated by a variable ISI. B. Group Motion. Correspondence between elements compatible with the group motion percept. C. Element Motion. Correspondence solution compatible with the element motion percept.

While the studies cited above are consistent with an object-level correspondence process, a challenge to testing between object-based and image-based theories of correspondence is that object representations are initially defined on the basis of information in the image. Therefore, manipulating information at the object-level usually involves image-level changes to displays. While conditions can be established such that local parts of the image are identical while object-level aspects differ, these strategies require different image-level information as context, such as different sets of transparent filters (Hein & Moore, 2014) or occluding surfaces (Hein & Moore, 2014; Moore et al., 2007). An alternative strategy that we take in the current study is to define object structure on the basis of spatio-temporal history, and measure how correspondence is established when displays are physically identical but have different object histories. Using this strategy, the possibility that any differences that are observed were driven by differences in image-level context can be ruled out because there will be no differences in image-level context at the time that correspondence is measured. To the best of our knowledge this is the first time this approach is used to investigate the influence of object-based information on object correspondence.

The aim of this study was to distinguish between theories that maintain that the correspondence process is entirely image-based and theories that maintain that it can be object-based. We measured perceived Ternus motion (i.e., element versus group motion) as a measure of correspondence, and manipulated the object history of the stimuli used in the Ternus displays. This provided a measure of correspondence under different object conditions, but identical image conditions. Following previous studies (Moore & Lleras, 2005; Mordkoff & Danek, 2011), we manipulated object history by presenting a short movie prior to the Ternus display that showed the Ternus stimuli as either spatio-temporally grouped together (common history, e.g., movie 1), or spatio-temporally independent from each other (separate history, e.g., movie 2). Any difference in how the Ternus displays are perceived following these two different types of history, would indicate that object-based information played a role in correspondence, consistent with object-based theories and inconsistent with pure image-based theories. If, however, perceived Ternus motion is the same across the two different object-history conditions, then it would indicate that object history was insufficient to drive correspondence, consistent with image-based theories.

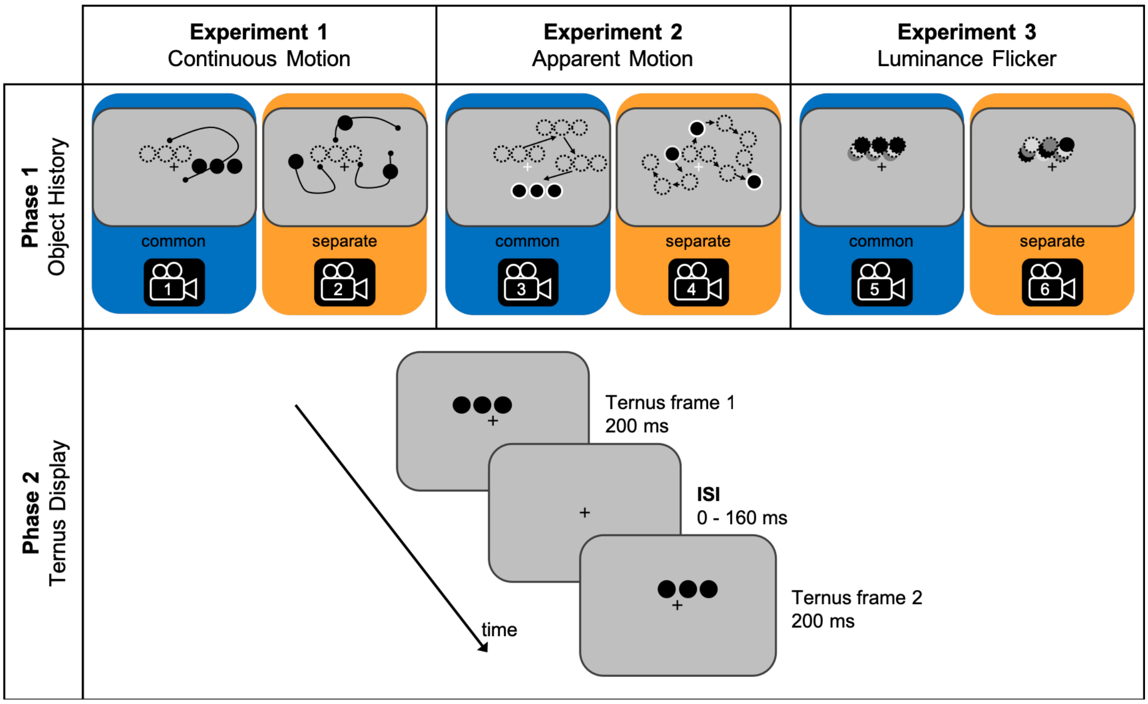

Figure 5.2: Overview experiments. Each trial consisted of two phases. In phase 1, the object history was manipulated in order to change the representations of the Ternus elements as being spatio-temporally grouped together (common history) or independent from each other (separate history). Examples of the type of movies used in this phase can be found online. In phase 2, after the object history, the Ternus display was presented and participants were asked to indicate whether they perceived element or group motion.

5.3 Experiment 1: Continuous motion object history

Each trial in Experiment 1 began with a short movie in which the three stimuli from a Ternus display were shown in smooth random motion. In the common history condition, the elements moved together along a single random trajectory, whereas in the separate history condition, the three elements moved independently along different random motion trajectories (see Figure 5.2 Experiment 1 and movies 1 and 2). The movie was then followed by a standard Ternus display, which was identical across the two history conditions. On the basis of the principle of grouping-by-common-fate (Wagemans et al., 2012; Wertheimer, 1923) the stimuli in the common history condition should be represented as a single group, whereas in the separate history condition, they should be represented as separate entities. If more group motion is perceived following common history movies than separate history movies, it will indicate that object-level information influenced correspondence.

5.3.1 Method

Participants

Thirty-nine individuals from the University of Tübingen and the surrounding community participated in the experiment, and received 8 € / h or course credit in compensation for their time. One participant was excluded because he/she had participated in a similar experiment and was therefore not naïve as to the purpose of this one. Two others were excluded because of technical errors during data collection. A total of 36 participants (27 female, 9 male; mean age = 25.4 years, range: 19-56), therefore, contributed to the reported data. All were naïve as to the purpose of the experiment and reported normal or corrected-to-normal vision.

Apparatus

The experiment was controlled by a PC running Windows XP, driving a 17-inch color cathode ray tube monitor (1024 x 768 pixel) with a refresh rate of 100 Hz. MATLAB software (Version R2012a, 7.14, Mathworks Inc., MA, USA) with Psychtoolbox 3 extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) was used to run the experiment. The experiment was conducted in a dimly lit individual testing room with a fixed viewing distance of 65 cm to the monitor.

Stimuli

A fixation cross was presented at the center of the screen, which had a line width of 0.08° and measured 0.2° x 0.2°. The Ternus display (Pikler, 1917; Ternus, 1926) consisted of two frames with three elements aligned next to each other, each with a diameter of 1.6°. The center-to-center distance between the elements was 2°. The elements were presented 2° above the fixation cross, the middle of the center elements being vertically aligned with it. The color of the Ternus elements and the fixation cross was black (RGB: 0, 0, 0; 0 cd/m^2). They were presented on a gray background (RGB: 127.5, 127.5, 127.5; 24 cd/m^2). For the motion history identical elements were used as in the Ternus display.

Continuous motion object history

To create the motion history of the elements, a sequence of five element positions was chosen. The starting position of the elements in the motion history was always the same as for the first Ternus frame. From there the elements moved away for five positions and then moved back taking the same positions in reversed order until the motion history ended again on the start position of the Ternus display. For each trial a new random motion sequence was calculated using a new set of five positions. Each element position of the motion sequence was calculated as follows: the x-value of an element position was calculated by multiplying the sine (in degrees) of a randomly chosen degree value (range: 0-359°) by a randomly chosen radius value (range: 30-60 pixels; element diameter: 58 pixels). For the y-value of this element position the cosine of the same degree was multiplied with the same radius used for the x-value. In the common history condition the same motion sequence was used for all three elements, which started and finished at the same time. For the separate history condition separate motion sequences were calculated for each element. The elements started their movement one after another in random order and finished in reversed order. To realize the sequential start of the elements in the separate history condition, while at the same time having the same movie duration in both conditions, two element presentations without a position change were added to the common history condition. The entire object history sequence lasted for 2,400 ms.

Task

During the presentation of the object history movie, participants only had to watch the history. For the following Ternus display participants had to report whether they perceived all elements in the Ternus display as moving together (group motion) or as one element jumping across the others (element motion). They gave their response by pressing the “j” or “f” key, respectively.

Procedure

The experiment lasted about 60 minutes and began with informing participants about the experimental procedure. After they gave informed consent according to the ethical principles of the World Medical Association (2013, Declaration of Helsinki) the experiment started with written instructions on the screen followed by clear demonstrations of group and element motion (using the most extreme interstimulus intervals (ISIs) tested of 0 and 160 ms) and a practice trial block of 12 trials. After asking questions if something about their task remained unclear participants completed 12 experimental blocks of 24 trials each. Each trial began with the presentation of the first frame of the three Ternus elements and the fixation cross for 200 ms. The fixation cross remained on the screen during the entire trial and participants were asked to fixate it. Then the motion history movie was presented (2,400 ms). After the motion history the fixation cross was presented alone for another 500 ms (short pause condition) or 1,200 ms (long pause condition), to test whether the time between the movie and the Ternus display influences the processing of the history. The pause was followed by the first Ternus frame, which was presented for 200 ms. After a variable ISI (0, 10, 20, 40, 80 or 160 ms), during which again only the fixation cross was presented, the second Ternus frame was shown for another 200 ms. The fixation cross stayed on the screen until the participants responded. The next trial started after 800 ms.

Design

A 6 (ISI: 0, 10, 20, 40, 80, 160 ms) x 2 (history: common, separate) x 2 (pause: short, long) within-subject design was used in which all factors were counterbalanced and randomly mixed within all trials. Each participant completed 288 trials, resulting in 12 observations per condition.

5.3.2 Results and discussion

Figure 5.3A shows mean group motion responses as a function of object history and ISI for Experiment 1. Trials with responses other than the two response keys were eliminated (0.85 %). In addition, all trials with RTs longer than 8,000 ms were excluded (0.06 %; mean RT: 1,046 ms). We performed a 6 x 2 x 2 repeated-measures ANOVA on the subject means with the factors ISI, object history and pause on the mean group motion responses. Effect sizes are reported throughout in terms of *adjusted partial eta-squared* ($adj\ \hat{\eta}_p^2$) which is an estimate of partial eta-squared that adjusts for known bias (Mordkoff, 2019). As expected mean group motion responses increased with increasing ISI, $F(5{,}175) = 115.57$, $p < .001$, $adj\ \hat{\eta}_p^2 = .76$, replicating the typical ISI effect found for the Ternus display in many studies (e.g., Breitmeyer & Ritter, 1986a; Petersik & Pantle, 1979). The pause showed no significant effect, $F(1{,}35) = 0.86$, $p = .361$, $adj\ \hat{\eta}_p^2 < .01$, but there was a significant interaction between pause and ISI, $F(5{,}175) = 3.86$, $p = .002$, $adj\ \hat{\eta}_p^2 = .07$. To examine this interaction further, we conducted Holm's-corrected paired *t*-tests for the pause condition for each ISI condition separately. They showed that at an ISI of 20 ms more group motion was perceived in the long than in the short pause condition (85 % vs. 80 %, $t(35) = 2.98$, $p_{\text{holm}} = .031$, $adj\ \hat{\eta}_p^2 = .18$). All other comparisons were not significant, $1.93 <= t <= 2.38$, $p_{\text{holm}} >= .116$, $adj\ \hat{\eta}_p^2 <= -.11$. Most importantly, however, we found that object history had no significant effect on the motion percept, $F(1{,}35) = 0.70$, $p = .409$, $adj\ \hat{\eta}_p^2 = -.01$, and there was also no significant interaction of object history with any of the other factors, $F_s <= 0.97$, $p_s >= .436$, $adj\ \hat{\eta}_p^2 <= -.01$.

The main result was that there was no influence of object history on perceived Ternus motion, which is consistent with pure image-based theories of correspondence. We manipulated object history in Experiment 1 using smooth continuous motion to establish (un)grouping by (un)common fate. Previous experiments that manipulated object history, however, used apparent motion (Moore & Lleras, 2005; Mordkoff & Danek, 2011). Because Ternus motion (group or element) is itself apparent motion, it is possible that the Ternus display was not perceived as related with the object-history part of the display, and information from it was therefore not used for resolving correspondence in the Ternus display. This possibility was examined in Experiment 2.

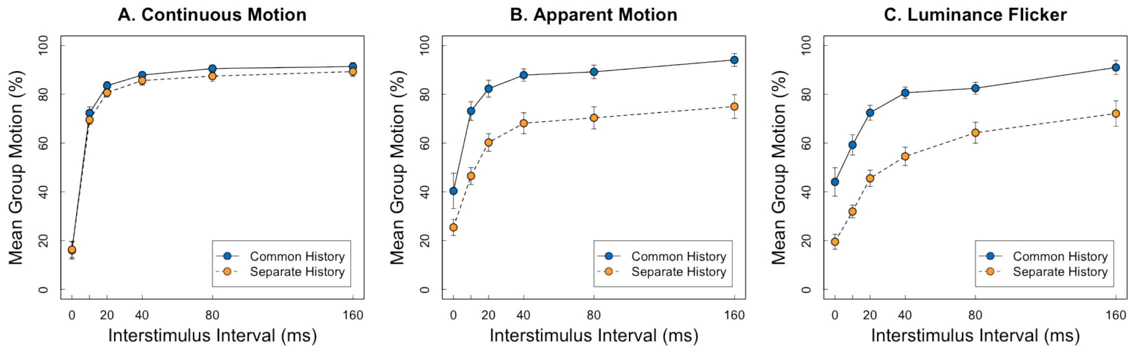

Figure 5.3: Results for all three experiments. Mean group motion responses as a function of ISI and object history for all three experiments (N = 36 in each). A. Experiment 1: Continuous motion object history. B. Experiment 2: Apparent motion object history. C. Experiment 3: Luminance flicker object history. Standard errors represent within-subject SEs after Cousineau-Morey (Cousineau, 2005; Morey, 2008).

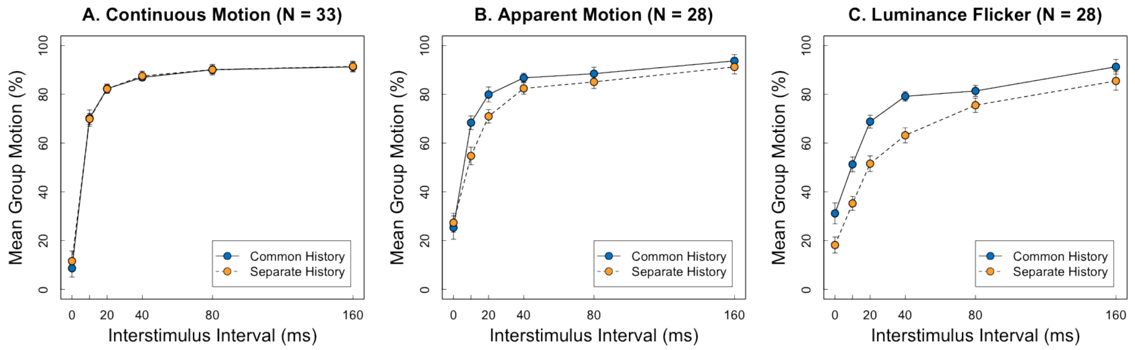

Figure 5.4: Results for all three experiments (subset of participants). Mean group motion responses as a function of ISI and object history for all three experiments for a subset of participants. Participants that showed no effect of ISI across both object history conditions were excluded from the data set for these graphs. A. Experiment 1: Continuous motion

object history (three participants excluded). B. Experiment 2: Apparent motion object history (eight participants excluded). C. Experiment 3: Luminance flicker object history (eight participants excluded). Standard errors represent within-subject SEs after Cousineau-Morey (Cousineau, 2005; Morey, 2008).

5.4 Experiment 2: Apparent motion object history

Experiment 2 was analogous to Experiment 1, except that object history was manipulated using apparent motion instead of smooth motion by including a blank screen between the stimulus frames (see Figure 5.2 Experiment 2 and movies 3 and 4). This change made the quality of motion in the object-history movies more similar to the quality of motion in the Ternus displays. If more group motion is perceived following common history movies than following separate history movies in Experiment 2, then it would suggest that the mismatch of motion quality across motion history and Ternus displays in Experiment 1 lead to object information not carrying over to the Ternus displays for use in the correspondence process. If, however, there is no difference in perceived group versus element motion as a function of object history, it would reinforce the possibility that correspondence is a purely image-based process.

Following the main experiment, we asked participants three short questions in order to assess the extent of the participants' understanding of the objective of the experiment. We used the questionnaire, because of the concern that participants might strategically base their responses "group" vs. "element" on the history part of the display, rather than only on the Ternus display as instructed. This possibility would seem to be more likely for participants who were aware of the objective than those who were not. We therefore sought to identify those subjects in order to be able to have an independent measure of this potential influence.

5.4.1 Method

Participants

Thirty-seven individuals participated in Experiment 2, and received 8€/h or course credit in compensation for their time. One participant was excluded because he/she had participated in a similar experiment and was therefore not naïve as to the purpose of this one. A total of 36 participants (26 female, 10 male; mean age = 24.5 years, range: 18-47), therefore, contributed to the reported data. All were naïve as to the purpose of the experiment and reported normal or corrected-to-normal vision.

Apparatus

The same apparatus was used as in Experiment 1.

Stimuli

With the following exceptions the stimuli used in this experiment were the same as in Experiment 1: The fixation cross had a line width of 0.05° and was 0.15° x 0.15° in size. The black Ternus elements were surrounded by a small white (RGB: 255, 255, 255; 150 cd/m^2) outline (linewidth: 1 pixel), in order to make sure that the individual elements were easily perceived even when they overlapped with each other. The fixation cross was first presented in white during the object history and then changed its color to green (RGB: 100, 255, 60; 103 cd/m^2) with the start of the Ternus display. The object history elements looked the same as the Ternus elements.

Apparent motion object history

The element positions for the motion sequence of the object-history elements were calculated in the same way as for Experiment 1, but the radius was increased compared to Experiment 1 and randomly chosen between 60 and 100 pixels. Additionally, a blank screen divided each presentation of the object history elements, which was presented for the same duration as the ISI used for the Ternus display in each trial. To keep the duration of the motion history similar to Experiment 1, a motion sequence of only six element positions was used (total object history time: 1,400-2,520 ms).

Task

The task was the same as in Experiment 1.

Questionnaire

The following three questions were used to get a more and more precise idea of what the participants understood of the purpose of the experiment: Question 1: “What do you think this experiment was about?”; Question 2: “Do you have any idea, why you have seen different movie sequences before the actual task? If yes, what is your idea?”; Question 3: “In your opinion, what was the author’s hypothesis?”. To analyze the questionnaire, one of the authors and an undergraduate research assistant rated the answers independently in the following way: for each answer to one of the three questions the raters gave a score of either 0 (no idea), 1 (a vague idea, but incorrect) or 2 (correct). As a measure for the agreement between both raters we used Cohen's (1968) weighted kappa (weight: squared). We used the average across both raters of the sum of the scores (ranging between 0 and 6) for each question as the final experimental understanding score of the participant.

Procedure

The procedure was the same as for Experiment 1 apart from the following exceptions: The practice trial block consisted of 24 trials and participants completed 10 experimental blocks of 24 trials. The trial sequence in Experiment 2 was similar to Experiment 1, but between the object history and the start of the Ternus display no different pause conditions were used, the time interval being the same as the ISI. To make the start of the Ternus display sequence easily recognizable, the color of the fixation cross changed from white to green simultaneously with the start of the Ternus display. In addition, we introduced two different motion directions of the Ternus display to make the Ternus direction unpredictable. The second frame of the Ternus display could therefore be presented to the right or the left of the first frame. At the end of the main experiment the questions of the questionnaire were presented on the screen. Participants responded to each question one after the other by typing their answer on the keyboard.

Design

A 6 (ISI: 0, 10, 20, 40, 80, 160 ms) x 2 (history: common, separate) x 2 (Ternus direction: left, right) within-subject design was used in which all factors were counterbalanced and randomly mixed within all trials. Each participant completed 240 trials, resulting in 10 observations per condition.

