The Effect of Spatial Attention on Multistable Motion Perception via the Depth Mechanism

By

Hua-Chun Sun (孫華君)

B.S., National Chengchi University, 2007

THESIS

Submitted to the Department of Psychology of

National Chengchi University

in

Partial Fulfillment of the Requirements for the Degree of

Master of Science

Thesis Advisor: Shwu-Lih Huang, Ph.D. (黃淑麗博士)

July, 2010

Abstract

Many studies have found that fixating or directing spatial attention to different regions can bias the perception of the Necker cube, but whether this effect of spatial attention is due to attended areas perceived as being closer have yet to be examined. This issue was directly investigated in this study. The stimulus used was the diamond stimulus, containing four occluders and four moving lines that can be perceived as coherent or separate motions. The results of Experiment 1 show that coherent motion was perceived more often under the attending-to-occluders condition than under the attending-to-moving-lines condition, indicating that spatial attention can bias multistable perception. The results of Experiment 2 show that the mean probability of reporting lines behind occluders in small binocular disparities was significantly higher under the attending-to-occluders condition than under the attending-to-lines condition, indicating that spatial attention can make attended areas look slightly closer. The results of Experiments 3 and 4 show that the effect of spatial attention on biasing multistable perception was weakened when there were binocular or monocular depth cues to define the depth relationship between the occluders and the lines. These results are all consistent with the notion that spatial attention can bias multistable perception through affecting depth perception, making attended areas look closer.

Table of Contents

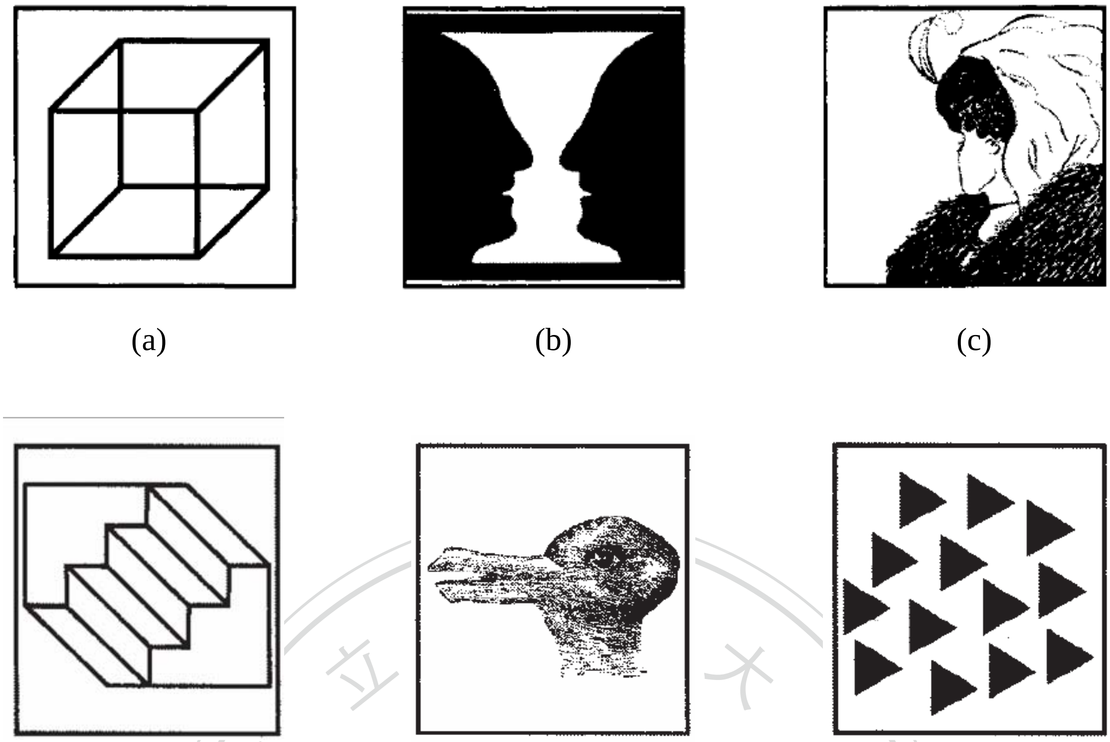
List of Figures

1. Introduction

1.1 Multistable Figures: Definitions, Properties, Examples, and Categories

A multistable figure is an ambiguous visual stimulus that can form at least two markedly different perceptual interpretations. When looking at a multistable figure, observers can perceive just one interpretation at one time. Some people have called this characteristic of multistable figures the "property of exclusivity" (Leopold & Logothetis, 1999). In other words, observers' perceptual systems seem unable to fixate on a single stable interpretation of a multistable figure. Instead, their perception fluctuates, alternating between different interpretations during a period of continuous viewing (Toppino & Long, 2005; Suzuki & Peterson, 2000).

The Necker cube is one of the famous examples of multistable figures, shown in Figure 1a, which can be seen as the front face down to the left or the front face up to the right. Other well-known examples include Rubin's faces/vase figure, which can be interpreted as a vase or two faces, and Boring's young girl/old woman figure, which can be interpreted as a young girl or an old woman. They are shown in Figures lb and lc, respectively.



(a) (b) (c)

(d) (e) (f)

Figure 1. This figure shows the following examples of multistable figures: (a) Necker cube; (b) Rubin’s face/vases figure; (c) Boring’s young girl/old woman figure; (d) Schroeder staircase; (e) duck/rabbit figure; and (f) equilateral triangles (Toppino & Long, 2005; Long & Toppino, 2004).