5.4.2 Results and discussion

Ternus percept

Figure 5.3B shows mean group motion responses as a function of object history and ISI for Experiment 2. As in Experiment 1, trials with responses other than the two response keys were eliminated (0.75 %). In addition, all trials with RTs larger than 8,000 ms were also excluded (0.02 %, mean RT 1,134 ms). We performed a 6 x 2 x 2 repeated-measures ANOVA on the subject means with the factors ISI, object history and Ternus direction on the mean group motion responses. We found no main effect of Ternus direction, $F(1,35) = 0.06, p = .816, adj\ \hat{\eta}_p^2 = -.03$. Ternus direction also did not interact with any of the other factors and the three-way interaction was also not significant, $F_s <= 0.77$, $p_s >= .413$,

$adj\ \hat{\eta}_p^2$<=-.01. As expected, there was a significant effect of ISI, $F(5,175) = 47.94, p < .001$, $adj\ \hat{\eta}_p^2 = .57$, as mean group motion responses increased with increasing ISI. Most importantly and in contrast to Experiment 1, we found a significant effect of object history, $F(1,35) = 12.77, p = .001$, $adj\ \hat{\eta}_p^2 = .25$, with more group motion responses in the common object history condition compared to the separate object history condition. We also found an interaction of object history and ISI, $F(5,175) = 3.55, p = .013$, $adj\ \hat{\eta}_p^2 = .07$. To examine this interaction further, we conducted Holm's-corrected *t*-tests for object history and each ISI condition separately. These post-hoc analyses revealed that the object history effect was reliable for all ISI conditions, including the 0 ms condition, *t*(35)>= 2.33, p_{holm}<=.026, $adj\ \hat{\eta}_p^2$<=.12.

A subset of eight participants showed no effect of the ISI on reports of Ternus motion (collapsed across both object history conditions). These participants may have used a strategy in which they based their responses on the history part of the display, which is independent of ISI, rather than on the Ternus display. Such a strategy would be especially problematic as it concerns our question because it would yield data consistent with the object history influencing the perception of Ternus motion, when in fact the responses are unrelated to the perception of Ternus motion. We therefore repeated our analyses excluding this subset of participants, and found a very similar pattern to the original analyses. Figure 5.4B shows the mean data for Experiment 2, excluding the eight subjects who showed no effect of ISI. Although the difference between the two object conditions was reduced, there was still a main effect of object history, $F(1,27) = 6.06$, $p = .021$, $adj\ \hat{\eta}_p^2 = .15$. The main effect of ISI was significant, not surprisingly given that inclusion in this analysis was contingent on this, $F(5,135) = 78.02$, $p < .001$, $adj\ \hat{\eta}_p^2 = .73$. In addition, the interaction between object history and ISI was significant, $F(5,135) = 5.26, p = .001$, $adj\ \hat{\eta}_p^2 = .13$. No other reliable effects were observed, F_{s}<=1.10, p_{s}>=.358, $adj\ \hat{\eta}_p^2$<=.004. Separate Holm's corrected *t*-tests for each ISI condition showed that the object history effect was only significant for the 10 ms condition, *t*(27) =3.67, p_{holm}=.006, $adj\ \hat{\eta}_p^2 = .31$, the condition, in which the Ternus motion is most ambiguous and therefore presumably most susceptible to be influenced from other sources of information such as object history instead of the ISI. No other comparisons were significant, $t_{\text{s}}(27)$<=2.38, $p_{\text{s;holm}}$>=.124, $adj\ \hat{\eta}_p^2$<=.14. For comparison, Figure 5.4A shows the analogous data for Experiment 1, with three participants who showed no effect of ISI excluded.

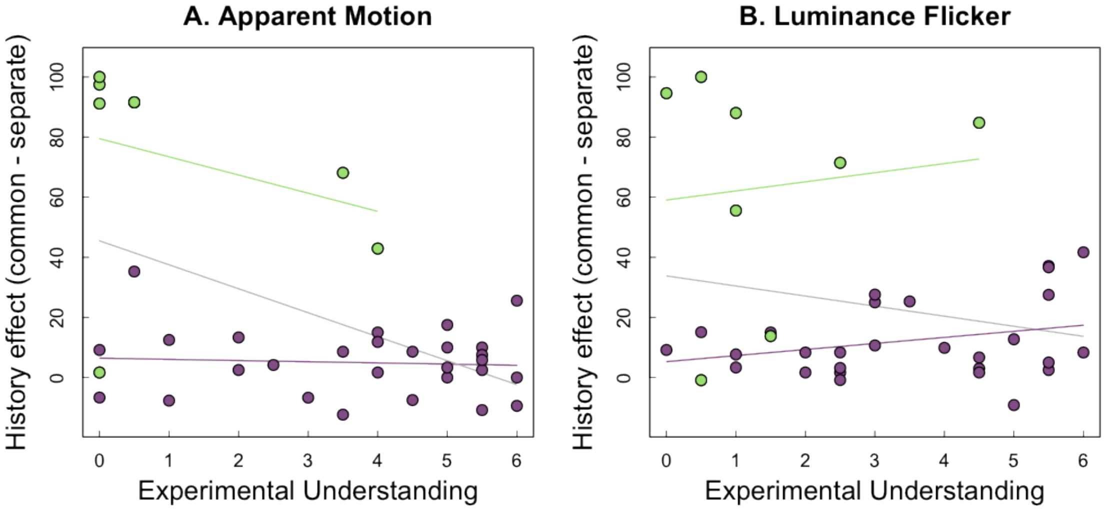

Figure 5.5: Object history effect and experimental understanding score. Comparison of the size of the object history effect (difference of group motion responses in the common versus separate history condition; higher values indicate larger object history effect) with the experimental understanding score of the participant (higher values indicate more understanding). Purple (darker) circles represent the participants that showed ISI effects and green (lighter) circles represent participants that showed no ISI effect. The gray (light) regression line is based on all participants (purple and green circles together), the green (lightest) regression line is based only on the subset of participants that showed no ISI effect, and the purple (dark) regression line represents the subset of participants that showed an ISI effect and that were included in the second analysis. Regression lines for illustrative purposes only. A. Experiment 2: Apparent motion object history. B. Experiment 3: Luminance flicker object history.

Questionnaire

We calculated the difference in mean group motion responses between the common and the separate history condition for each participant and compared this value with the experimental understanding score from the questionnaire. As can be seen in Figure 5.5A, participants' understanding of the experiment varied strongly, ranging from 0 to 6 with a mean of 3.17 and a standard deviation of 2.17. The agreement between raters across all questions was very good, K=0.81, and much greater than would be expected by chance, Z=5.38, $p<.001$ (Question 1: K=0.84, Z=5.18, $p<.001$; Question 2: K=0.73, Z=4.99, $p<.001$; Question 3: K=0.73, Z=5.01, $p<.001$). If all participants are included the size of the history effect is negatively correlated with the experimental understanding (gray regression line), as participants that had a better understanding of the experiment showed smaller history effects, Kendall's $tau=-0.27$, $z=-2.20$, $p=.028$. This significant correlation, however, depends on the inclusion of the eight participants that showed no ISI effect (green circles) with the rest of the participants. If only the participants that showed an ISI effect were included (purple circles), we did not find evidence that the history effect was related to the experimental understanding of the participants (purple regression line), as the same history effects were found for all levels of experimental understanding, Kendall's $tau = -0.04$, $z = 0.30$, $p = .762$.

Discussion

In contrast to Experiment 1, participants perceived more group motion in the common compared to the separate history condition. The analysis of the questionnaire revealed that with the subjects who showed no ISI effect included, the size of the history effect was related to the understanding of the experiment, albeit in the opposite direction of what we expected, as participants with larger knowledge of the experiment showed smaller history effects. Nevertheless, this finding is consistent with the concern that participants' understanding might affect the size of the history effect, therefore highlighting the importance of the separate analyses we conducted. For the subset of participants that showed the usual ISI effect, there was no evidence that the understanding of the experiment had an effect on the size of the history effect, suggesting that for this subset of participants at least the history effect is not likely due to strategic responses biased by the object-history movie. This suggests that the object history can affect how correspondence is solved, if the appropriate, i.e., similar enough, motion information is given during the history.

5.5 Experiment 3: Luminance flicker object history

In Experiment 2 we used apparent motion to manipulate object history and found an influence of it on perceived motion in the Ternus display. Both, the dependent and independent variables therefore involved motion. Given this, a possible concern is that participants may have strategically based their response to the Ternus display on the type of motion that they perceived in the history displays, especially when they were unsure about their percept in the Ternus display. Consistent with this idea the questionnaire based on all participants, including the eight participants without any ISI effect, revealed a correlation between the experimental understanding and the size of the history effect (even though the correlation disappeared for the subset of participants with ISI effect alone). Thus, the concern remains that the results of Experiment 2 might have at least in part affected by strategic responses due to the similarity between the motion during the object-history movie sequence and the subsequent Ternus motion. We therefore conducted a third experiment in which we replaced the object-motion history with an object-luminance history (grouping by common luminance changes, Sekuler & Bennett, 2001). During the history part of the trial, stimuli were stationary but changed in luminance (see Figure 5.2 Experiment 3 and movies 5 and 6), either all together (common history) or independently (separate history). Otherwise, the logic was identical to that of Experiment 2. If the results of Experiment 2 were driven by the matched motion-type decision strategy described above, then perceived Ternus motion should not depend on the object history in Experiment 3. If, however, they were driven by differences in the perceived object organization that was established by object history, then we should observe the same pattern of results in Experiment 3 as we did in Experiment 2.

5.5.1 Method

Participants

Thirty-seven individuals participated in Experiment 3, and received 8 €/h or course credit in compensation for their time. One participant was excluded because he/she had accidentally participated in a similar experiment and was therefore not naïve as to the purpose of this one. A total of 36 participants (24 female, 12 male; mean age = 23.4 years, range: 18-47), therefore, contributed to the reported data. All were naïve as to the purpose of the experiment and reported normal or corrected-to-normal vision.

Apparatus

The same apparatus was used as in the previous two experiments.

Stimuli

The same stimuli were used as in Experiment 2 except that the Ternus elements had no white contours.

Luminance flicker object history

For the luminance object history used in this experiment the elements were presented stationary at the later start position of the first Ternus frame. To create the luminance flicker the color of the object history elements changed between black (RGB: 0, 0, 0; 0 cd/m^2), light gray (RGB: 200, 200, 200; 82 cd/m^2) and dark gray (RGB: 80, 80, 80; 6 cd/m^2). For the common history condition all elements had the same luminance change at the same time in the following sequence: black - light gray - dark gray - black - light gray - dark gray - black. For the separate history the elements started their luminance change one after the other in a random order. Each element passed the following sequence: light gray - dark gray - black - light gray - dark gray. In order to maintain the similarity in appearance across the history and Ternus parts we created a flicker in the luminance history by presenting a blanc frame between each element frame with the same durations as the ISI in the Ternus display in each trial. In total the luminance history lasted about 1,400-2,520 ms and included seven presentations of luminance changes and ISIs.

Task

Participants had the same task as in the other two experiments. At the end of the experiment, they were asked to answer the same questionnaire as in Experiment 2.

Procedure

The procedure was the same as for Experiment 2. Apart from the type of motion history the trial sequence was the same as in Experiment 2.

Design

As in Experiment 2, a 6 (ISI: 0, 10, 20, 40, 80, 160 ms) x 2 (history: common, separate) x 2 (motion direction: left, right) within-subject design was used, in which all factors were counterbalanced and randomly mixed within all trials. Again, each participant completed 240 trials, resulting in 10 observations per condition.

5.5.2 Results and discussion

Ternus percept

Figure 5.3C shows mean group motion responses as a function of object history and ISI for Experiment 3. As before, trials with responses other than the two response keys were eliminated (1.94 %), as well as all trials with RTs larger than 8,000 ms (0.02 %, mean RT 1,193 ms). On this data set we performed a 6 x 2 x 2 repeated-measures ANOVA with the factors ISI, object history and Ternus direction on the mean group motion responses. There was no significant effect of the Ternus direction, $F(1{,}35)=1.29$, $p=.264$, $adj\ \hat{\eta}_p^2=.01$, but a significant interaction with ISI, $F(5{,}175)=2.51$, $p=.032$, $adj\ \hat{\eta}_p^2=.04$. This interaction was due to the ISI of 10 ms, in which the largest difference between the Ternus direction from right to left compared to the other direction was found (6.45 %), but this difference did not reach significance in a Holm's-corrected post-hoc comparison, $t(35)=2.15$, $p_{\text{holm}}=.231$ ($p_{\text{uncorrected}}=.038$), $adj\ \hat{\eta}_p^2=.09$. All other comparisons were not significant, $-1.67>=t_{\text{s}}<=1.32$, $p_{\text{s;}holm}>=.517$, $adj\ \hat{\eta}_p^2<=.05$ (direction differences ranging from -2.5 % for the ISI of 160 ms to 5.5 % for the ISI of 20 ms). No other factor interacted significantly with the Ternus direction, $F_{\text{s}}<=0.89$, $p_{\text{s}}>=.472$, $adj\ \hat{\eta}_p^2<=-.01$. As in the previous experiments we found a significant effect of ISI, $F(5,175)=48.72, p<.001$, $adj\ \hat{\eta}_p^2=.57$. In addition, and most importantly, we found again a significant effect of object history, $F(1,35)=22.96$, $p<.001$, $adj\ \hat{\eta}_p^2=.38$, as well as a significant interaction of both factors, $F(5,175)=3.62$, $p=.008$, $adj\ \hat{\eta}_p^2=.07$. As in the previous experiment we examined this interaction further using Holm's-corrected *t*-tests for each ISI condition separately. These analyses showed that the comparisons were significant for all ISIs, $t_{\text{s}}>=3.40, p_{\text{s;holm}}<=.003$, $adj\ \hat{\eta}_p^2>=.23$.

As in Experiment 2, we repeated the analysis excluding the subset of participants that showed no effect of ISI (N=8; Figure 5.4C). The general pattern of results was the same as before, with the exception that the interaction between the ISI and direction was not reliable $F(5,135)=2.12, p=.067$, $adj\ \hat{\eta}_p^2=.04$ (largest direction difference of 6.7 % was again found for the ISI of 10 ms, $t(27)=1.77$, $p_{holm}=.528$, ($p_{\text{uncorrected}}=.088$), $adj\ \hat{\eta}_p^2=.07$. Most relevant to our question, the main effect of object history was still significant, $F(1,27)=26.24, p<.001$, $adj\ \hat{\eta}_p^2=.47$. Of course, given that the analysis was contingent on it, the main effect of ISI was also significant, $F(5,135)=78.78, p<.001$, $adj\ \hat{\eta}_p^2=.74$. And the interactions between these two factors was significant, $F(5,135)=4.09$, $p=.002$, $adj\ \hat{\eta}_p^2=.10$. Holm's corrected post-hoc comparisons for the history effect at each ISI level showed that with the exception of the 80 ms ISI all comparisons were significant, $t_{\text{s}}>=2.59$, $p_{\text{s;holm}}<=.030$, $adj\ \hat{\eta}_p^2>=.14$. The object history effect at the ISI of 80 ms showed a trend, $t(27)=1.90$, $p_{\text{holm}}=.069$, $adj\ \hat{\eta}_p^2=.07$. No other significant effects were found, $F_{\text{s}}<=1.13$, $p_{\text{s}}>=.333$, $adj\ \hat{\eta}_p^2<=.01$.

Questionnaire

As in the previous experiment we calculated the difference in mean group motion responses between the common and the separate history condition for each participant and compared this value with the experimental understanding score. As can be seen in Figure 5.5B, participants' understanding of the experiment varied again strongly, ranging from 0 to 6 with a mean of 3.03 and a standard deviation of 2.00. The agreement between raters across all questions was substantial, $K=0.61$, and greater than would be expected by chance, $Z=4.71$, $p<.001$ (Question 1: $K=0.52$, $Z=3.95$, $p<.001$; Question 2: $K=0.61$, $Z=4.24$, $p<.001$; Question 3: $K=0.58$, $Z=4.32$, $p<.001$). Furthermore, small and large history effects were found for each level of experimental understanding and there was no evidence for a correlation between both factors. This is true for all participants, including those eight participants that showed no ISI effect (green circles), Kendall's $tau=-0.06$, $z=-0.48$, $p=.631$, as well as the subset of participants that showed an ISI effect (purple circles), Kendall's $tau=0.14$, $z=1.04$, $p=.300$.

Discussion

As in Experiment 2, we found an effect of object history with more group motion reported in the common compared to the separate object history condition. In contrast to Experiment 2, even with all participants included, there was no evidence that the size of the object-history effect was related to the understanding of the experiment, suggesting that it is unlikely that participants adopted a strategy of basing their Ternus responses on the appearance of the object-history displays. Indeed, the effect of the luminance object history was even stronger than the effect of the motion object history. One could speculate that this difference comes from the fact that, in contrast to the motion history, in the luminance flicker history the elements were always presented at the position of the first Ternus frame. This could have strengthened the spatial connection between the elements in the history and the Ternus display and thereby the role of object-based information, i.e., the common or separate luminance change, was more pronounced. Experiment 6 of He and Ooi (1999) is in line with this idea. They presented the two inner Ternus elements flickering together twice prior to the Ternus display. Their results showed more element motion compared to a Ternus display without the flicker history. He and Ooi (1999) suggested that this could be due to a stronger spatial grouping between the flicker elements and the inner Ternus elements leading to a 'no motion' trace for the inner Ternus elements. The results of Experiment 3, therefore, reinforce our interpretation of those from Experiment 2 and together suggest that object history can affect perceived Ternus motion.

5.6 General discussion

This study investigated whether correspondence is resolved based entirely on image-based information (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993) or whether higher-order object-based information is used as well (e.g., Enns et al., 2010; Hein & Cavanagh, 2012; Hein & Moore, 2012, 2014). We used Ternus motion (group versus element) as a measure of the correspondence process and manipulated the object history of the elements used in the Ternus display to bias a perception of them as spatio-temporally grouped together (common object history) or spatio-temporally independent from each other (separate object history). This allowed us to compare correspondence across displays that were identical in terms of image-level information, but differed in terms of object-level information. While the smooth-motion history (Experiment 1) did not influence subsequent Ternus motion, the apparent-motion history (Experiment 2) did. Moreover, the luminance flicker history similarly influenced Ternus motion. Together the results indicate that object-level information can play a role in resolving correspondence.

Why did smooth-motion history fail to influence correspondence (Experiment 1)? We suspect it is because the quality of smooth motion is so different from the quality of apparent motion that the two parts of the trial (object history and Ternus display) were perceived as distinct and therefore did not interact. The fact that object history defined by apparent motion, which is the same type of motion as the Ternus motion, did influence correspondence in the Ternus display (Experiment 2) supports this interpretation. Unfortunately, the other way of matching motion type—smooth motion for both object history and measuring correspondence—cannot be tested with our design because our measure of motion is Ternus motion, a kind of apparent motion. Furthermore, the influence of the luminance flicker history (Experiment 3) also supports this interpretation that there must be a certain element similarity between the history and the Ternus part. In this case both parts could be linked by the similarity of the presentation frequency of the flicker in the history part and the ISI in the Ternus part. Future work that uses a different measure of correspondence or manipulates the similarity between the object history and Ternus display could provide converging evidence and give insight how similar the history part and the Ternus part have to be for the representations to be integrated. In the meanwhile, it is important to note that it is the fact that any version of object history influenced correspondence that indicates that the correspondence process can incorporate object-level information.

A general concern regarding the approach used in this study is that differences in reports of element versus group motion could reflect strategical responses based on the object history part of the trial, rather than differences in object correspondence and the perception of the displays as our logic assumes. Specifically, it is possible that when participants did not have a clear perception of the Ternus motion type, they responded based on the nature of the object history displays, which were clearly separate or common history displays. Because separate and common are conceptually similar to element and group, it may be natural to associate the two types of history display with the two alternative responses to the Ternus display. While this is a concern, there are several aspects of our results that reduce it. First, in Experiments 2 and 3, we administered a questionnaire at the end of the experiment, that asked participants about their understanding of the experiment. There was no correlation between the size of the object history effect and the level of understanding of the relationship between the history conditions and Ternus responses for the subset of participants that showed an ISI effect in Experiment 2, and no correlation at all for Experiment 3. This indicates that at least no conscious strategy of basing responses on object history drove the pattern of responses. Second, the strategy was available in all three experiments, yet object history influenced reports of element versus group motion in Experiments 2 and 3, but not in Experiment 1. Finally, the conceptual similarity between the separate versus common history conditions and the element versus group response choices seems less strong in Experiment 3 (luminance change) than 2 (apparent motion), and yet the effect of object history was largest in Experiment 3. In addition to conscious response biases, the object history may have induced implicit, i.e., unconscious, response biases which are difficult to access. The typical ISI effect that we found in all experiments reduces this concern by showing that participants must have based their responses at least to a certain extent on how they perceived the motion in the Ternus display. Future work that identifies alternative measures of correspondence from which converging evidence can be sought, would be most helpful in assuaging remaining concerns. In the meanwhile, however, the results from these three experiments taken as a whole, suggest that object history can influence the correspondence process separate from response strategies.

The current findings extend our understanding of the correspondence process. First, it is important to note that the conclusion that object-level information influences correspondence does not imply that image-level information does not. Numerous studies have demonstrated the use of image-level information in correspondence, as unprocessed retinal information e.g., luminance, spectral content, textures, that affect motion energy and persistence (e.g., Adelson & Bergen, 1985; Breitmeyer & Ritter, 1986b, 1986a; Petersik & Pantle, 1979; van Santen & Sperling, 1985; Werkhoven et al., 1993). In addition the effect of ISI in the current (and previous) Ternus studies (Breitmeyer & Ritter, 1986a; Petersik & Pantle, 1979) is a demonstration of image-level information influencing correspondence. Second, while previous studies have shown that higher-order information can influence correspondence (Chen & Zhou, 2011; He & Nakayama, 1994; He & Ooi, 1999; Hein & Moore, 2014; Hsu et al., 2015; Ramachandran & Anstis, 1983; Yu, 2000), none before this one has done so without manipulating the level of potentially influencing information by changing the image-level context in which it was embedded. The current findings demonstrate that identical displays can be resolved differently in terms of object correspondence depending on higher-order information that is no longer present.

In summary, we found strong evidence that our visual system uses object-based information as a basis for solving correspondence, further strengthening object-based theories of correspondence (Hein & Cavanagh, 2012; Hein & Moore, 2014). Identical displays were resolved differently in terms of object correspondence depending on their history. Pure image-based theories such as motion-energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993) cannot account for these findings. Overall, it looks like the correspondence process happens at different levels of visual processing, including low but also higher, object-based processing levels. Further studies need to examine more closely the circumstances under which different correspondence processes are engaged.

5.7 Acknowledgements

We thank Jannis Plöger for conducting an experiment for his B.Sc. thesis that greatly informed the experimental choices of Experiment 1. In addition, we thank Maren Klimm and Tana Glemser for help with data collection. We also would like to thank the editor and the two reviewers for their very constructive feedback. This research was supported by DFG project HE 7543/1-1.

5.8 Open Practices Statement

None of the data or materials for the experiments reported here is available online, but can be obtained on request. None of the experiments was preregistered.

Chapter 6 How voluntary spatial attention influences feature biases in object correspondence (Study 3)

Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Attention, Perception, & Psychophysics, How voluntary spatial attention influences feature biases in object correspondence, Stepper, M. Y., Rolke, B., & Hein, E., 2020, advance online publication, 28 June 2019 (https://doi.org/10.3758/s13414-019-01801-9)

The official citation that should be used in reference to this material is:

Stepper, M. Y., Rolke, B., & Hein, E. (2020). How voluntary spatial attention influences feature biases in object correspondence. *Attention, Perception, & Psychophysics, 82,* 1024-1037. https://doi.org/10.3758/s13414-019-01801-9

Attention, Perception, & Psychophysics is available online at:

https://www.springer.com/journal/13414

6.1 Abstract

Our visual system is able to establish associations between corresponding images across space and time and to maintain the identity of objects, even though the information our retina receives is ambiguous. It has been shown that lower-level factors, as for example spatio-temporal proximity, can affect this correspondence problem. In addition, higher-level factors, as for example semantic knowledge, can influence correspondence, suggesting that correspondence might also be solved at a higher object-based level of processing, which could be mediated by attention. To test this hypothesis, we instructed participants to voluntarily direct their attention to individual elements in the Ternus display. In this ambiguous apparent motion display, three elements are aligned next to each other and shifted by one position from one frame to the next. This shift can be either perceived as all elements moving together (group motion) or as one element jumping across the others (element motion). We created a competitive Ternus display, in which the color of the elements was manipulated in such a way that the percept was biased toward element motion for one color and toward group motion for another color. If correspondence can be established at an object-based level, attending toward one of the biased elements should

increase the likelihood that this element determines the correspondence solution and thereby that the biased motion is perceived. Our results were in line with this hypothesis providing support for an object-based correspondence process that is based on a one-to-one mapping of the most similar elements mediated via attention.

6.2 Introduction

Imagine several kittens playing, jumping across each other and all looking very much alike. Keeping track of one of these kittens subjectively seems to be easy for us. But the task is not as easy as it seems, as the information our visual system receives is ambiguous and not continuous, for example because the kittens could occlude each other while moving around or they could reappear spatially shifted from behind a cupboard. Building up correspondence, i.e., establishing associations between images across space and time and maintaining the identity of an object (like our individual kittens), is therefore difficult. How our visual system solves this “correspondence problem” (Ullmann, 1979) has been a topic of research for decades.

Much of this research has used apparent motion displays (Wertheimer, 1912), in which no physical motion is present, but two successively presented objects are perceived as one single moving object. It has been shown that spatio-temporal factors, like the specific spatial distance and the time interval between the occurrences of the objects is important to establish correspondence between the objects and to perceive apparent motion (e.g., Korte, 1915). The correspondence problem is especially obvious for ambiguous apparent motion displays, like for example the motion quartet (von Schiller, 1933), for which different correspondence solutions are possible. The motion quartet consists of two elements presented at opposing edges of a fictive square alternating with two more elements at the other two edges. Depending on the distance and the temporal interval between successively presented elements, the elements can be perceived as moving horizontally or vertically. For example, reducing the horizontal distance between the elements results in the perception of more horizontal movements, and reducing the vertical distance results in the perception of more vertical movements (e.g., Hock, Kelso, & Schöner, 1993; von Schiller, 1933). Temporal factors have also been shown to strongly influence another ambiguous apparent motion display, the Ternus display (Pikler, 1917; Ternus, 1926). The Ternus display usually consists of three elements presented next to each other, shifted by one element position in the next frame (Figure 6.1A). Depending on how correspondence is solved, all elements can be perceived as moving together, i.e., each element moving to the position of the adjacent element (group motion), or one element can be perceived as jumping across the others that remain stationary (element motion). Which type of motion is perceived strongly depends on the time between the successively presented stimuli frames (ISI), as the probability to perceive group motion increases with increasing ISI (Pantle & Petersik, 1980; Petersik & Pantle, 1979). Taken together, it has been shown that the spatio-temporal relationship between stimulus occurrences strongly influences the way correspondence is established and apparent motion is perceived.

Another factor that influences correspondence is feature information (e.g., Alais & Lorenceau, 2002; Casco, 1990; Dawson, Nevin-Meadows, & Wright, 1994; Hein & Cavanagh, 2012; Hein & Moore, 2012; Kramer & Rudd, 1999; Kramer & Yantis, 1997; Moore & Enns, 2004; Petersik & Rice, 2008; Wallace & Scott-Samuel, 2007). For example, Alais and Lorenceau (2002) used a Ternus display with Gabor patches as Ternus elements. These patches could be oriented either collinearly (i.e., gratings oriented horizontally) or parallel (i.e., gratings oriented vertically). The authors showed that more group motion was perceived for the collinearly oriented elements compared to the elements oriented in parallel, suggesting that the feature information of the elements within a frame could influence the correspondence solution. Hein and Moore (2012) manipulated the appearance of the individual elements in the Ternus display, in a way that the elements were either compatible with the element motion percept (element bias, Figure 6.1B top display) or compatible with the group motion percept (group bias, Figure 6.1B bottom display). They showed that the motion percept is shifted in the direction of the bias: For the group bias more group motion was perceived and for the element bias more element motion was perceived compared to a display without such biases, i.e., all elements were identical (Figure 6.1A). Different feature biases, i.e., color, polarity, orientation, hue and luminance, all strongly influenced the correspondence solution (Hein & Moore, 2012). These findings suggest that the identity of the elements across frames also strongly influences correspondence.

Finally, there is evidence that in addition to these rather lower-level factors - spatio-temporal and feature information - even more complex, higher-level information can influence how correspondence is determined. For example, lexical information (Chen & Zhou, 2011; Tse & Cavanagh, 2000), the global context (He & Ooi, 1999; Ramachandran & Anstis, 1983b), the perceived size and lightness (He & Nakayama, 1994; Hein & Moore, 2014), semantic information (Hsu, Taylor, & Pratt, 2015; Yu, 2000), as well as attention

(Aydın et al., 2011; Kohler et al., 2008; Suzuki & Peterson, 2000; Wertheimer, 1912; Xu et al., 2013) modulate the perception of apparent motion. Regarding the influence of attention, Kohler et al. (2008), for example, instructed participants to voluntarily control the perceived motion direction in the motion quartet (von Schiller, 1933). The results showed that participants were able to hold an intended motion direction twice as long than in a passive viewing condition, in which participants were instructed to just report their motion percept. Moreover, they were also able to switch between the vertical and horizontal moving directions twice as fast as compared to the automatic switching in a passive viewing condition. Specifically for the Ternus display, Aydın, Herzog and Öğmen (2011) investigated if the availability of attentional resources influences the apparent motion percept using a dual-task paradigm. In the dual-task condition participants had to detect and count the occurrence of a particular form in a stream of different forms at fixation, in addition to judging the motion of the Ternus display in the periphery. They showed that in the dual-task condition less group motion was perceived compared to a control condition, in which attention was fully available for the Ternus display. The authors concluded that more attention is needed for perceiving group motion compared to element motion. Thus, studies have shown that besides spatio-temporal factors and feature information, even higher-level factors, as for example attention directed toward a particular motion percept, can influence how correspondence is solved.

To explain the influence of these different factors several theories have been developed. Some of these theories emphasize the importance of spatio-temporal factors as for example motion energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993). According to these models, low-level motion detectors, i.e., Reichardt detectors (Reichardt, 1961), compute motion energy based on spatio-temporal activation changes. The direction of these changes then constitutes the basis for determining apparent motion. These theories thus can account particularly well for effects of the ISI and the spatial distance. In line with motion energy models other theories, as for example the spatio-temporal priority theory (Flombaum, Scholl, & Santos, 2012) or the object-file theory (Kahneman, Treisman, & Gibbs, 1992) have highlighted spatio-temporal information as the most important factor to establish correspondence, whereas the identity of an object in terms of its feature information should play no or only a minor role.

To account more directly for the influence of feature information on correspondence, grouping theories have been proposed (e.g., Alais & Lorenceau, 2002; He & Ooi, 1999; Kramer & Yantis, 1997). These theories suggest that correspondence depends on how strongly the objects are associated or grouped with one another based on their features, following for example general grouping principles, as the similarity or proximity of the objects (e.g., Wertheimer, 1923). In particular, Kramer and Yantis (1997) suggested that the stronger the spatial grouping of the elements within a Ternus frame is, the more group motion should be perceived. Moreover, the stronger the temporal grouping of the overlapping elements across Ternus frames is, the more element motion should be perceived. This grouping mechanism could explain for example the findings by Alais and Lorenceau (2002) that collinearly oriented elements increased group motion percepts, as these elements should lead to more spatial grouping due to facilitated contour interactions. In addition, element biases, as shown for example by Hein and Moore (2012), could be easily explained as the spatial grouping (i.e., within a Ternus frame) should be decreased and the temporal grouping (i.e., across both Ternus frames) increased in these conditions.

Finally, to account for the influence of higher-level factors and feature-based biases, Hein and colleagues suggested an object-based theory of correspondence (Hein & Cavanagh, 2012; Hein & Moore, 2014). According to this theory correspondence is established by a one-to-one mapping, i.e., each individual element in one frame is connected with the perceptually most similar element in the next frame. Thus, in contrast to grouping theories, perceived motion is not based on the similarity of all elements within a frame (spatial grouping), but all individual elements across frames are matched based on their identity. Such an object-based theory could explain feature biases, as well as high-level influences of lexical or semantic knowledge on correspondence that are difficult to explain with grouping or motion energy theories. Hein and Cavanagh (2012) suggested that attentional pointers (Cavanagh, 1992; Cavanagh, Hunt, Afraz, & Rolfs, 2010), i.e., spatiotopically organized location pointers that are based on identity information, could connect the most similar elements across frames and track them over space and time, attention thus being a key mechanism for correspondence. Such a correspondence process could happen at a relatively high level of processing such that the similarity of the objects, even in terms of lexical or semantic knowledge and the global context could be taken into account by this type of correspondence process.

The aim of the current study was to directly test the object-based theory of correspondence (Hein & Cavanagh, 2012; Hein & Moore, 2014) by further investigating the potential influence of spatial attention on object correspondence. Following this theory orienting attention to an object should make this object more likely to determine the correspondence solution, as it should orient the attentional pointers toward that object. To test this idea, we run two experiments in which we used a biased Ternus display and instructed participants to direct their attention to one of the elements. In particular, we created a Ternus display containing a competitive bias (Hein & Schütz, 2019), for which differently colored elements were arranged in a way that across both frames the percept was biased toward group motion by one color and element motion by another color (Figure 6.1C). Additionally, we used a classic Ternus display, in which all elements had the same color (Figure 6.1A). Attention was manipulated by using a precue to one of the Ternus elements (e.g., Posner, 1980). The precue consisted of a written word presented at the beginning of each trial that indicated which element of the first Ternus frame participants should attend (left, center, right or all). Participants had to indicate whether they perceived group or element motion in the Ternus display (Ternus task). As orienting attention was not necessary to solve the Ternus task, we used an additional discrimination task to independently verify whether attention had been oriented successfully. The Ternus task was identical in the two experiments, but they differed concerning this additional discrimination task. In Experiment 1, participants were asked to discriminate the orientation of a Landolt C that was briefly presented on one of the Ternus elements. Ternus task and discrimination task were randomly intermixed and thus participants could not anticipate the specific task of a given trial when processing the cue. In Experiment 2, we separated the Ternus task from the discrimination task to avoid potential dual-task costs. In addition, instead of a difficult Landolt C discrimination task, we used a simple gap detection task.

According to the object-based correspondence theory with its attentional pointers (Hein & Cavanagh, 2012; Hein & Moore, 2014), we expected an influence of attention in the competitive display condition, as attending a specific element should make this element more likely to be connected with the element of that particular color feature in the next frame, and thus it should be more likely that this element determines the correspondence solution. In particular, we expected more perceived group motion, when the element containing the group bias was attended (GB-Match condition; center element; cyan element in Figure 6.1C) compared to when the element containing the element bias was attended (EB-Match condition; left element; green element in Figure 6.1C). If attention was oriented to the third element, no particular effect was expected, as there was no direct feature-based match toward the element or group motion percept. In the classic display condition, with all elements being the same, orienting attention to one of the elements should not have any specific effect, as there would be no particular good match of the attended object in the first frame with one of the objects in the second frame in this case.

In contrast to the object-based theory, motion energy models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993) and grouping theories (Alais & Lorenceau, 2002; He & Ooi, 1999; Kramer & Yantis, 1997) do not rely on attention. We nevertheless expected that attention should have a general effect on motion energy or grouping and thus also affect correspondence. In particular, attention studies have shown that the orienting of attention might affect the appearance of the attended stimulus by increasing its contrast (e.g., Carrasco, Ling, & Read, 2004; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Posner, 1980) and its perceived duration (Rolke et al., 2006, 2008; Yeshurun, 2004; Yeshurun & Levy, 2003). These attentional effects on the stimulus should affect the motion energy and grouping of the elements in the Ternus display in the same way in both displays, as these attention effects should be independent of the features of the elements. In particular, for grouping theories, if one of the elements is attended and thus appears to have a higher contrast than the other two elements, the spatial grouping strength between the elements should decrease, thereby increasing the amount of perceived element motion in all cue position and display conditions. For motion energy models, the temporal effect of attention should be most important. When attended, the elements in the first frame should be perceived as longer lasting, thus orienting attention to the second or third element of the first Ternus frame, should “close the temporal gap” between the two successive frames. Because this should decrease the perceived length of the ISI, we expected that group motion percepts should decrease. It is less clear, what one would predict for the situation when attention is oriented to the first element, as motion energy theories are usually based on the central elements. We think, however, that in that case attention should rather increase group motion percepts, as there is no element at that location in the second frame, and thus the system should signal motion of the first element to the adjacent element.

To summarize our hypothesis, for motion energy and grouping theories we predict general attention effects independent of the particular features of the elements, i.e., the same effects for the two display conditions and no interaction between the cue position (first, second, third, or all elements) and the display type (classic or competitive display). In contrast, for the object-based correspondence theory we expect such an interaction, as for the competitive display we predict more element motion percepts in the EB-match condition (attention oriented to the first element) compared to the GB-match condition (attention oriented to the second element), while for the classic display, we predict no particular effect of the attentional manipulation.

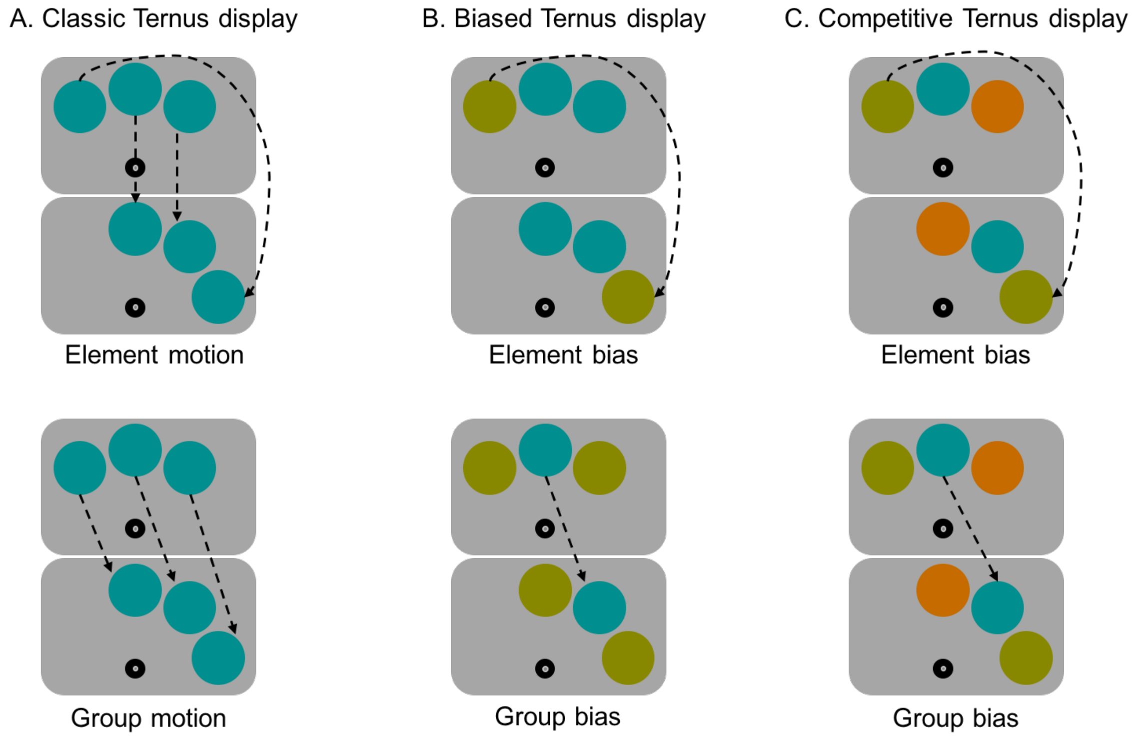

Figure 6.1. Three different Ternus display types. A. Classic Ternus display, in which all elements have the same color and either element motion or group motion can be perceived. B. Biased Ternus display, in which one differently colored element is either compatible with the element motion percept (element bias, here olive green) or the group motion percept (group bias, here cyan). C. Competitive Ternus display, in which all elements have different colors, arranged in such a way that the display contains an element bias (here olive green) and a group bias (here cyan) at the same time.

6.3 Experiment 1

6.3.1 Methods

Participants

A group of 14 participants (9 female) took part in the experiment. The sample size was chosen based on previous studies investigating correspondence and in particular attentional effects on correspondence (Hein & Moore, 2014; Kohler et al., 2008). Their ages ranged between 19 and 25 (M=20.57, SD=1.88 years) and they were mostly students of the University of Tübingen. For their participation, they were compensated with money (8 € per hour) or course credit. All of them were naïve as to the purpose of the experiment and reported normal or corrected-to-normal vision.

Apparatus

The experiments were controlled by a PC with Windows XP as operating system, on which a self-written program was running in MATLAB (Version R2012a, 7.14, Mathworks Inc., MA, USA) using the Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and the EyeLink Toolbox (Cornelissen et al., 2002). A desk-mounted video-based eye tracker (EyeLink 1000 Desktop Mount, SR Research Ltd., Ontario, Canada) was used to monitor central fixation. Eye movements were measured monocular on the right eye with a sampling frequency of 500 Hz. Stimuli were presented on a 17-inch color cathode ray tube monitor (1024 x 768 pixel) with a refresh rate of 100 Hz. Participants conducted the experiment in a dimly lit individual testing room with a fixed viewing distance of 60 cm and their heads stabilized by a chin rest with forehead support.

Stimuli

We use a modified version of a Ternus display (Pikler, 1917; Ternus, 1926) that consisted of two frames with three elements each with a diameter of 1.6° (see Figure 6.2). The elements were aligned on a fictive circle with a diameter of 5.6° centered on the fixation point in the middle of the screen to ensure equal distances from the fixation point to each of the elements. Elements were separated by a center-to-center distance of 2°. The first frame was always presented horizontally centered around the fixation point. The second frame was shifted by one element position to the left or to the right. In the classic Ternus display all elements had the same color (Figure 6.1A) and were cyan (RGB: 0, 142, 142; 15.2 cd/m^2), green (RGB: 136.5, 136.5, 4.5; 15.2 cd/m^2) or orange (RGB: 197, 107, 0; 15.2 cd/m^2), the color being randomly assigned across trials. In the competitive Ternus display the three elements were presented in different colors in the following way: The first element in the first frame and the last element in the second frame were identical, the second elements in both frames were identical and the last element in the first frame and the first element in the second frame were identical (Figure 6.1C). The same three colors as described above for the classic Ternus display were used (cyan, green and orange) and which element was given which color was randomly assigned across trials. The background was presented in gray (RGB: 130.5, 130.5, 130.5; 14.7 cd/m^2) with a luminance set to be as equal as possible to the colors of the Ternus elements. The fixation point was black (0.07 cd/m^2) with a diameter of 0.59° and a smaller gray point in the center (0.15°) to facilitate precise fixations. As an attention cue the word “left”, “center”, “right” or “all” (Arial with a font size of 14) was presented centered 1.5° above the fixation point.

To test whether attention was successfully oriented we replaced the Ternus task with a discrimination task in one third of the trials. For this discrimination task, a Landolt C with a diameter of 0.5° and a line width of 0.03° (1 pixel) was presented centered on one of the elements in the first Ternus frame. The gap of the Landolt C pointed to the left or to the right with a fixed gap size. Gap size depended on the individual performance of the participants in a pretest and ranged between 0.06° (2 pixel) and 0.19° (6 pixel). During the answering period two Landolt Cs with a diameter of 1.6° and a gap size of 0.22° (7 pixel) were presented 3° to the left and right of fixation, one with a gap to the right and one with a gap to the left. Which Landolt C was presented on which side was randomly chosen. All Landolt Cs were black.

Task

For the Ternus task participants had to judge if they perceived all elements as moving together (group motion) or one element as moving separately, jumping across the other two elements (element motion) by pressing the “j” or “f” key, respectively. For the discrimination task, participants had to indicate as correctly as possible the side, on which the Landolt C with the same orientation as the one they saw previously was presented, by pressing the “j” or “f” key for the right or left side, respectively.

Procedure

Participants were informed about the experimental procedure and gave informed consent according to the ethical principles of the World Medical Association (2013; Declaration of Helsinki) prior to their participation. The experiment comprised two sessions of about 2 to 2.5 h, run on two different days. In the first session, a pretest was conducted prior to the main experiment in order to determine individual performance for the Landolt C discrimination. A “1 up 3 down” adaptive staircase (Kaernbach, 1991) was used to find the gap size, for which participants’ performance was about 75 % correct in discriminating the Landolt C. This gap size was used for the first session and further adjusted for the second session, if the error rate in the first session was less than 10 % or above 40 %, by decreasing or increasing the gap size by 1 pixel. After this, each session began with written instructions and clearly distinguishable demonstrations of group and element motion (using extreme ISI of 0 and 160 ms). Each session started with two practice trial blocks of 20 trials. Central fixation was not monitored during the first practice block to familiarize participants stepwise with the experimental procedure and the eye tracker. Participants completed 24 experimental blocks of 20 trials in each of the two sessions.

The time-course of a Ternus task trial is displayed in Figure 6.2. Each trial began with a fixation point. Participants were asked to fixate it and then confirm their fixation by pressing the “j” key. Following this confirmation, the fixation point was presented for another 500 ms, after which the cue (left, right, center or all) was added to the display for 400 ms. Following the cue, the fixation point was presented alone again for another 600 ms. The first Ternus frame then added for 200 ms. After a variable ISI of 0-160 ms, during which again only the fixation point was presented, the second Ternus frame was added for another 200 ms. During the answering period only the fixation point was then presented until the participants gave their response. The next trial started after 500 ms.

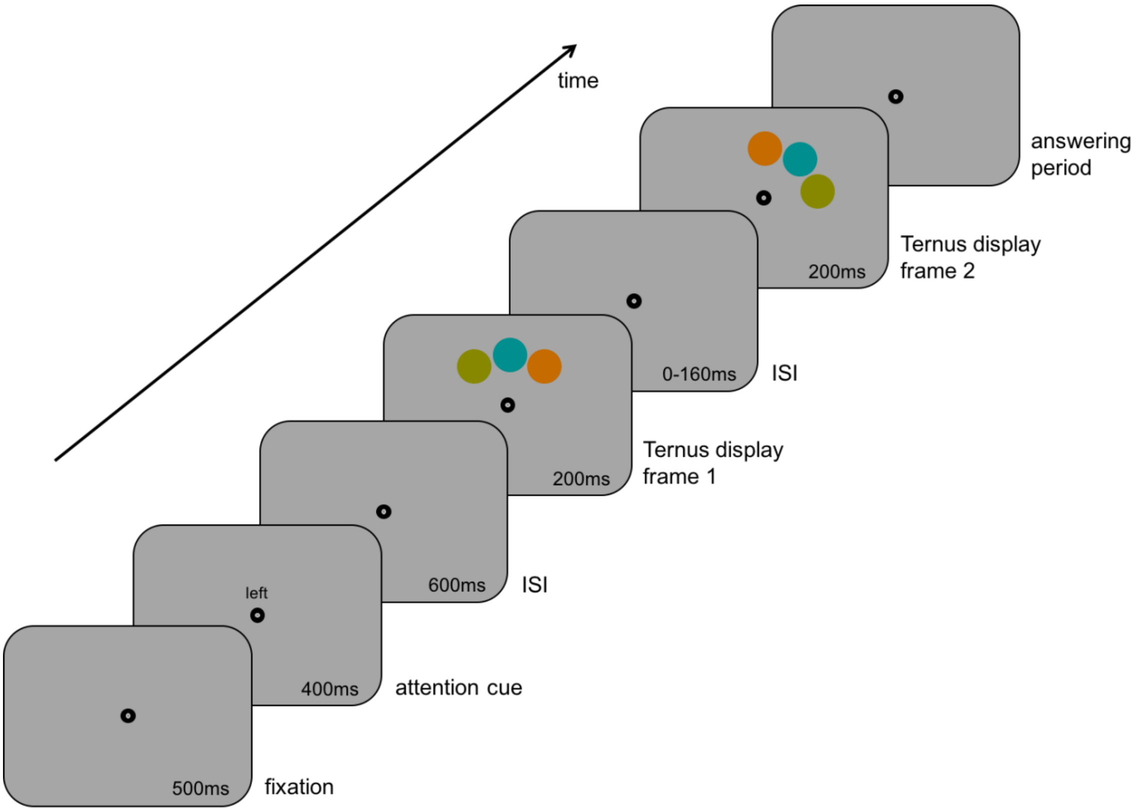

Figure 6.2. Ternus task trial. Illustration of the time course of a single Ternus task trial (shown here is a competitive Ternus display, with a cue to the left element, and a motion direction of the Ternus display to the right).

The trial sequence for the discrimination task was similar to the Ternus task with the following exceptions: After the first Ternus frame was presented for 100 ms, a Landolt C with a gap either to the left or to the right side was added to the display at one of the

Ternus elements for 40 ms. After the Landolt C disappeared, the Ternus frame was presented for another 60 ms. During the answering period a screen with the two possible Landolt Cs and the fixation point was presented until a response was recorded.

Participants’ fixations on the fixation point were controlled from the beginning of each trial (key press by the participant) until the answering period started. To this end a fictive square (2.5°) that was centered around the fixation point was defined, within which participants had to fixate for the fixation to be accepted as valid. Between blocks and, if necessary, within a block the eyetracker was calibrated with a five-point calibration. If fixation was lost a written message reminded participants to fixate (presented for 1,500 ms). Trials, in which fixation was lost, were aborted and immediately repeated.

Design

For the Ternus task a 2 (display type: competitive, classic) x 5(ISI: 0, 20, 40, 80, 160 ms) x 4 (cue position: first, second, third, all) x 2 (motion direction: right, left) within-subjects design was used. All factors were counterbalanced and randomly mixed within all trials. Each participant completed 640 Ternus task trials, resulting in eight observations for each condition. For the discrimination task a 2 (display type: competitive, classic) x 4 (cue position: first, second, third, all) x 2 (target position: cued, non-cued) within-subjects design was used. All factors were counterbalanced and randomly mixed within all trials. For the factor target position and all cue position conditions, but the “all” condition, in the cued condition the Landolt C was presented at one of the elements indicated by the cue (valid condition). In the non-cued condition, the target was randomly presented at one of the two elements not indicated by the cue (invalid condition). If the cue oriented attention to all elements, the target was randomly presented at one of them (neutral condition). Each participant completed 320 discrimination trials, resulting in 80 observations for the neutral condition and 120 observations each for the valid and invalid condition for both display types together.

6.3.2 Results

All statistical analysis were done with R (R Development Core Team, 2008). For analysis of variance (ANOVA), whenever necessary, Greenhouse-Geisser corrections were computed to account for violations of the sphericity assumption. For further post-hoc analysis Holm corrected *t*-tests were conducted. Prior to the inferential analysis trials in which other keys were used than one of the two possible response keys were excluded from the data, 20 trials in the Ternus task (< 1 %) and two trials in the discrimination task (< 1 %).

Ternus Task

For the analysis of the Ternus task, we first submitted the individual mean percentage of group motion responses to a three-factorial analysis of variance (ANOVA) with the factors ISI (0, 10, 20, 40, 80, 160 ms), cue position (first, second, third, all) and display type (competitive, classic). In a next step we looked at the competitive and the classic Ternus display separately, using a two-factorial analysis of variance (ANOVA) with the factors ISI (0, 10, 20, 40, 80, 160 ms) and cue position (first, second, third, all). For the competitive Ternus display, if attention was directed toward the first element, i.e., the element containing the element bias, we will refer to it as the EB-Match condition. If attention was directed to the second element, i.e., the element containing the group bias, we will refer to it as the GB-Match condition. This distinction makes no sense for the classic Ternus display, as there are no biases in this display, we will therefore just refer to these conditions as attention being oriented to the first and second element. For both display types, if attention is directed to all elements, we will refer to it as the neutral condition and if attention is directed to the third element as the third element condition.

The omnibus analysis across all factors revealed a significant main effect of ISI, $F(4,52) = 28.52$, $p < .001$, $\eta_P^2 = .69$ and of cue position, $F(3,39) = 14.93$, $p < .001$, $\eta_P^2 = .53$, but no interaction between both factors, $F(12,156) = 1.47$, $p = .142$, $\eta_P^2 = .10$. In addition, there was a trend for the display type, $F(1,13) = 3.94$, $p = .069$, $\eta_P^2 = .23$, and there was an interaction between the factor display type and ISI, $F(4,52) = 40.29$, $p < .001$, $\eta_P^2 = .76$. Most importantly, there was an interaction between display type and cue position, $F(3,39) = 9.93$, $p < .001$, $\eta_P^2 = .43$. Finally, a trend for the three-way interaction, $F(12,156) = 2.10$, $p = .063$, $\eta_P^2 = .14$, occurred.

To further investigate these interactions separate ANOVAs were conducted for the classic and the competitive Ternus display. Figure 6.3 shows the mean percentages of group motion responses as a function of the attention manipulation and the ISI for both display types separately. The analysis for the classic Ternus display revealed the typical main effect of ISI, $F(4,52) = 42.91$, $p < .001$, $\eta_P^2 = .77$, with an increase of group motion percepts with increasing ISI. The main effect of the factor cue position was also significant, $F(3,39) = 6.29$, $p = .001$, $\eta_P^2 = .33$, but the interaction between ISI and cue position was not significant, $F(12,156)=0.85$, $p=.536$, $\eta_P^2=.06$. Post-hoc tests for the cue position showed that in the neutral condition more group motion was perceived than when attention was oriented to the first element, $t(13)=4.53$, $p_{\mathrm{Holm}}=.003$, $d=1.21$. There was also a trend that more group motion was perceived in the neutral compared to the third element condition, $t(13)=2.99$, $p_{\mathrm{Holm}}=.052$, $d=0.80$. All other comparisons did not reach significance, $t_s <= 2.26$, $p_{s;\mathrm{Holm}} >= .168$, $d_s <= 0.60$.

In contrast to the classic Ternus display the analysis of the competitive display revealed no significant effect for the factor ISI, $F(4,52)=2.09$, $p=.157$, $\eta_P^2=.14$, as the overall percentage of group motion percepts were very similar for all ISI conditions (Figure 6.3, right graph). There was, however, a trend for an interaction between ISI and cue position, $F(12,156)=2.21$, $p=.052$, $\eta_P^2=.15$. To further investigate this interaction, we conducted ANOVAs with the factor ISI for each cue position separately. This revealed a significant effect of the ISI for the GB-Match condition, $F(4,52)=2.57$, $p=.049$, $\eta_P^2=.17$, and the neutral condition, $F(4,52)=2.88$, $p=.031$, $\eta_P^2=.18$. For the other cue positions no significant effects of the ISI were found, $F_s(4,52) <= 2.18$, $p_s >= .134$. Most importantly, the analysis revealed a main effect of the factor cue position, $F(3,39)=15.52$, $p<.001$, $\eta_P^2=.54$. Post-hoc tests for this factor revealed significant differences between all conditions: As can be seen in Figure 6.3 (right graph), the percentage of group motion percepts for the GB-Match condition was higher than for all other conditions: neutral condition, $t(13)=2.60$, $p_{\mathrm{Holm}}=.045$, $d=0.70$, third element, $t(13)=3.41$, $p_{\mathrm{Holm}}=.019$, $d=0.91$, and most importantly EB-Match condition, $t(13)=4.82$, $p_{\mathrm{Holm}}=.002$, $d=1.29$. In addition, more group motion was perceived in the neutral condition compared to the third element condition, $t(13)=2.81$, $p_{\mathrm{Holm}}=.045$, $d=0.75$, and the EB-Match condition, $t(13)= 5.18$, $p_{\mathrm{Holm}}=.001$, $d=1.39$, as well as for the third element condition compared to the EB-Match condition, $t(13)=2.75$, $p_{\mathrm{Holm}}=.045$, $d=0.73$.

Discrimination Task

For the analysis of the discrimination task a two-factorial analysis of variance (ANOVA) with the factors display type (classic, competitive) and cueing condition (valid, invalid, neutral) was conducted for each of the two dependent variables, our main dependent variable error rates, but also on reaction times. We found no significant effects, neither for the mean error rates nor for the reaction times. In particular, for the error rates the results showed no differences between the different cueing conditions (valid: 24.23 %, invalid:

23.72 %, neutral: 22.68 %), $F(2,26) = 0.72$, $p = .497$, $\eta_P^2 = .05$. There were also no other significant effects, $F_s <= 1.38$, $p_s >= .261$. For the reactions times the results also showed no differences between the different cueing conditions (valid: 1003 ms, invalid: 1019 ms, neutral: 1014 ms), $F(2,26) = 0.44$, $p = .652$, $\eta_P^2 = .03$ and no other significant effects, $F_s <= 2.40$, $p_s >= .145$.

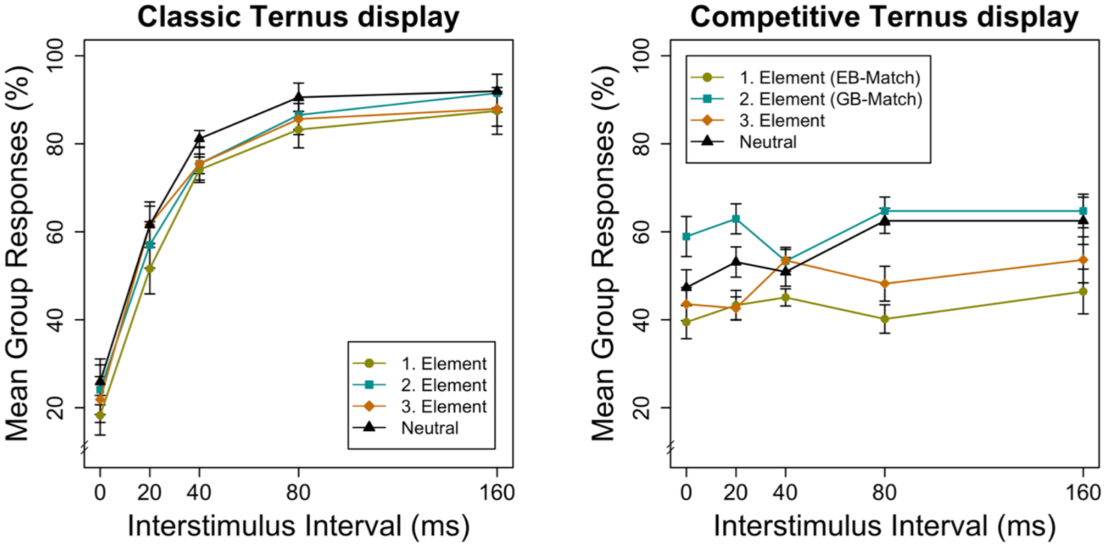

Figure 6.3. Results of Experiment 1. Mean percentage of group motion responses as a function of ISI and cue position. The left graph shows the classic Ternus display and the right graph shows the competitive Ternus display. The error bars represent the within-subject standard errors of the means in each condition (Cousineau, 2005; Morey, 2008).

6.3.3 Discussion

The pattern of results for the Ternus task differed depending on the display type. In particular, participants reported more group motion percepts in the GB-Match than in the EB-Match condition in the competitive display, but there was no difference between the comparable cueing conditions in the classic display. The results therefore support the assumptions under the object-based correspondence theory. In addition to this difference between the GB-Match and the EB-Match condition, the specific pattern of attentional influences in the competitive Ternus display is also interesting. The results showed more group motion in the neutral compared to the EB-Match condition and less group motion in the neutral compared to the GB-Match condition. The difference between the neutral and the GB-Match condition was, however, much smaller than the difference between the neutral and the EB-Match condition. This could be due to more attentional ressources being available for the whole Ternus frame in the neutral condition. Based on Aydın et al. (2011) more attention is needed to perceive group motion than to perceive element motion. Regarding the GB-Match condition this could have led to less group motion percepts due to less attentional ressources available for the whole display, working against the bias toward more group motion. In contrast, regarding the EB-Match condition, both effects should go in the same direction, i.e., one would expect less group motion percepts due to less attentional ressources, as well as due to the element bias. The finding that the neutral and the GB-Match conditions were much more similar to each other than the neutral and the EB-Match conditions could also be explained within the framework of grouping theories (e.g., Kramer & Yantis, 1997), if one assumes that orienting attention toward an element changes its appearance, i.e., making it more salient by increasing its contrast (e.g., Carrasco, Ling, & Read, 2004; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Posner, 1980). This would reduce spatial grouping and therefore would decrease group motion percepts in the EB-Match and the GB-Match condition compared to the neutral condition. The general difference between the GB-Match and EB-Match condition, however, should remain constant as the availability of less attentional ressources or the reduced spatial grouping should have led to the same decrease of group motion percepts in both conditions, which means our main conclusion that attentional pointers seem to strongly influence the correspondence solution in the competitive display is supported by the specific pattern of results.

In addition, the results for the competitive display also showed more group motion in the neutral compared to the third element condition, the third element condition being more similar to the EB-Match condition. As described above directing attention to an individual element could have influenced the availability of general attentional resources or the spatial grouping which could explain this effect at least to some degree. The third element condition, however, is also a special case concerning its feature, as the third element in the first frame is compatible with the first element in the second frame. Thus, if these elements are connected across frames via attentional pointers this could be perceived as a special case of element motion, in which the center elements swap places, at least for some participants. Further studies have to investigate this possibility more closely.

Unexpected under the object-based correspondence theory, the analysis also revealed that attention had an effect in the classic Ternus display, as more group motion was perceived in the neutral condition, in which attention was directed to all elements, compared to all other conditions, although this difference only reached significance compared to when attention was directed to the first element. This attentional effect is in line with the study by Aydın et al. (2011), as if more attentional ressources were available for the whole Ternus frames in the neutral condition when attention was directed to all elements, this could have lead to more perceived group motion, than when attention was directed toward an individual element. This attentional effect in the classic display could also be explained within the framework of grouping theories (e.g., Kramer & Yantis, 1997), as more group motion should be perceived in the neutral condition due to stronger spatial grouping compared to when attention is oriented to individual elements.

As a control for the allocation of attention on the Ternus element, participants had to perform a discrimination task at cued and non-cued elements. We found, however, no effect of the attentional manipulation on discrimination performance. This was unexpected, as it has been shown that orienting attention to a target evokes faster responses and better performance in similar discrimination tasks (e.g., Posner, 1980; Posner, Snyder, & Davidson, 1980; Yeshurun & Carrasco, 1999). We assume that we did not obtain the expected cueing effect because the discrimination task was intermixed with the Ternus task. This dual task situation might have produced switch costs (e.g., Monsell, 2003), as participants might have focused on the main task, i.e., the Ternus task which occurred in two-thirds of all trials. This assumption is supported by the rather high RTs in the discrimination task. To test this idea that the intermixing of the discrimination task with the Ternus task prevented the attentional effect to be measurable in the discrimination task, we run Experiment 2 and blocked the two tasks.

6.4 Experiment 2

In this experiment, we used a blocked instead of a mixed design for the discrimination and the Ternus task. Moreover, to increase the importance of the cue, we made the cue predictive by presenting the target in 75 % of the trials at the cued position (valid condition), instead of 50% as in Experiment 1. Finally, we made the discrimination task easier and focused on reaction times as a measure of attentional allocation. In particular, we asked participants to detect a large cut-out on the top or the bottom of one of the Ternus elements, instead of a difficult Landolt C discrimination as in Experiment 1.

6.4.1 Methods

Participants

A new sample of 20 participants (13 female) contributed. We increased the sample size compared to Experiment 2 for the following reasons: First, we balanced the order of the tasks (Ternus and discrimination task) as well as the finger-to-key assignment for the RT based discrimination task, which resulted in a multiple-of-four sample size number. Second, we expected that the block-wise separation of the Ternus task and the discrimination task might weaken the attentional effect in the Ternus task block, as directing attention was not necessary to solve this task. In increasing the sample size, we aimed to discover the potentially smaller effect. Participants' ages ranged between 19 and 33 (*M*=24.15*,* *SD*=3.80 years). Originally, 24 participants took part in this experiment. We excluded three participants from our analysis, because they could not maintain fixation in more than 30% of the trials. One additional participant was excluded because this participant showed an atypical decrease of group motion responses with increasing ISI in the neutral condition of the classic Ternus display. This pattern is in the opposite direction of the typical increase of group motion with increasing ISI, suggesting that this participant might have mixed up the response keys.

Apparatus

The apparatus was the same as in Experiment 1.

Stimuli

The stimuli were the same as in Experiment 1 for the Ternus task. For the discrimination task we used a circular cut-out at the top or bottom of one of the Ternus elements. This cut-out was created by presenting a background-colored circle (diameter of 1.2°) on top, centered either at the top or at the bottom edge of the Ternus element.

Task

The Ternus task was identical to Experiment 1. For the discrimination task, participants had to indicate as quickly and as correctly as possible with their index fingers, whether the cut-out in the Ternus element appeared at the top or the bottom, by pressing the "z" key for top ("z" on the German keyboard corresponds to "y" on the American keyboard) and the “b” key for bottom. The assignment of the fingers to the keys was counterbalanced across participants.

Procedure

The general procedure was identical to Experiment 1 with the following exceptions. First, no pretest for the discrimination task was necessary, as we used an easy cut-out discrimination. Second, the Ternus task and the discrimination task were run in two different sessions on two different days (order balanced across participants). In each session participants completed 32 experimental blocks of 20 trials.

The time-course of the Ternus task was identical to Experiment 1 (Figure 6.2). For the discrimination task, after the first Ternus frame was presented for 100 ms, the cut-out at one of the Ternus elements was presented for 100 ms, before the Ternus display disappeared. The next trial started 500 ms after a response was recorded.

Design

The design for the Ternus task was exactly the same as for Experiment 1. For the discrimination task, the design was the same with the exception that the cue was now predictive, as the target was presented at the cued position in 75 % of the trials (valid condition). Participants completed 640 discrimination task trials. This resulted in 160 observations for the neutral condition, 120 observations for the invalid condition and 360 trials for the valid condition for both display types together.

6.4.2 Results

Prior to the inferential analysis we excluded trials in which other keys were used than one of the possible response keys. These were 71 trials in the Ternus task (< 1 %) and 44 trials in the discrimination task (< 1 %).

Ternus task

As in Experiment 1, the omnibus analysis with the factors ISI, cue position and display type revealed a significant main effect for ISI, $F(4,76) = 23.53$, $p < .001$, $\eta_P^2 = .55$ and for cue position, $F(3,57) = 9.16$, $p = .001$, $\eta_P^2 = .33$, but no interaction between both factors, $F(12,228) = 0.83$, $p = .616$, $\eta_P^2 = .04$. In addition, and in contrast to Experiment 1, the factor display type was significant, $F(1,19) = 9.18$, $p = .007$, $\eta_P^2 = .33$, with overall more group motion percepts in the classic ($M=75.54\%$) compared to the competitive display ($M=58.45\%$). As in Experiment 1 there was an interaction between the factor display type and ISI, $F(4,76)=33.06$, $p<.001$, $\eta_P^2=.64$. Most importantly and replicating Experiment 1, we found a significant interaction between display type and cue position, $F(3,57)=4.40$, $p=.019$, $\eta_P^2=.19$. The three-way interaction between all three factors was also significant, $F(12, 228)=1.83$, $p=.044$, $\eta_P^2=.09$.

As in Experiment 1 we conducted separate ANOVAs for the classic and the competitive Ternus display to gain insights into the specific pattern of results for each display type. Figure 6.4 shows the mean percentages of group motion responses for the Ternus task as a function of the attention manipulation and the ISI for each display type condition. For the classic Ternus display there was an effect of ISI, $F(4,76)=35.22$, $p<.001$, $\eta_P^2=.65$, with an increasing percentage of group motion percepts with increasing ISI. There was also a main effect of cue position, $F(3,57)=4.30$, $p=.019$, $\eta_P^2=.18$. Descriptively the pattern of results was similar to those of Experiment 1, as the largest difference in group motion percepts was between the neutral condition and the first element, followed by the third element condition and the second element condition. Holm corrected post-hoc tests, however, revealed no significant difference between any of the individual comparisons, $t_s<=2.54$, $p_{s;\mathrm{Holm}}>=.119$, $d_s<=0.56$. As in Experiment 1 the interaction of the factor ISI and cue position was not significant, $F(12,228)=1.69$, $p=.132$, $\eta_P^2=.08$.

For the competitive Ternus display, consistent with Experiment 1, there was no main effect for the factor ISI, $F(4,76)=0.61$, $p=.559$, $\eta_P^2=.03$. In contrast to Experiment 1, there was no trend for an interaction between the factor ISI and cue position, $F(12,228)=1.06$, $p=.393$, $\eta_P^2=.05$. Most importantly, as in Experiment 1, there was a main effect of the factor cue position, $F(3,57)=8.52$, $p=.002$, $\eta_P^2=.31$ (see Figure 6.4, right graph). Holm-corrected post-hoc tests for this factor revealed the following differences: Most importantly, group motion percepts in the GB-Match condition were higher than in the EB-Match condition, $t(19)=2.73$, $p_{\mathrm{Holm}}=.040$, $d=0.61$. In addition, the group motion percepts were higher in the GB-Match compared to the third element condition, $t(19)=3.29$, $p_{\mathrm{Holm}}=.019$, $d=0.74$, and higher for the neutral compared to the EB-Match condition, $t(19)=3.16$, $p_{\mathrm{Holm}}=.020$, $d=0.71$, and the third element condition, $t(19)=3.77$, $p_{\mathrm{Holm}}=.008$, $d=0.84$. In contrast to Experiment 1, the GB-Match condition did not differ from the neutral condition, $t(19)=0.76$, $p_{\mathrm{Holm}}=.919$, $d=0.17$, and there was also no difference between the EB-Match and the third element condition, $t(19)=0.48$, $p_{\mathrm{Holm}}=.919$, $d=0.11$.

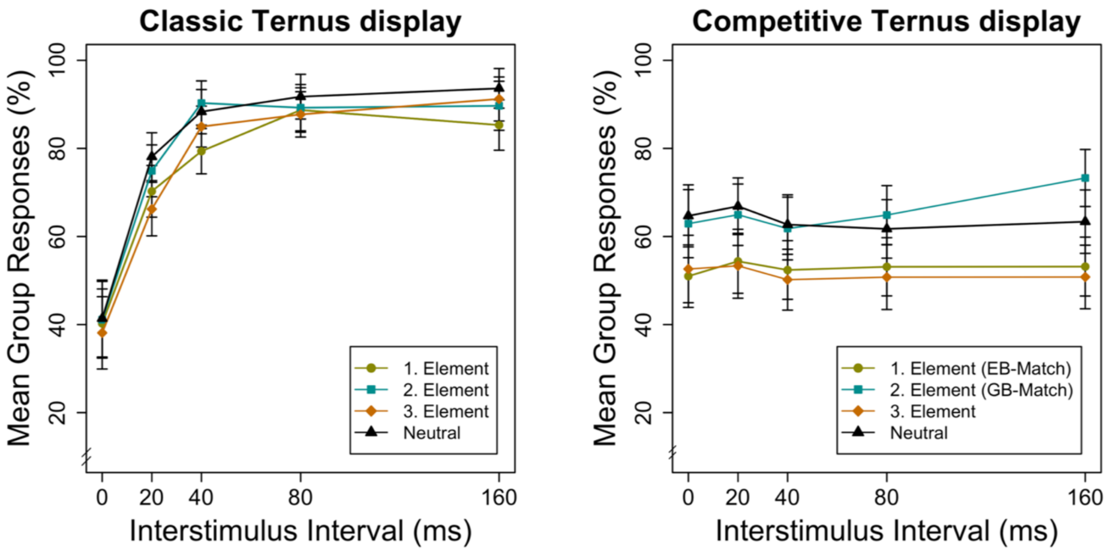

Figure 6.4. Results of Experiment 2. Mean percentage of group motion responses as a function of ISI and cue position. The left graph shows the classic Ternus display and the right graph shows the competitive Ternus display. The error bars represent the within-subject standard errors of the means in each condition (Cousineau, 2005; Morey, 2008).

Discrimination task

We excluded trials in which participants made an error (6.06 %). We additionally excluded trials with RTs +/- 3 * *SD*s of the mean RT for each participant and condition (1.29 %). In contrast to Experiment 1, there was a significant main effect for the factor display type, $F(1,19)=20.89$, $p<=.001$, $\eta_P^2=.52$ and, most importantly, a significant effect of the factor cueing condition (valid: 457 ms, invalid: 496 ms, neutral: 465 ms), $F(2,38)=13.29$, $p=.001$, $\eta_P^2=.41$. Post-hoc tests revealed that participants reacted significantly faster in the valid compared to the invalid cueing condition, $t(19)=3.71$, $p_{\mathrm{Holm}}=.003$, $d=0.83$, and in the neutral compared to the invalid cueing condition, $t(13)=3.93$, $p_{\mathrm{Holm}}=.003$, $d=0.88$. There was a trend for a difference in RT between the valid compared to the neutral cueing condition, $t(19)=1.81$, $p_{\mathrm{Holm}}=.085$, $d=0.41$. The interaction between the factors cueing condition and display type was not significant, $F(2,38)=1.07$, $p=.354$, $\eta_P^2=.05$. The analysis for the mean error rates revealed no effect for the factor cueing condition (valid: 6.05 %, invalid: 5.81 %, neutral: 6.35 %), $F(2,38)=0.58$, $p=.566$, $\eta_P^2=.03$. There was also no difference in error rates between the two display types, $F(1,19)=0.13$, $p=.725$, $\eta_P^2=.01$, but a trend for an interaction between display type and cueing condition, $F(2,38)=2.87$, $p=.069$, $\eta_P^2=.13$.

6.4.3 Discussion

In this Experiment there was a cueing effect in the RT-based discrimination task. This result shows that the cue was in principle able to orient attention toward specific elements of the Ternus display. Why did we find the expected attentional effect in this experiment, but not in Experiment 1? We think that several factors might have contributed to the occurrence of the cueing effect. First, we strenghtened the impact of the cue by enhancing its predictive value, second, we measured RT in a simple discrimination task instead of error rates, and third, we separated the discrimination task from the Ternus task. This latter change might have reduced switching costs (e.g., Monsell, 2003) between the Ternus task and the discrimination task, which might have masked attentional effects in Experiment 1. No matter which change might be the most important factor for establishing the cueing effect, we successfully showed that the cue employed in the two experiments had the potential to orient attention in the Ternus display.

Importantly, in this experiment, we replicated our most interesting result for the Ternus task, i.e., the differential effects of attention in the competitive compared to the classic Ternus display. In particular, there were only small influences of attention in the classic Ternus display, but much larger attention effects in the competitive display. The reason why the attentional influence in the classic display and partly in the competitive display condition were slightly reduced in Experiment 2 might be due to the blocked design employed in Experiment 2. Here, an attentional orientation was task-irrelevant for the Ternus task session and thus participants might have sometimes neglected the cueing instruction in this session. Overall, however, the most important effect of the attentional orientation, i.e., that in the competitive display more group motion percepts were reported for the GB-Match condition compared to the EB-Match condition, was replicated and is in line with the object-based correspodence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014).

6.5 General Discussion

The object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014) suggests that correspondence is established by a one-to-one mapping of the elements that are perceived as most similar across both frames and that attention could mediate this process. In order to test this theory, we investigated if directing attention voluntarily to a specific object influences the correspondence solution. We used the Ternus display (Pikler, 1917; Ternus, 1926), an ambiguous apparent motion display, for which either element or group motion can be perceived depending on how correspondence is solved, and which has been shown to be strongly influenced by feature information (Casco, 1990; Hein & Moore, 2012; Kramer & Yantis, 1997; Petersik & Rice, 2008). In particular, we created a competitive Ternus display containing a bias toward element motion and a bias toward group motion within the same display by changing the color of the elements (Hein & Schütz, 2019). We also used a classic Ternus display, in which all elements had the same color. Attention was either directed to one individual or to all Ternus elements in both display conditions. Based on the object-based correspondence theory, we expected that attending an individual element would increase the impact of that element for solving correspondence in the competitive Ternus display, but not in the classic Ternus display. For the competitive Ternus display this should lead to more group motion percepts, if the element containing the group bias was attended (GB-Match condition) compared to less group motion percepts, if the element containing the element bias was attended (EB-Match condition). In the classic display condition, however, no effect of attention was expected under the object-based correspondence theory, as all elements had the same color and thus a similarity based one-to-one mapping across frames would not find a specific match. Therefore, directing attention to a particular element should not affect the correspondence solution. The results were in line with the predictions under the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014), as we found different effects of the attentional manipulation in the two display conditions in both experiments. In particular, across all ISI, more group motion was perceived in the GB-Match compared to the EB-Match condition in the competitive display, while no difference in group motion responses was found, when attention was oriented to the first and the second Ternus element in the classic display. This suggests that the attended element was weighted stronger for solving correspondence, i.e., the corresponding one-to-one mapping across frames was more likely to be selected.

Interestingly, we also found different effects of the ISI in the two display types for the Ternus task. For the classic Ternus display we found the typical ISI effect (e.g., Pantle & Petersik, 1980; Petersik & Pantle, 1979), suggesting that spatio-temporal factors had a strong effect on correspondence in this case (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993; Flombaum, Scholl, & Santos, 2012; Kahneman, Treisman, & Gibbs, 1992). In the competitve Ternus display, however, the motion percept was nearly independent of the ISI. The competitive feature biases in this display seem to mostly override the ISI effect, which provides further evidence that under the right circumstances features can have a strong effect on solving correspondence (e.g., Alais & Lorenceau, 2002; Casco, 1990; Dawson, Nevin-Meadows, & Wright, 1994; Hein & Cavanagh, 2012; Hein & Moore, 2012; Kramer & Rudd, 1999; Kramer & Yantis, 1997; Moore & Enns, 2004; Petersik & Rice, 2008; Wallace & Scott-Samuel, 2007). In line with the object-based theory of correspondence (Hein & Cavanagh, 2012; Hein & Moore, 2014), the independence from the ISI in the competitive display could be explained in the way that the feature information (i.e., the color of the elements) was more dominant than spatio-temporal factors and therefore correspondence might have been established mainly on the one-to-one mapping of the elements in this display condition.

Moreover, these different ISI effects in the two display conditions were modulated by the attentional manipulation in different ways. While for the classic display there was a very strong increase of group motion percepts with increasing ISI for all cueing conditions, for the competitive display only in the neutral condition a reduced (Experiment 1) or even no (Experiment 2) influence of ISI was found. This finding suggests that the influence of attention on correspondence was rather minimal when no features differentiated the Ternus elements and correspondence was more likely to be mediated by motion energy (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Werkhoven et al., 1993) and/or temporal grouping (He & Ooi, 1999; Kramer & Yantis, 1997). When the features made the elements distinguishable, however, the effect of attention on correspondence became very strong, suggesting that correspondence was more likely to be mainly mediated by object-based mechanisms. Thus, overall, the effect of attention on correspondence seems to be dependent on the correspondence mechanism. And which correspondence mechanism(s) is/are at work seems to be dependent on the complexity of the display, i.e., in our case whether the elements are distinguishable by different features or not.

With the discrimination task we used in Experiment 2 we showed that attention was oriented successfully, which is in contrast to the lack of a cueing effect in Experiment 1.

However, as the cue was the same in both Experiments and as the main results in the Ternus task, especially the difference between the EB-Match and the GB-Match condition in the competitive Ternus display were the same in both Experiments, one can assume that the attentional manipulation was also successful in Experiment 1. We hypothesize that we could not measure an attentional effect in the discrimination task because of switch costs between the Ternus task and the discrimination task, concealing the attention effects in this experiment. In contrast to the failure to measure an attentional effect in the discrimination task, the attentional manipulation seems to have even been stronger in Experiment 1 than in Experiment 2 for two reasons: First, concerning the effects in the competitive display in Experiment 1 the difference between the EB-Match and the GB-Match was larger and there were more modulations of the group motion percepts in the other cueing conditions compared to Experiment 2. Second, in the classic Ternus display there was a general attentional effect in Experiment 1, which can be explained by grouping theories (e.g., Kramer & Yantis, 1997) or effects of the availability of attentional resources (Aydın et al., 2011) that was smaller in Experiment 2. It is possible that due to the blocked design in Experiment 2 attention was oriented less strongly in the Ternus task, in which the attentional orienting did not help to do the task, reducing the strength of the attentional effects in the competitive display, and even eliminating some of the smaller, more general attentional effects in both display conditions.

To summarize, we found that spatial attention influences how feature biases are weighted to determine correspondence. Up to now studies have mainly shown that attention can influence correspondence, if participants voluntarily envision a particular motion percept or voluntarily track a certain motion path (Kohler et al., 2008; Suzuki & Peterson, 2000; Wertheimer, 1912; Xu et al., 2013). Our results extend these findings by showing that voluntarily attending to a certain object also influences how correspondence is determined. Moreover, we found small general attention effects, on the motion percept in the Ternus display, especially in Experiment 1, which were present in both display types and could be explained by grouping theories (e.g., Kramer & Yantis, 1997) or a general effect of the availability of attentional resources (Aydın et al., 2011). Finally, our findings of an increase of the motion percepts in the direction of the bias for the competitive Ternus display in both Experiments support the object-based theory of correspondence (Hein & Cavanagh, 2012; Hein & Moore, 2014), which suggests that correspondence in this display condition is based on the perceived similarity of the individual objects in a one-to-one mapping, likely connecting these objects across space and time via attentional pointers

(Cavanagh, 1992; Cavanagh, Hunt, Afraz, & Rolfs, 2010). Taken together, correspondence seems to be a complex process that can happen at different levels of processing depending on the specific requirements and the complexity of the particular display the visual system has to interpret. Moreover, the effect of orienting attention toward individual elements seems to be dependent on the type of correspondence mechanism that is engaged.

6.6 Acknowledgements

We thank Michael Hoyer, Maren Klimm, Marius Mackamul and Carlos Medina for help with data collection. We also would like to thank the editor Michael D. Dodd and three anonymous reviewers for their very constructive feedback helping to clarify our ideas and suggesting the changes for the discrimination task in the second experiment. This research was supported by DFG project HE 7543/1-1.

6.7 Open Practices Statement

None of the data or materials for the experiments reported here is available online, but can be obtained on request. None of the experiments was preregistered.

Chapter 7 General discussion

This dissertation is concerned with investigating the influence of higher-level factors on solving the correspondence problem (Dawson, 1991; Ullmann, 1979), specifically, the question of how our visual system knows which object went where. The primary aims were to identify the level of visual processing at which correspondence is established and to test the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). According to this theory, the correspondence process takes place at an intermediate to high level of visual processing (see Figure 2.3) based on object representations. In contrast, image-based theories, including motion-based (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992) and feature-based theories (e.g., Kramer & Yantis, 1997), suggest that only image-based information directly available from the retina at relatively low levels of visual processing is used for solving correspondence. With regard to this main question, we identified three specific questions to be investigated: (1) Does object-based information at an intermediate level of visual processing influence correspondence? (2) Is such an influence independent from information available at low levels of visual processing? (3) Could attention be the mechanism to connect corresponding objects across space and time? More precisely, the questions can be described as follows: (1) Although, in line with the object-based correspondence theory, there are several findings regarding the influence of higher-level object information on correspondence (e.g., Chen & Zhou, 2011; He & Nakayama, 1994; Hein & Moore, 2014; Hsu et al., 2015; Yu, 2000), these are relatively rare compared to systematic research regarding the influence of low-level information (e.g., Adelson & Bergen, 1985; Flombaum & Scholl, 2006; Kahneman, Treisman, & Gibbs, 1992; Scholl, 2001; van Santen & Sperling, 1985; Werkhoven, Sperling, & Chubb, 1994). Therefore, these higher-level findings are less systematic with respect to the (higher) level of visual processing they investigate and focus most on the influence of object-independent higher-level factors such as spatial attention (e.g., Aydın et al., 2011; Xu et al., 2013) and short-term memory content (e.g., Hein et al., 2021; Scocchia et al., 2013). In particular, there is a lack of research investigating the intermediate level of higher-level visual processing. This intermediate level refers to the level of visual processing up to which image-based information is processed, where object representations are established and contextual information available in a scene is incorporated. Since research on this intermediate level of visual processing is underrepresented, the first two studies of this thesis investigated the influence of object-based information at this processing level, looking at the influence of perceived size (Study 1: Stepper, Moore, et al., 2020a, Chapter 4) and object history (Study 2: Stepper, Moore, et al., 2020b, Chapter 5). (2) A clear differentiation between the influence of low-level image-based information and mid-level object-based information is often difficult, since by manipulating object-based information, image-based information is often affected as well. For example, additional filters to manipulate the perceived lightness, as used by Hein and Moore (2014), could possibly lead to changes in image-based information, such as the luminance contrast, in the area surrounding the to-be-judged element, hence possibly influencing its perception at a low level. Such differences in image-based information could at least partially explain the results through image-based theories (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997) suggesting a correspondence mechanism based on, for example, motion energy. This issue was addressed in Study 2 (Stepper, Moore, et al., 2020b, Chapter 5), which investigated whether different object representations are able to influence the correspondence solution in identical Ternus displays (Pikler, 1917; Ternus, 1926). (3) Finally, to further test a specific proposition of the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014), mainly that attentional pointers (Cavanagh, 1992; Cavanagh et al., 2010) could establish correspondence between elements at the object level, we asked whether voluntary spatial attention could influence the correspondence solution. Previous research had shown that the availability of attention and focus on a certain motion percept can influence correspondence (Aydın et al., 2011; Kohler et al., 2008; Suzuki & Peterson, 2000; Xu et al., 2013), but no study as yet has tested the specific predictions that attention could serve as a mechanism for establishing correspondence. Therefore, Study 3 (Stepper, Rolke, et al., 2020, Chapter 6) of this thesis was concerned with investigating this potential role of attention by orienting attention to specific elements of the Ternus display. In the following paragraphs, the findings will be summarized and discussed with regard to the initial aims of the present work.

7.1 Findings and contribution to open research questions

The aim of Study 1 (Stepper, Moore, et al., 2020a) was to find further evidence indicating whether object-based information, especially information at an intermediate level of visual processing, can be used to establish correspondence. To manipulate such object-based information, a Ponzo-like illusion was used, which alters the perceived size of objects depending on where they are presented on a background with pictorial depth-cues (Gregory, 2009; Rock, 1983). We presented a Ternus display (Pikler, 1917; Ternus, 1926) on such a background either at the perceived near or the perceived far position in order to manipulate the perceived size of the Ternus display. Therefore, while the perceived size at an intermediate level of visual processing should be different, the physical size, the input at a lower level of visual processing, was identical. Participants' task was to judge the Ternus display according to whether they perceived element or group motion. Results showed that the motion percept in the Ternus display was influenced by the perceived size of the Ternus display and not its physical size. This strongly indicates that the perceived size and therefore the percept at an intermediate level of visual processing, where information about depth and the background structure has been processed, can influence object correspondence. So far, it has been shown that correspondence can be solved at the level of amodal completion, the perceptual completion of objects that appear to extend behind occluding surfaces (He & Nakayama, 1994; Hein & Moore, 2014, Exp. 1), and lightness perception, after accounting for the luminance context in the scene (Hein & Moore, 2014, Exp 2). This work (Stepper, Moore, et al., 2020a) extends the existing literature by showing that the level of size constancy, after a Ponzo-like depth illusion was established, is also used for solving correspondence. Moreover, evidence that the Ponzo illusion involves higher-level information (Brown et al., 2018) additionally indicates that correspondence could be established even after the level at which amodal completion takes place. However, image-based information could also play a role, as in order to manipulate the perceived size of the Ternus display, a Ponzo-like illusion background was presented. The converging lines in this background differentiate the image-based information, such as the luminance contrast between the perceived near and the perceived far condition. One aspect of Study 1 (Stepper, Moore, et al., 2020a) that makes the influence of this image-based information less likely is that the illusory background is always present, and we did not compare conditions with and without this background. For example, such a comparison was performed in the study by Hein and Moore (2014), in which conditions with and without a filter/occluder were compared. In addition, a control background containing no depth information was used, but this background was divided into three areas of different color, similar to the Ponzo background. This at least ensured that this subdivision of the background could not explain any differences found with the Ponzo-like background. On the other hand, the possibility exists that the differences in image-based information may have influenced the correspondence solution at least in part and that these differences could be explained by image-based theories (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997).

The aim of Study 2 (Stepper, Moore, et al., 2020b) was therefore to investigate whether information available at an intermediate level of processing can influence correspondence independently of image-based information. To access the correspondence process, we again used the Ternus display. Prior to presenting the Ternus display, we presented a history of the Ternus elements in which the elements moved or changed their luminance (depending on the experiment) either together or independently of each other. This object history was used to bias the perception of the elements as either spatio-temporally grouped together (common) or spatio-temporally independent (separate). If this kind of information, most likely stored at a mid-level object representation, can influence correspondence, then the bias of the object history (common vs. separate) should affect the perception of the Ternus display presented after the object history. Importantly, in this experimental design, the image-based information at the time of the Ternus display was exactly the same for both history conditions. Results showed that the object history had an influence on the correspondence solution with more group motion percepts in the common history compared to the separate history condition. Furthermore, we found that the spatio-temporal properties of the history needed to be similar to the one in the apparent motion Ternus display in order to have an effect. This suggests that elements in the history and Ternus display will only share an object representation if they appear to be similar, meaning that only in that particular case will the history have an influence. Our findings (Stepper, Moore, et al., 2020b) show that correspondence in identical Ternus displays can be established differently based on information about objects, which is not present in the image but is stored in a representation of the object. Therefore, the results indicate that information at the object level, independent from image-based information, can be used for the correspondence process. Strong evidence that the information stored in object representations was indeed used to establish correspondence comes from the finding that the object history was only found to have an effect if its spatio-temporal information was similar to the spatio-temporal properties of the Ternus display, which was the case for the apparent motion history (Exp. 2) and the luminance flicker history (Exp. 3). This suggests that in order for information stored in the object representation to have an influence on correspondence, the elements in the Ternus display have to be identified as the same elements as in the history. Therefore, with regard to the first research question, both studies

(Study 1: Stepper, Moore, et al., 2020a & Study 2: Stepper, Moore, et al., 2020b) provide further evidence for the influence of information at an intermediate level of visual processing. They also extend previous findings, the majority of which investigated higher levels of visual processing, such as semantic or lexical knowledge (e.g., Chen & Zhou, 2011; Hsu et al., 2015), by showing that the perceived size and the spatio-temporal information stored in object representations can influence correspondence. A major contribution of Study 2 (Stepper, Moore, et al., 2020b) was to find evidence for the second research question, which was to differentiate between the levels of visual processing at which correspondence is established. By showing that correspondence can be established at the intermediate level of visual processing, independent from image-based information, Study 2 provides additional support for the influence of higher processing levels demonstrated in Study 1 (Stepper, Moore, et al., 2020a) and previous work (e.g., He & Nakayama, 1994; Hein & Moore, 2014).

The aim of Study 3 (Stepper, Rolke, et al., 2020) was to investigate whether attention, specifically attentional pointers (Cavanagh, 1992; Cavanagh et al., 2010), could be a possible mechanism mediating the correspondence process, as suggested by the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). In particular, attentional pointers are suggested to connect corresponding elements across space and time based on the elements' identity. To test this idea, we investigated whether directing voluntary attention to a specific element in the Ternus display influences how correspondence is established. We used a special competitive version of the Ternus display, in which differently-colored elements across frames were ordered so that they contained a bias toward group and element motion at the same time. Directing attention to one element should lead to a stronger weighting of that element and influence the correspondence solution based on its identity, leading to more motion percepts in the direction of the bias of this element. In addition, a classic Ternus display was used, in which all elements had the same color and therefore no bias toward a specific motion percept. Following the attentional pointer idea, directing attention to a specific element should not lead to a difference in perceived motion within the classic display. In line with this prediction, for the competitive display, results showed more group motion when the element containing the group bias was attended compared to more element motion when the element containing the element bias was attended. No such differences were found attending those elements in a classic Ternus display. Therefore, the results of Study 3 (Stepper, Rolke, et al., 2020) suggest that attention is involved in establishing correspondence and thus could mediate the connection of corresponding elements, as suggested by the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). Image-based theories (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997) cannot account for such findings, as the attentional effects cannot be explained solely by the differences available at the image-based level, in this case, the different colors. Also, the results could not have arisen because of low-level changes of the element due to the allocation of attention, like an increase of contrast (e.g., Carrasco et al., 2004; Posner, 1980) or perceived duration increase (e.g., Rolke et al., 2008; Yeshurun, 2004) as such differences would have led to the same changes in the competitive and the classic Ternus display. Thus, this study provides the first evidence for an attention-based correspondence mechanism in line with the attentional pointer idea (Cavanagh, 1992; Cavanagh et al., 2010) as suggested in the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014).

In sum, all three studies (Stepper, Moore, et al., 2020a, 2020b; Stepper, Rolke, et al., 2020) provide further evidence that correspondence can be established at a higher level of visual processing, using attention and object-based information such as the perceived size of objects and information about the history of objects to solve correspondence. Therefore, besides answering the specific questions, all three studies of this thesis could provide further evidence in line with the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014).

7.2 The Ternus display – a special case?

The Ternus display (Pikler, 1917; Ternus, 1926) was used for all studies conducted in the context of this dissertation because it is well-suited for investigating the factors that determine how correspondence is solved (for an overview see Hein, 2017; Petersik & Rice, 2006). One reason for this is that the Ternus display is an apparent motion display. In such displays, the correspondence problem, the question which object went where, becomes especially salient given that motion can be perceived despite no motion being physically present. Furthermore, the Ternus display is an apparent motion display that is ambiguous, meaning that very different motion percepts (element or group motion) can be perceived depending on how correspondence is established. Therefore, the correspondence process can be investigated in an elegant way by manipulating factors of interest and testing whether they influence the motion percept. Despite these more general reasons why the Ternus display is well-suited to investigating object correspondence, the Ternus display has also shown influences of information at different levels of visual processing. It has been shown that spatio-temporal information, like the timing between the Ternus frames (ISI; e.g., Pantle & Petersik, 1980; Petersik & Pantle, 1979) or the spatial distance between the Ternus elements (e.g., Casco, 1990; Pantle & Petersik, 1980), and also feature information such as color and orientation of the Ternus elements (e.g., Alais & Lorenceau, 2002; Hein & Moore, 2012) can be used for solving correspondence in the Ternus display. In addition to the influence of this information available at a low level of visual processing, which can be explained by image-based theories (motion-based theories: e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; feature-based theories: e.g., Kramer & Yantis, 1997), the Ternus display has also been used to collect evidence that object-based information available at a higher level of visual processing can influence the correspondence solution (e.g., Aydın et al., 2011; Chen & Zhou, 2011; He & Nakayama, 1994; He & Ooi, 1999; Hein & Moore, 2014; Hsu et al., 2015; Ramachandran & Anstis, 1983a; Yantis, 1995; Yu, 2000), in line with the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). Since the goal of this thesis was to further investigate the influence of higher-level processes and to test the object-based correspondence theory, the Ternus display was clearly the best option for all studies conducted as part of this dissertation.

Although the Ternus display seems to be the best option for the present investigation, the question arises whether the findings from this work can be transferred to other correspondence situations or are specific for the Ternus display. One claim could be that whether object-based correspondence mechanisms influence the correspondence solution, as shown within the present work for the Ternus display, could depend on whether the correspondence problem is actually one of *object* correspondence. Moore et al. (2019) point out that a distinction has to be made between object correspondence and motion correspondence. Despite being related, they can be distinguished by the breadth of their constructs. Motion correspondence always involves a direct perception of motion and is limited to a relatively short time scale, as in apparent motion, where a motion percept is either present or not. In contrast, in object correspondence, motion over a longer period of time could also be implicitly assumed, as in the case of an object that disappears, reappears after a certain time and at a certain distance, and is recognized as the same object that must have moved even though this motion was not directly perceived. For example, if we meet a fellow student in a café an hour after we both left the university, then we know that he, too, has moved there without us directly perceiving his motion. Moore et al. (2019) suggest that the Ternus display could be more complex than simple apparent motion displays and therefore, for the Ternus displays, object-based correspondence processes could be used in addition to motion correspondence processes. On the basis of this proposal, it is possible that the complexity of a display could determine whether the correspondence problem is one of motion correspondence or object correspondence. Therefore, the type of motion (apparent motion vs. continuous motion) the display comprises and the complexity of the display (simple vs. complex) could span a continuum from motion correspondence to object correspondence. At one end of the continuum are most apparent motion displays. At the other end of the continuum, continuous motion displays, like the object reviewing paradigm (Kahneman et al., 1992) or the perceived causality display (Michotte, 1963) are generally agreed to be a problem of object correspondence because the question is not on the focus in continuous motion itself but rather on the maintenance of object identity over a longer time scale (Moore et al., 2019). Thus, how transferable the findings of the present work are could depend on the position of the Ternus display along this continuum relative to other correspondence situations being investigated.

So where do we place the Ternus display on this continuum? There are several indications that the Ternus display is actually presenting a problem of object correspondence, as suggested by Moore et al. (2019). First, the findings of this thesis, which can confirm and extend previous findings, provide a strong argument that the Ternus display presents a problem of object correspondence by showing that object-based correspondence mechanisms are involved in establishing correspondence. Furthermore, in line with Moore et al. (2019), although the Ternus display is an apparent motion display and therefore should be giving participants a motion correspondence problem, there is evidence that it has a special status among apparent motion displays due to its complexity. Compared to other apparent motion displays, the Ternus display consists of more elements compared to simple apparent motion displays (e.g., Kolers, 1972; Korte, 1915; Wertheimer, 1912), which consist of only two elements, and even split motion displays (e.g., Nishida & Takeuchi, 1990; Werkhoven et al., 1993, 1994), in which one element is presented in the first frame and two in the second frame. In addition, the motion percept also seems to be more complex in the Ternus display than in other apparent motion displays. Compared to simple apparent motion displays, in which movement between two objects is either perceived or not, the Ternus display is an ambiguous apparent motion display in which different motion percepts can be perceived. Moreover, the type of motion is more complex even compared to other ambiguous apparent motion displays. While very simple motion percepts often compete regarding their motion direction, like horizontal versus vertical motion in the motion quartet (Navon, 1976; von Schiller, 1933), the motion direction in the Ternus is always the same, but the motion type differs. In most cases, one of the two "classic" motion types is perceived: either all elements move together (group motion) or one element jumps across the other(s) (element motion). In principle, however, any element in the first frame can be connected to any element in the second frame, giving eight different possibilities for a three-element Ternus display. An example of another motion percept is the three-dimensional motion perception in a two-element Ternus display found by Dodd, McAuley, and Pratt (2005). They connected the two Ternus elements with a line expecting less element motion due to a within-frame grouping of the elements, but instead of perceiving more group motion, participants perceived one element rotating around the other.

In sum, the higher complexity of the Ternus display relative to other apparent motion displays means that the Ternus display might be special with respect to the influence of object-based correspondence mechanisms: it is presenting a problem of object correspondence, like continuous motion display, despite being an apparent motion display.

In view of these arguments for a special status of the Ternus display, we return to the question posed above: to what extent are the findings of this work transferable to other correspondence situations? According to the distinction between motion correspondence and object correspondence based on the motion type and complexity of the display as discussed above, the findings regarding the influence of object-based correspondence mechanisms within this work should be transferable to displays with continuous motion, like, for example, the causality display (e.g., Moore et al., 2020), as with their associated higher complexity they are grouped as problems of object correspondence. In contrast, the situation seems to be less clear for apparent motion displays other than the Ternus display. With their lower complexity, it is possible that the results from the Ternus display are not transferable given that correspondence in these displays might be solved with image-based correspondence mechanisms only. However, previous findings (e.g., He & Nakayama, 1994; Ramachandran & Anstis, 1983b; Tse & Cavanagh, 2000), have shown that higher-level factors such as amodal surface completion, the context, or the lexical information may also have an influence in apparent motion displays other than the Ternus display. For example, Tse and Cavanagh (2000) have shown that the knowledge about writing a Chinese character can influence the apparent motion percept. In the first phase the Chinese character was presented to the participant as a whole. In the second phase it was presented again, but this time stroke by stroke. Participants had the task to judge which type of motion, (no motion, left to right, or right to left) they perceived for the presentation of the last stroke. Chinese participants were more likely to perceive motion in the direction the stroke would be drawn by hand. Such findings indicate that certain apparent motion displays other than the Ternus display may also be considered as presenting an object correspondence problem. Given that these other apparent motion displays seem to be less complex then the Ternus display, this suggests that factors other than the complexity of the display itself could determine the complexity of the correspondence situation. Therefore, the results obtained within the present work and other studies using the Ternus display might not be purely Ternus-specific with regard to apparent motion.

The overall complexity of a correspondence situation could also depend on the context in which the display is presented or on the participants' task, which could also be interrelated with the complexity of the display itself. The participants' task itself could influence whether the correspondence problem to be solved is one of motion or object correspondence. For simple apparent motion, the task is often to report whether motion was perceived or not. The task thus results in a focus on the direct perception of movement, favoring motion correspondence. In contrast, the task in ambiguous apparent motion displays is to decide *which* motion percept was perceived, for example in the Ternus display, “group motion” or “element motion.” Further, in continuous motion displays, the task may not even be to name the motion perceived. For example, in the object reviewing paradigm, the task is to name a target letter, appearing on objects in motion, as quickly as possible (Kahneman et al. 1992). Here, the focus of the task moves more and more toward the objects: which objects belong together or their identity, and so the problem becomes more and more one of object correspondence. Thus, it would be conceivable that whether a study is investigating motion or object correspondence depends not only on the complexity of the display itself, including how many elements it contains and which motion type it has, apparent or continuous, but also on the task to be solved. Another point that could influence the complexity of the correspondence situation is the context in which the display is presented, whether or not there is additional information given to solve correspondence and the type of this information. The context's influence can be seen directly with the Ternus display. The perceived motion percept in a classic version of the Ternus display, consisting of three elements in each frame which are all identical, is mainly influenced by the spatio-temporal information between frames, like the ISI (e.g., Pantle & Petersik, 1980; Petersik & Pantle, 1979). In such a case, object-based mechanisms seem unnecessary to the correspondence solution and the correspondence problem could be simply one of motion correspondence, explained by image-based mechanisms alone (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997). In contrast, if the context becomes more complex, object-based correspondence mechanisms might be used for establishing correspondence. For example, complexity can be added by introducing a scene containing additional information, like depth cues in Study 1 (Stepper, Moore, et al., 2020a) by the Ponzo-illusion background, or by simply adding more color introducing a bias toward one or the other (or both, as in the case of the competitive display; see Hein et al., 2021) motion percept(s) (Hein & Moore, 2012; see Figure 2.2B), and this might cause participants to use object-based correspondence mechanisms. In sum, the overall complexity and the resulting correspondence mechanism that is accessed, motion or object correspondence, could therefore strongly depend on the available information, and the task in addition to the general display type, including its complexity and motion type. This in turn should also apply to continuous motion displays.

In summary, the Ternus display could be a special case as it is more complex than other (ambiguous) apparent motion displays and therefore a problem of object correspondence for which object-based correspondence mechanisms apply. However, this does not necessarily imply a limitation in the transferability of the evidence found within this work, heavily based on evidence from Ternus displays, because the complexity of a correspondence situation seems not only to be determined by the display itself, but may also depend on other factors such as the task or the context. Therefore, it can be assumed that the object-based correspondence mechanisms for which evidence was found here using the Ternus display will also apply to other correspondence situations, as long as they are object correspondence problems.

7.3 Model of the correspondence process

With the aim of developing an overall model for the correspondence process and the two previously-described correspondence mechanisms, image-based and object-based, we are faced with three possibilities: correspondence is solved (1) purely based on image-based correspondence mechanisms, (2) purely based on object-based correspondence mechanisms, or (3) based on both together. The idea that only image-based correspondence mechanisms are used for establishing correspondence as proposed by image-based theories (motion-based theories: e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; feature-based theories: e.g., Kramer & Yantis, 1997), can, however, immediately be considered disproved. Many previous studies have demonstrated that information processed up to the object-level, for example, the perceived attributes of objects (e.g., perceived surface: He & Nakayama, 1994; perceived lightness: Hein & Moore, 2014), the motion context (He & Ooi, 1999; Yantis, 1995) or even higher-level factors like attention (e.g., Kohler et al., 2008; Suzuki & Peterson, 2000; Xu et al., 2013) and visual working memory content (e.g., Hein et al., 2021; Scocchia et al., 2013) can influence the correspondence solution. Moreover, within the present work, we found further evidence for the influence of such object-based information especially at an intermediate level of visual processing (Study 1: Stepper, Moore, et al., 2020a; Study 2: Stepper, Moore, et al., 2020b) and for an attention-mediated object correspondence mechanism (Study 3: Stepper, Rolke, et al., 2020). In these cases, the correspondence problem seems to be one of object correspondence, as suggested by Moore et al. (2019), and image-based theories alone cannot explain how correspondence is solved, as the information used for establishing correspondence within these studies is not available at these low levels of visual processing. Therefore, image-based correspondence mechanisms can only explain how correspondence is established if the correspondence problem is one of motion correspondence, as discussed above and suggested by Moore et al. (2019). A model of object correspondence therefore cannot rely purely on image-based correspondence mechanisms.

The opposite possibility regarding a model of the correspondence process is that only object-based correspondence mechanisms, taking place at an intermediate to high level of visual processing, determine the correspondence solution, as is suggested in the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014). This seems plausible, as evidenced by Study 1 (Stepper, Moore, et al., 2020a) of this work. In this case the physical size of an object, as image-based information, is further processed up to the perceived size of that object, becoming object-based information and incorporating context information like depth-cues. Based on this object-based information, correspondence is then established between the objects perceived as most similar. In such cases, information at higher levels of visual processing is used, and the correspondence solution can be explained by object-based correspondence mechanisms alone. However, there are also studies that have shown that information available at a lower level of visual processing can influence correspondence, like the ISI (e.g., Pantle & Petersik, 1980; Petersik & Pantle, 1979) or the physical size (e.g., Breitmeyer & Ritter, 1986b, 1986a; Casco, 1990; Petersik & Grassmuck, 1981) of the elements in the Ternus display. These studies, however, were not designed to differentiate the level of visual processing at which correspondence is solved, and this information was also further processed up to an object-level of the visual system. As long as image-based information is not changed by further processing, the information is still available at an object level. The physical size information, for example, is available unchanged at an object level as long as there is no further information like depth cues available. Therefore, one possibility regarding a model of the correspondence process is that only object-based correspondence mechanisms are determining correspondence based on object-based information. Support for this possibility comes from Study 1 (Stepper, Moore, et al., 2020a) and Study 2 (Stepper, Moore, et al., 2020b) of this work, showing evidence for solving correspondence based on object-based information at an intermediate level of visual processing. In Study 1, it has been shown that changes in the perceived size of the Ternus display had an influence in the same direction as for the physical size (e.g., Casco, 1990; Exp. 2). This suggests that both correspondence solutions, whether based on the perceived size or on the physical size, could be determined at the same intermediate level of visual processing. In Study 2, it was shown that the object history, consisting of spatio-temporally grouped or separated information, can influence correspondence. This indicates that spatio-temporal information that is further processed up to the object level and stored within object representations, like the ISI leading to the typical ISI effect within the Ternus display, could be used for solving correspondence at a higher level of visual processing. Therefore, regardless of whether image-based or object-based information is available, the correspondence problem could always be solved by object-based correspondence mechanisms. The possibility of this one-level correspondence process model is depicted in Figure 7.1A. All information is processed up to the object level of visual processing, at least an intermediate level, is then stored in object representations and can be used for solving correspondence based on object-based correspondence mechanisms.

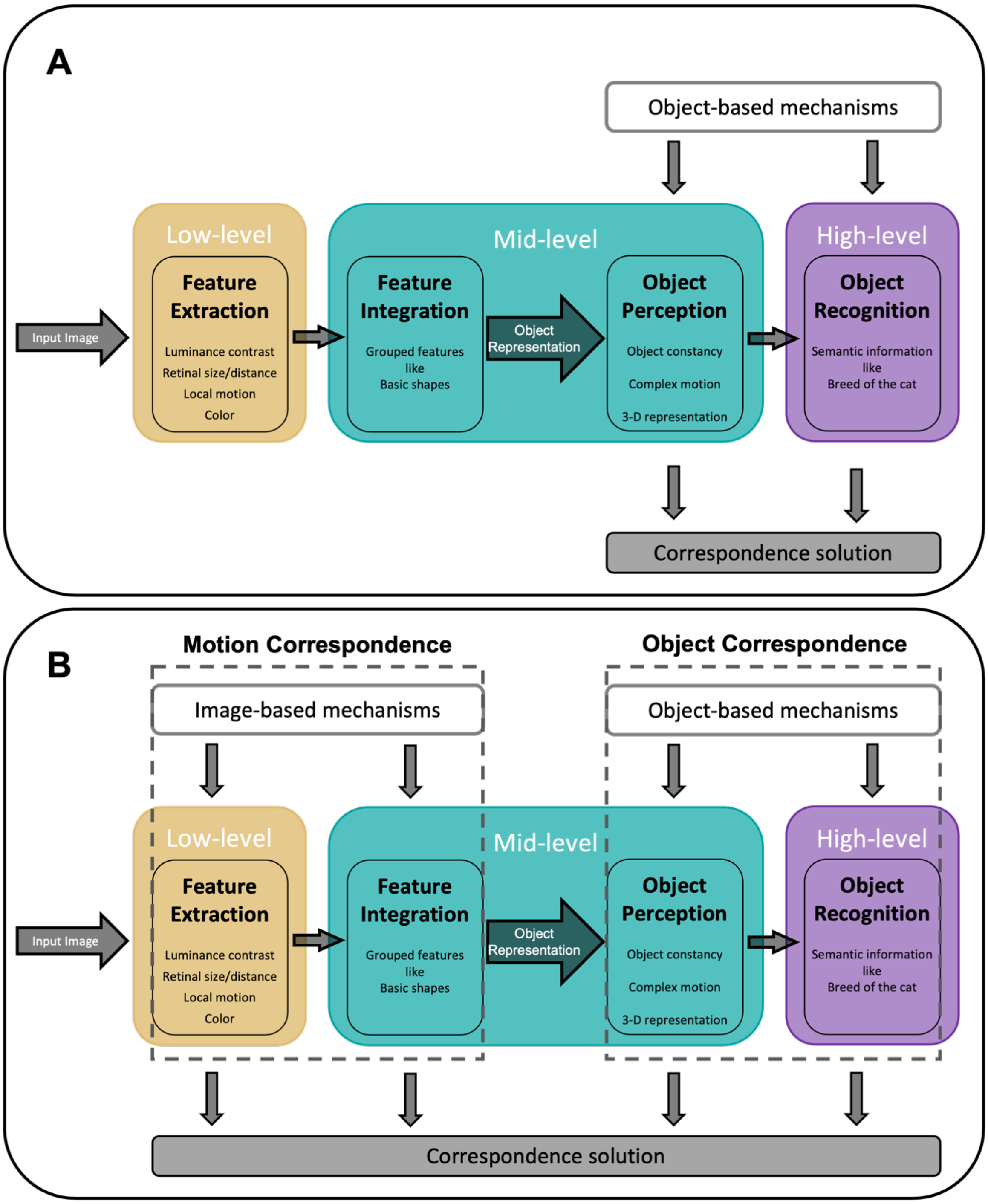

Figure 7.1: Models of the correspondence process. **A.** One-level correspondence process model. It shows the possibility that the whole correspondence process could rely solely on object-based correspondence mechanisms. Within this model, image-based information is further processed up to the object level at higher levels of visual processing. Correspondence is then solved based on all information available about the objects stored in object representations. **B.** Two-level correspondence process model. Within this model, the correspondence process can take place at different levels of visual processing. Motion correspondence can be solved solely based on low-level information, whereas more complex object correspondence is solved based on all information available at the object level. The final correspondence solution in this case could depend on the information available and on how reliable the information is.

The third possibility of a model of the correspondence process is that both, image-based as well as object-based correspondence mechanisms, are used to determine correspondence. This would mean that correspondence can be established at low levels as well as at higher levels of visual processing. Image-based correspondence mechanisms, suggested in image-based theories (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997) use information available at low levels, and object-based correspondence mechanisms, as proposed in the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014) use information at the object level. Assuming that both correspondence mechanisms are used to establish correspondence, the question arises as to *when* each correspondence mechanism takes effect and how they might interact with each other to determine the final solution. The choice of correspondence mechanism used by the visual system could depend on the information available in the visual scene, especially on its complexity. In line with this idea is Moore et al.'s (2019) suggestion, which points out that a distinction has to be made between two correspondence processes: motion correspondence and object correspondence. This distinction, discussed in the chapter above, could be linked to the correspondence mechanisms in such a way that image-based correspondence mechanisms could be sufficient to establish motion correspondence, while more complex object correspondence would also require mechanisms at the object level – that is, object-based correspondence mechanisms. Whether the actual correspondence problem is one of motion or of object correspondence could depend on the type of motion and the complexity of the display as well as the complexity of the task or the context. For example, if the information available exists at the image-level, as in a simple apparent motion task, viewers would be faced with a problem of motion correspondence requiring only image-based correspondence mechanisms, which take place at a low level of visual processing. In contrast, with increasing complexity of the display and especially the information available, as in Study 1 (Stepper, Moore, et al., 2020a) with its additional depth-cue information altering the perceived size of the Ternus elements, viewers encounter an object correspondence problem, requiring object-based mechanisms taking place at a higher level of visual processing. The distinction between motion correspondence and object correspondence as pointed out by Moore et al. (2019) therefore aligns with the idea that both image-based and object-based correspondence mechanisms are used for establishing correspondence, and which one is used depends on the complexity of the available information. A motion correspondence problem is solved using lower-level image-based information and an object correspondence problem is solved using higher-level object-based information. This two-level model of the correspondence process is depicted in Figure 7.1B.

The question remains: which of these two models, one-level or two-level, best describes the correspondence process? According to Occam's razor, the simplest theory should be preferred above all others, the one-level correspondence process model should be chosen, as a model consisting of one (object-based) theory seems to be much simpler than a model consisting of two (image-based and object-based) theories. Furthermore, one could argue that, in the course of visual processing, low-level information is in any case processed up to higher levels for object perception and recognition. Therefore, it seems plausible to use this information at the object level, consisting of both further processed information and unchanged low-level information. On the other hand, a two-level model might seem more plausible if there are indeed two types of correspondence, motion correspondence and object correspondence, as suggested by Moore et al. (2019). The authors suggest that motion correspondence operates on a short time scale and involve a direct perception of motion. Concerning the limitation to a short time scale it makes sense that motion correspondence is solved at low levels of visual processing using image-based mechanisms as low-level information is available first, and image-based mechanisms like Reichardt detectors (Reichardt, 1961) are simple and fast. Establishing motion correspondence could therefore serve to rapidly provide us with a perception of motion. This could be important for our survival, for example when judging the right moment to cross a busy road between fast-moving cars. Object correspondence, on the other hand, is suggested to take place at the object level, recognizing motion between objects even over longer time scales and when no direct motion is perceived. The two-level correspondence model therefore seems to be more appropriate than the one-level correspondence model. It better reflects the evidence regarding two different correspondence processes, motion correspondence and object correspondence (Moore et al., 2019), and the influence of different processing levels, image-based and object-based, on how correspondence is solved. Further evidence needs to be collected in order to be able to distinguish between the one-(object-)level and the two-level model of the correspondence process. However, this thesis has provided convincing evidence that a model representing the correspondence process has to contain at least object-based mechanisms taking place at higher levels of visual processing.

7.4 Open questions & Outlook

This work has provided evidence for the influence of mid-level information in establishing correspondence (Study 1: Stepper, Moore, et al., 2020a & Study 2: Stepper, Moore, et al., 2020b), a solid start in the process of filling in the lack of research on the precise level of visual processing at which correspondence is solved. Overall, the number of studies regarding the influence of higher levels on correspondence is still small compared to the decades of research on the influence of low-level information. More systematic research of different levels of higher-level visual processing is needed to complete the picture and understand which levels are involved in solving correspondence. This could be done by using paradigms targeting specific levels of visual processing. For example, different masking paradigms, like crowding and interocular suppression, could be useful in teasing apart these levels. Crowding (Levi, 2008; Whitney & Levi, 2011) describes the phenomenon that an element in the periphery of the visual field cannot be recognized if it is presented simultaneously with adjacent flanking stimuli that are close to the target but irrelevant to the participant’s task. In interocular suppression (Tong et al., 2006), the dominant eye is presented with a salient stimulus and the non-dominant eye is simultaneously presented with a less salient stimulus. The less salient stimulus is suppressed and only the salient stimulus is perceived. Evidence was found that crowding occurs after interocular suppression, meaning at a later level of visual processing (Breitmeyer, 2015). Combining a measure of correspondence, for example the Ternus display, with these masking paradigms could give a more fine-grained insight into the particular level of visual processing where correspondence is established. If a masking paradigm has an influence on the correspondence solution, the correspondence is created after the masking has taken place, hence at the same or a later level of visual processing. If a masking paradigm has no influence, the correspondence was already completed at a level of processing before the masking took place. Such approaches could be used to pinpoint more exactly the processing levels of the visual system at which correspondence can be solved.

Another main finding within this thesis is that voluntary attention has been shown to be able to influence the correspondence process (Study 3: Stepper, Rolke, et al., 2020). Though this is the first evidence for the attentional pointer idea (Cavanagh, 1992; Cavanagh et al., 2010) within the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014), which states that attention could be a key mechanism for establishing correspondence, this finding comes with a possible limitation, as it could be that participants could have tracked the path of the element they had to attend to, leading to more motion percepts in the direction of the attended bias. This would also show that directing attention to specific elements can influence the correspondence solution, but not in the sense of an automatic correspondence mechanism like the attentional pointer idea. Such a voluntary tracking of the motion path would mean that participants must recognize the observed element's color within the presentation time of the Ternus display and consciously perceive its position across both frames in order to match its position over time to one of the possible motion percepts. This seems unlikely, as it would require additional effort that is not necessary to complete the task. However, it remains possible, so future research is needed to show that the influence of attention as a mechanism for resolving correspondence is automatic, possibly by examining whether similar attentional effects can be found when attention is directed automatically to the Ternus elements.

Questions also remain open regarding the suggested two-level correspondence process model. In this model, the final correspondence solution is determined by both image-based and object-based correspondence mechanisms. This model is in line with the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014), which also allows for the influence of image-based information on correspondence, including correspondence processes taking place at a low level of visual processing, as is proposed in image-based theories (e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; Kramer & Yantis, 1997). The distinction between situations in which each correspondence mechanism is used could be one of motion correspondence vs. object correspondence, as suggested by Moore et al. (2019). This leads to several open research questions: When is the current correspondence problem one of motion correspondence and when is it one of object correspondence? And how could the two correspondence processes interact? One approach to answer these questions could be to systematically investigate the influence of different display types. It would be important to understand how a display's complexity and motion type determine if it is motion correspondence or object correspondence as suggested by Moore et al. (2019). Further, within this work (see Chapter 7.3), is has been suggested that both the participants' task and the context information could affect the complexity, independent of the display type. The way these factors determine the correspondence type should be carefully investigated as well. In case of the task it has to be noted, that with increasing complexity of the display, the task also becomes more complex (simple apparent motion display: Motion yes/no?; ambiguous apparent motion display: What kind of motion?, e.g., Ternus display; continuous motion display: Name attributes of an object, e.g., object reviewing paradigm). To investigate the influence of the task independent of the display type, different tasks could be used with the same display. In this way, complexity can be gradually increased, for example shifting the focus from the motion percept toward objects. The Ternus display would be suitable for this purpose because the standard task, selecting the type of motion perceived, is of medium complexity. As a less complex task, one in which the focus is less on the objects, an indirect measure of motion perception could be used. For example, Boi, Öğmen, Krummenacher, Otto, and Herzog (2009) used a variation of the Ternus display which included a small dot on the central element positioned so that it rotates clockwise or counterclockwise if - and only if - group motion is perceived. The task of the participants was to indicate the motion direction of the small dot. For a more complex task, one could use the idea of the object reviewing paradigm (Kahneman et al., 1992) to construct a task in which participants have to name a target letter as fast as possible. In a preview of the three Ternus elements of the first frame, a letter could appear on the center element. After the letter disappears, the classic Ternus display is shown. Then, on the second Ternus frame, a target letter appears, and participants have to name it as fast as possible. An object specific preview benefit (OSPB) should be found if the target letter appears on the same object as in the preview, which in turn depends on the perceived motion. For example, a target letter on the left element should be named faster if it is the same as in the preview and, more importantly, if element motion was perceived, as the target letter would then have appeared on the same object. By using different tasks on the same display, the influence of the complexity of the task could be clarified.

Another aspect to investigate is the interplay between the image-based and object-based correspondence mechanisms: do they interact? If so, which information from which level is used to establish correspondence, and how is the information weighted for the final correspondence solution? One possibility could be that the different types of information, spatio-temporal, feature, and object-based information, are weighted depending on how reliable the information is and the level of visual processing at which it is available. Several studies' findings indicate that the available information is not always weighted in the same way for solving correspondence; in fact, the weighting could depend on how reliable information is. Hein and Moore (2012), for example, showed that with a stronger feature bias (see Figure 2.2B), meaning that more of the Ternus elements are compatible with either a group or an element motion percept, the ISI's influence becomes weaker. Hein and Schütz (2019) showed that, for solving correspondence, the visual system not only differentiates between the reliability of spatio-temporal and feature information but also between different types of feature information. They used a competitive Ternus display like the one used in Study 3 (Stepper, Rolke, et al., 2020), including an element and a group bias at the same time, though their competitive display consisted of different features, with a luminance and a color bias competing against each other. They showed that the feature bias with the higher contrast was more likely to be used for the correspondence solution and therefore seemed to be the more reliable (or salient) information for establishing correspondence. To investigate whether the reliability of information has an influence on the correspondence process, and through this which mechanism is used to solve correspondence, the dominance of image-based or object-based information could be manipulated. For example, the proportion of a particular ISI or a bias in trials can be manipulated to alter the reliability of the image-based information (ISI) relative to the object-based information (group or element bias). If the reliability has an influence on the correspondence solution, this would provide insight into the circumstances under which the different correspondence processes are more frequently used. This could contribute a better understanding of the behavior of the correspondence process in different correspondence problem situations.

7.5 Summary

This dissertation investigated the level of visual processing at which correspondence is established and found evidence in line with the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014) suggesting that correspondence can be solved at an intermediate level of visual processing using object-based information. This fills in a gap as research on this intermediate level (e.g., perceived surface: He & Nakayama, 1994; perceived lightness: Hein & Moore, 2014) was lacking relative to studies investigating higher-level influences (e.g., attention: Aydın et al., 2011; Scocchia et al., 2013; semantic and lexical information: Aydın et al., 2011; Chen & Zhou, 2011; Hsu et al., 2015; Yu, 2000). Study 1 and Study 2 both contributed to this by showing that the perceived size (Study 1: Illusion Ternus, Stepper, Moore, et al., 2020a), as well as the spatio-temporal history shown a priori (Study 2: History Ternus, Stepper, Moore, et al., 2020b) can influence the motion percept in the Ternus display and therefore the way correspondence is established. In addition, in Study 2 we found evidence that the object history stored at a mid-level object representation can influence the correspondence process independent of image-based information. Because correspondence can be established independently of such low-level information, this new evidence strongly contradicts image-based theories (motion-based theories: e.g., Adelson & Bergen, 1985; Kahneman et al., 1992; feature-based theories: e.g., Kramer & Yantis, 1997), which suggest that correspondence can be solved only using image-based information. Furthermore, with Study 3 (Stepper, Rolke, et al., 2020), we were able to show that voluntary attention directed to a specific element can influence the correspondence solution. This shows an influence of attention as a mechanism of establishing correspondence between elements, as is suggested by the object-based correspondence theory. Thus, attention could be used in the form of attentional pointers (Cavanagh, 1992; Cavanagh et al., 2010) tracking the most similar perceived elements over space and time based on their identity, creating a connection, a correspondence, between them. In line with the findings of this work and the idea of the object-based correspondence theory (Hein & Cavanagh, 2012; Hein & Moore, 2014) a two-level model of the correspondence process has been suggested, taking into account the potential distinction of two different correspondence processes (motion correspondence and object correspondence) as suggested by Moore et al. (2019).

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*(2), 284–299. https://doi.org/10.1364/JOSAA.2.000284

Alais, D., & Lorenceau, J. (2002). Perceptual grouping in the Ternus display: evidence for an "association field" in apparent motion. *Vision Research*, *42*(8), 1005–1016. https://doi.org/10.1016/S0042-6989(02)00021-4

Aydın, M., Herzog, M. H., & Öğmen, H. (2011). Attention modulates spatio-temporal grouping. *Vision Research*, *51*(4), 435–446. https://doi.org/10.1016/j.visres.2010.12.013

Berbaum, K., Lenel, J. C., & Rosenbaum, M. (1981). Dimensions of figural identity and apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(6), 1312–1317. https://doi.org/10.1037/0096-1523.7.6.1312

Boi, M., Öğmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision*, *9*(13), 1–11. https://doi.org/10.1167/9.13.5

Boring, E. G. (1930). A new ambiguous figure. *The American Journal of Psychology*, *42*(3), 444–445. https://doi.org/10.2307/1415447

Bowling, A., & Lovegrove, W. (1981). Two components of visible persistence: Effects of orientation and contrast. *Vision Research*, *21*(8), 1241–1251. https://doi.org/10.1016/0042-6989(81)90229-7

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Breitmeyer, B. G. (2015). Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing. *Consciousness and Cognition*, *35*, 234–250. https://doi.org/10.1016/j.concog.2015.01.012

Breitmeyer, B. G., Levi, D. M., & Harwerth, R. S. (1981). Flicker masking in spatial vision. *Vision Research*, *21*, 1377–1385. https://doi.org/10.1016/0042-6989(81)90243-1

Breitmeyer, B. G., & Ritter, A. (1986a). The role of visual pattern persistence in bistable stroboscopic motion. *Vision Research*, *26*(11), 1801–1806. https://doi.org/10.1016/0042-6989(86)90131-8

Breitmeyer, B. G., & Ritter, A. (1986b). Visual persistence and the effect of eccentric viewing, element size, and frame duration on bistable stroboscopic motion percepts. *Perception & Psychophysis*, *39*(4), 275–280. https://doi.org/10.3758/BF03204935

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*(6), 719–722. https://doi.org/10.1016/0042-6989(75)90290-4

Brown, J. M., Breitmeyer, B. G., Hale, R. G., & Plummer, R. W. (2018). Contrast sensitivity indicates processing level of visual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(10), 1557–1566. https://doi.org/10.1037/xhp0000554

Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology*, *4*(3), 121–138. https://doi.org/10.1080/17470215208416611

Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, *88*(2), 171–195. https://doi.org/10.1037/0033-295X.88.2.171

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*, 308–313. https://doi.org/10.1038/nn1194

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF : Support for signal enhancement. *Vision Research*, *40*, 1203–1215. https://doi.org/10.1016/S0042-6989(00)00024-9

Casco, C. (1990). The relationship between visual persistence and event perception in bistable motion display. *Perception*, *19*(4), 437–445. https://doi.org/10.1068/p190437

Cavanagh, P. (1992). Attention-based motion perception. *Science*, *257*(5076), 1563–1565. https://doi.org/10.1126/science.1523411

Cavanagh, P., Arguin, M., & von Grünau, M. (1989). Interattribute apparent motion. *Vision Research*, *29*(9), 1197–1204. https://doi.org/10.1016/0042-6989(89)90065-5

Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*(4), 147–153. https://doi.org/10.1016/j.tics.2010.01.007

Chen, L., & Zhou, X. (2011). Visual apparent motion can be modulated by task-irrelevant lexical information. *Attention, Perception, & Psychophysics*, *73*(4), 1010–1015. https://doi.org/10.3758/s13414-010-0083-5

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Coren, S., & Girgus, J. S. (1972). A comparison of five methods of illusion measurement. *Behavior Research Methods & Instrumentation*, *4*(5), 240–244. https://doi.org/10.3758/BF03210006

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617. https://doi.org/10.3758/BF03195489

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*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, *98*(4), 569–603. https://doi.org/10.1037/0033-295X.98.4.569

Dawson, M. R., Nevin-Meadows, N., & Wright, R. D. (1994). Polarity matching in the Ternus configuration. *Vision Research*, *34*(24), 3347–3359. https://doi.org/10.1016/0042-6989(94)90069-8

De Bono, E. (2005). *The six value medals: The essential tool for success in the 21th century*. Ebury Press.

Di Lollo, V. (1977). Temporal characteristics of iconic memory. *Nature*, *267*, 241–243. https://doi.org/10.1038/267241a0

Dodd, M. D., McAuley, T., & Pratt, J. (2005). An illusion of 3-D motion with the Ternus display. *Vision Research*, *45*(8), 969–973. https://doi.org/10.1016/j.visres.2004.10.011

Enns, J. T., Lleras, A., & Moore, C. M. (2010). Object updating: A force for perceptual continuity and scene stability in human vision. In R. Nijhawan & B. Khurana (Eds.), *Space and time in perception and action* (pp. 503–520). Cambridge University Press. https://doi.org/10.1017/CBO9780511750540.028

Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 840–853. https://doi.org/10.1037/0096-1523.32.4.840

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2012). Spatiotemporal priority as a fundamental principle of object persistence. In B. Hood & L. Santos (Eds.), *The Origins of Object Knowledge* (pp. 135–164). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199216895.003.0006

Fracasso, A., Caramazza, A., & Melcher, D. (2010). Continuous perception of motion and shape across saccadic eye movements. *Journal of Vision*, *10*(13), 1–17. https://doi.org/10.1167/10.13.14

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Green, M. (1986). What determines correspondance strenght in apparent motion? *Vision Research*, *26*(4), 599–607. https://doi.org/10.1016/0042-6989(86)90008-8

Green, M. (1989). Color correspondence in apparent motion. *Perception & Psychophysics*, *45*(1), 15–20. https://doi.org/10.3758/BF03208027

Gregory, R. L. (2009). *Seeing through illusions*. Oxford University Press.

He, Z. J., & Nakayama, K. (1994). Perceived surface shape not features determines correspondence strength in apparent motion. *Vision Research*, *34*(16), 2125–2135. https://doi.org/10.1016/0042-6989(94)90322-0

He, Z. J., & Ooi, T. L. (1999). Perceptual organization of apparent motion in the Ternus display. *Perception*, *28*(7), 877–892. https://doi.org/10.1068/p2941

Hein, E. (2017). The Ternus effect. In A. G. Shapiro & D. Todorović (Eds.), *The Oxford compendium of visual illusions* (pp. 685–691). Oxford University Press.

Hein, E., & Cavanagh, P. (2012). Motion correspondence in the Ternus display shows feature bias in spatiotopic coordinates. *Journal of Vision*, *12*(7), 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*(4), 975–988. https://doi.org/10.1037/a0028197

Hein, E., & Moore, C. M. (2014). Evidence for scene-based motion correspondence. *Attention, Perception, & Psychophysics*, *76*(3), 793–804. https://doi.org/10.3758/s13414-013-0616-9

Hein, E., & Schütz, A. C. (2019). Competition between color and luminance in motion correspondence. *Attention, Perception, & Psychophysics*, *81*(1), 310–322. https://doi.org/10.3758/s13414-018-1589-5

Hein, E., Stepper, M. Y., Hollingworth, A., & Moore, C. M. (2021). Visual working memory content influences correspondence processes. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(3), 331–343.

https://doi.org/10.1037/xhp0000890

Hock, H. S., Kelso, S. J. A., & Schöner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 63–80. https://doi.org/10.1037/0096-1523.19.1.63

Hollingworth, A., & Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition*, *113*(2), 150–166. https://doi.org/10.1016/j.cognition.2009.08.004

Hsu, P., Taylor, J. E. T., & Pratt, J. (2015). Frogs jump forward: Semantic knowledge influences the perception of element motion in the Ternus display. *Perception*, *44*(7), 779–789. https://doi.org/10.1177/0301006615596903

Johnson, S. P. (2018). Object Perception. *Oxford Research Encyclopedia of Psychology*, 1–17. https://doi.org/10.1093/acrefore/9780190236557.013.62

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–229. https://doi.org/10.3758/BF03214307

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219. https://doi.org/10.1016/0010-0285(92)90007-O

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Supplement), 1–16.

Kohler, A., Haddad, L., Singer, W., & Muckli, L. (2008). Deciding what to see: The role of intention and attention in the perception of apparent motion. *Vision Research*, *48*(8), 1096–1106. https://doi.org/10.1016/j.visres.2007.11.020

Kolers, P. A. (1972). *Aspects of Motion Perception*. Pergamon Press. https://doi.org/10.1016/C2013-0-05617-4

Kolers, P. A., & Pomerantz, J. R. (1971). Figural change in apparent motion. *Journal of Experimental Psychology*, *87*(1), 99–108. https://doi.org/10.1037/h0030156

Kornmeier, J., & Bach, M. (2012). Ambiguous figures – what happens in the brain when perception changes but not the stimulus. *Frontiers in Human Neuroscience*, *6*(51), 1–23. https://doi.org/10.3389/fnhum.2012.00051

Korte, A. (1915). Kinematoskopische Untersuchungen [Kinematoscopic examinations]. *Zeitschrift Für Psychologie*, *72*, 194–296.

Kramer, P., & Rudd, M. (1999). Visible persistence and form correspondence in Ternus apparent motion. *Perception & Psychophysics*, *61*(5), 952–962.

https://doi.org/10.3758/BF03206909

Kramer, P., & Yantis, S. (1997). Perceptual grouping in space and time: Evidence from the Temus display. *Perception & Psychophysics*, *59*(1), 87–99. https://doi.org/http://dx.doi.org/10.3758/BF03206851

Levi, D. M. (2008). Crowding - An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009

Lleras, A., & Moore, C. M. (2003). When the target becomes the mask: Using apparent motion to isolate the object-level component of object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 106–120. https://doi.org/10.1037/0096-1523.29.1.106

Michotte, A. (1963). *The perception of causality*. Methuen.

Michotte, A., Thinès, G., & Crabbè, G. (1991). Amodal completion of perceptual structures. In G. Thinès, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 140–167). Routledge.

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review*, *14*(6), 1199–1204. https://doi.org/10.3758/BF03193113

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Moore, C. M., & Enns, J. T. (2004). Object updating and the flash-lag effect. *Psychological Science*, *15*(12), 866–871. https://doi.org/10.1111/j.0956-7976.2004.00768.x

Moore, C. M., & Lleras, A. (2005). On the role of object representations in substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1171–1180. https://doi.org/10.1037/0096-1523.31.6.1171

Moore, C. M., Mordkoff, J. T., & Enns, J. T. (2007). The path of least persistence: Object status mediates visual updating. *Vision Research*, *47*(12), 1624–1630. https://doi.org/10.1016/j.visres.2007.01.030

Moore, C. M., Stephens, T., & Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin & Review*, *17*(5), 731–736. https://doi.org/10.3758/PBR.17.5.731

Moore, C. M., Stephens, T., & Hein, E. (2020). Object correspondence : Using perceived causality to infer how the visual system knows what went where. *Attention, Perception, & Psychophysics*, *82*, 181–192. https://doi.org/10.3758/s13414-019-

01763-y

Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. *Advances in Methods and Practices in Psychological Science*, *2*(3), 228–232. https://doi.org/10.1177/2515245919855053

Mordkoff, J. T., & Danek, R. H. (2011). Dividing attention between color and shape revisited: Redundant targets coactivate only when parts of the same perceptual object. *Attention, Perception, & Psychophysics*, *73*, 103–112. https://doi.org/10.3758/s13414-010-0025-2

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Navon, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human, Perception, and Performance*, *2*(1), 130–138. https://doi.org/10.1037/0096-1523.2.1.130

Necker, L. A. (1832). LXI. Observations on some remarkable optical phænomena seen in Switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *1*(5), 329–337. https://doi.org/10.1080/14786443208647909

Nishida, S., & Takeuchi, T. (1990). The effects of luminance on affinity of apparent motion. *Vision Research*, *30*(5), 709–721. https://doi.org/10.1016/0042-6989(90)90097-5

Pantle, A. J., & Petersik, T. J. (1980). Effects of spatial parameters on the perceptual organization of a bistable motion display. *Perception & Psychophysics*, *27*(4), 307–312. https://doi.org/10.3758/BF03206119

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Petersik, T. J., & Grassmuck, J. (1981). High fundamental spatial frequencies and edges have different perceptual consequences in the ‘group/end-to-end’ movement phenomenon. *Perception*, *10*(4), 375–382. https://doi.org/10.1068/p100375

Petersik, T. J., & Pantle, A. J. (1979). Factors controlling the competing sensations produced by a bistable stroboscopic motion display. *Vision Research*, *19*(2), 143–154.

https://doi.org/10.1016/0042-6989(79)90044-0

Petersik, T. J., & Rice, C. M. (2006). The evolution of explanations of a perceptual phenomenon: A case history using the Ternus effect. *Perception*, *35*(6), 807–821. https://doi.org/10.1068/p5522

Petersik, T. J., & Rice, C. M. (2008). Spatial correspondence and relation correspondence: Grouping factors that influence perception of the Ternus display. *Perception*, *37*(5), 725–739. https://doi.org/10.1068/p5900

Pikler, J. (1917). *Sinnesphysiologische Untersuchungen [Sensoryphysiological examinations]*. Barth.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145–3161. https://doi.org/10.1016/S0042-6989(01)00173-0

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Poth, C. H., Herwig, A., & Schneider, W. X. (2015). Breaking object correspondence across saccadic eye movements deteriorates object recognition. *Frontiers in Systems Neuroscience*, *9*, 176. https://doi.org/10.3389/fnsys.2015.00176

Poth, C. H., & Schneider, W. X. (2016). Breaking object correspondence across saccades impairs object recognition: The role of color and luminance. *Journal of Vision*, *16*(1), 1–12. https://doi.org/10.1167/16.11.1

Prinzmetal, W., Shimamura, A. P., & Mikolinski, M. (2001). The Ponzo illusion and the perception of orientation. *Perception & Psychophysics*, *63*(1), 99–114. https://doi.org/10.3758/BF03200506

R Development Core Team. (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org

Ramachandran, V. S., & Anstis, S. M. (1983a). Extrapolation of motion path in human visual perception. *Vision Research*, *23*(1), 83–85. https://doi.org/10.1016/0042-6989(83)90044-5

Ramachandran, V. S., & Anstis, S. M. (1983b). Perceptual organization in moving patterns. *Nature*, *304*, 529–531. https://doi.org/10.1038/304529a0

Ramachandran, V. S., Armel, C., Foster, C., & Stoddard, R. (1998). Object recognition can drive motion perception. *Nature*, *395*, 852–853. https://doi.org/10.1038/27573

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory Communication* (pp. 303–317). MIT Press. https://doi.org/10.7551/mitpress/9780262518420.001.0001

Richard, A. M., Luck, S. J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition*, *109*(1), 66–88. https://doi.org/10.1016/j.cognition.2008.07.004

Rock, I. (1983). *The logic of perception*. MIT Press.

Rolke, B., Dinkelbach, A., Hein, E., & Ulrich, R. (2008). Does attention impair temporal discrimination? Examining non-attentional accounts. *Psychological Research*, *72*, 49–60. https://doi.org/10.1007/s00426-006-0092-0

Rolke, B., Ulrich, R., & Bausenhart, K. M. (2006). Attention delays perceived stimulus offset. *Vision Research*, *46*, 2926–2933. https://doi.org/10.1016/j.visres.2006.02.022

Scholl, B. J. (2001). Spatiotemporal priority and object identity. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *20*(5), 359–371.

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind & Language*, *22*(5), 563–591. https://doi.org/10.1111/j.1468-0017.2007.00321.x

Scocchia, L., Valsecchi, M., Gegenfurtner, K. R., & Triesch, J. (2013). Visual working memory contents bias ambiguous structure from motion perception. *PLOS ONE*, *8*(3: e59217). https://doi.org/10.1371/journal.pone.0059217

Sekuler, A. B., & Bennett, P. J. (1996). Spatial phase differences can drive apparent motion. *Perception & Psychophysics*, *58*(2), 174–190. https://doi.org/10.3758/BF03211874

Sekuler, A. B., & Bennett, P. J. (2001). Generalized common fate: Grouping by common luminance changes. *Psychological Science*, *12*(6), 437–444. https://doi.org/10.1111/1467-9280.00382

Shechter, S., Hochsteinn, S., & Hillman, P. (1988). Shape similarity and distance disparity as apparent motion correspondence cues. *Vision Research*, *28*(9), 1013–1021. https://doi.org/10.1016/0042-6989(88)90078-8

Stepper, M. Y., Moore, C. M., Rolke, B., & Hein, E. (2020a). Illusory size determines the perception of ambiguous apparent motion. *Psychonomic Bulletin & Review*, *27*, 1230–1238. https://doi.org/10.3758/s13423-020-01786-9

Stepper, M. Y., Moore, C. M., Rolke, B., & Hein, E. (2020b). The role of object history in establishing object correspondence. *Attention, Perception, & Psychophysics*, *82*(3), 1038–1050. https://doi.org/10.3758/s13414-019-01923-0

Stepper, M. Y., Rolke, B., & Hein, E. (2020). How voluntary spatial attention influences feature biases in object correspondence. *Attention, Perception, & Psychophysics*, *82*(3), 1024–1037. https://doi.org/10.3758/s13414-019-01801-9

Suzuki, S., & Peterson, M. A. (2000). Multiplicative effects of intention on the perception of bistable apparent motion. *Psychological Science*, *11*(3), 202–209. https://doi.org/10.1111/1467-9280.00242

Tas, A. C., Moore, C. M., & Hollingworth, A. (2012). An object-mediated updating account of insensitivity to transsaccadic change. *Journal of Vision*, *12*(11), 1–13. https://doi.org/10.1167/12.11.18

Ternus, J. (1926). Experimentelle Untersuchungen über phänomenale Identität. [Experimental studies on phenomenal identity]. *Psychologische Forschung*, *7*, 81–136. https://doi.org/10.1007/BF02424350

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*(11), 502–511. https://doi.org/10.1016/j.tics.2006.09.003

Tse, P. U., & Cavanagh, P. (2000). Chinese and Americans see opposite apparent motions in a Chinese character. *Cognition*, *74*(3), B27–B32. https://doi.org/10.1016/S0010-0277(99)00065-7

Ullmann, S. (1979). *The interpretation of visual motion*. MIT Press.

van Santen, J. P. H., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, *2*(2), 300–321. https://doi.org/10.1364/JOSAA.2.000300

von Schiller, P. (1933). Stroboskopische Alternativversuche [Stroboscopic alternative experiments]. *Psychologische Forschung*, *17*, 179–214. https://doi.org/doi:10.1007/BF02411959

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wallace, J. M., & Scott-Samuel, N. E. (2007). Spatial versus temporal grouping in a modified Ternus display. *Vision Research*, *47*(17), 2353–2366. https://doi.org/10.1016/j.visres.2007.05.016

Watson, B. (1986). Apparent motion occurs only between similar spatial frequencies. *Vision Research*, *26*(10), 1727–1730. https://doi.org/10.1016/0042-6989(86)90059-3

Werkhoven, P., Sperling, G., & Chubb, C. (1993). The dimensionality of texture-defined motion: A single channel theory. *Vision Research*, *33*(4), 463–485. https://doi.org/10.1016/0042-6989(93)90253-S

Werkhoven, P., Sperling, G., & Chubb, C. (1994). Perception of apparent motion between dissimilar gratings: Spatiotemporal properties. *Vision Research*, *34*(20), 2741–2759. https://doi.org/10.1016/0042-6989(94)90230-5

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung [Experimental studies on seeing movement]. *Zeitschrift Für Psychologie Und Physiologie Der Sinnesorgane*, *61*(1), 161–265.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II [Principles of perceptual organization]. *Psychologische Forschung*, *4*(1), 301–350. https://doi.org/10.1007/BF00410640

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *The Journal of the American Medical Association*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Xu, Y., Suzuki, S., & Franconeri, S. L. (2013). Shifting selection may control apparent motion. *Psychological Science*, *24*(7), 1368–1370. https://doi.org/10.1177/0956797612471685

Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science*, *6*(3), 182–186. https://doi.org/10.1111/j.1467-9280.1995.tb00329.x

Yeshurun, Y. (2004). Isoluminant stimuli and red background attenuate the effects of transient spatial attention on temporal resolution. *Vision Research*, *44*(12), 1375–1387. https://doi.org/10.1016/j.visres.2003.12.016

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*(2), 293–306. https://doi.org/10.1016/S0042-6989(98)00114-X

Yeshurun, Y., & Levy, L. (2003). Transient spatial attention degrades temporal resolution. *Psychological Science*, *14*(3), 225–231. https://doi.org/10.1111/1467-9280.02436

Yu, K. (2000). Can semantic knowledge influence motion correspondence? *Perception*, *29*(6), 693–707. https://doi.org/10.1068/p3063