Garcia-Perez (1992) classified multistable figures into five types of percept-changing properties. The first type is the figure-ground reversible figure, such as the Rubin’s vase/face figure and the diamond stimulus, which will be introduced later. Reversals of this type of figure are related to figure-ground organization changes.

The second type is the perspective-reversible figure, such as the Necker cube and the Schroeder staircase (Figure 1d). Reversals of this type of figure are related to fluctuations in perspective (i.e., change in perceived depth).

The third type is the meaning-ambiguous figure, such as the young girl/old woman figure and the duck/rabbit figure (Figure 1e). The perceptual instability of this type of figure is associated with changes in meaning (Long & Toppino, 2004).

The fourth type is the orientation-reversible figure, such as the equilateral triangles in Figure 1f. The multistability of this type of figure involves the assignment of reference frames for the description of the shapes.

The last type is the stereokinesis moving pattern, such as a two-dimensional ellipse rotating in the frontal plane, which can be perceived as a disc oscillating in three-dimensional space or an elongated egg slanted in three-dimensional space and describing a circular trajectory in the frontal plane. It is the analysis of motion that causes the interpretation change (Garcia-Perez, 1992).

To simplify the classification of multistable figures, it is feasible to classify them in terms of whether the perception change is related to the alternation of depth perception. Based on this definition, all figure-ground reversible figures, perspective-reversible figures, and stereokinesis moving patterns belong to the depth-reversible figure category. Conversely, meaning-ambiguous figures and orientation-reversible figures do not belong to this category because their reversals have nothing to do with depth reversal. This study is aimed at the investigation of depth-reversible figures.

Since the multistable figure itself does not change at all, the perceptual interpretation of a multistable figure changing with time indicates that perceptual processing must be affected by some internal factors, which may be related to perceptual processing mechanisms (Pitts, Gavin, & Nerger, 2008). Therefore, many researchers use multistable figures to determine what factors affect our perceptual processing and how to affect the processing. In other words, the operation of perceptual mechanisms might be understood better by studying the nature of multistable figure perception (Toppino, 2003).

On the other hand, many researchers believe that ambiguity is the hallmark of retinal stimulus in almost all visual perception, so the visual system must solve the ambiguity to adapt to the environment effectively (Long & Toppino, 2004; Peterson & Gibson, 1991). The ambiguity comes from the inverse problem of the visual system. That is, the perception of three-dimensional environments is underdetermined by the two-dimensional image on the retina. In general, our visual system usually uses a heuristic process, producing the most likely interpretation to solve the inverse problem (Palmer, 1999). However, when viewing a multistable figure, the visual system produces more than one interpretation. Instead of fixating on a single, stable interpretation, the observers' perception alternates between different interpretations of a multistable figure during a period of continuous viewing, called instability. Some psychologists think that the perception of multistable figures is just like the perception of other stimuli in a normal environment; they are all ambiguous (Slotnick & Yantis, 2005; Peterson & Gibson, 1991), and they all involve the observers' perception of form, except that the ambiguity of multistable figures is available to consciousness while the ambiguity of other stimuli is not (Peterson & Gibson, 1991). In other words, the problem of ambiguity is solved in the initial period of the perception of normal stimuli, so the ambiguity does not enter the consciousness. However, in the perception of multistable figures, the ambiguity comes into the consciousness dramatically due to the instability of multistable figure perception. Nevertheless, the underlying processes are believed to be similar to the perception of normal stimuli (Long & Toppino, 2004; Peterson & Gibson, 1991; Leopold & Logothetis, 1999).

1.2 Theoretical Accounts of Multistable Figure Perception

Different studies have found different factors that can influence multistable figure perception, such as adaptation, fixation, and intention. Consequently, a variety of theories have been proposed to account for multistable figure perception. Some psychologists have classified these theories into two types of approaches according to the nature of the assumed underlying mechanisms—bottom-up theory and top-down theory (e.g., Kornmeier & Bach, 2005; Toppino & Long, 2005; Kornmeier, Hein, & Bach, 2009). The former emphasizes the influence of bottom-up processes and the latter emphasizes the influence of top-down processes that influence multistable figure perception. These two types of approaches have been supported by respective evidence.

Bottom-up theory

In bottom-up theory, the proposed explanation of multistability is in terms of relatively passive, automatic, sensory, stimulus-driven, and bottom-up mechanisms. Figure reversals are assumed to reflect the fatigue and recovery of neural channels that underlie the alternate perceptions of a multistable figure (Toppino & Long, 2005; Kornmeier & Bach, 2005; Toppino & Long, 1987; Kornmeier, Hein, & Bach, 2009). When viewing a multistable figure for a long time, one set of channels underlying one percept of a multistable figure is adapted or fatigued. At the same time, the second set of channels underlying the other percept becomes more and more dominant, causing a reversal. Afterwards, as the second set of channels becomes fatigued with extended viewing and the first set of channels recovers from fatigue, the percept will reverse again accordingly.

There is a great deal of evidence supporting bottom-up theory. A reliable finding of multistable figure literature that has been replicated many times is as follows: with prolonged viewing of a multistable figure, reversals become more frequent. Bottom-up theorists attribute this phenomenon to each channel fatiguing more quickly with increased viewing time. Therefore, reversals take place more quickly, and the channels do not fully recover before the next reversal begins (Toppino & Long, 2005).

Toppino and Long’s (1987) study supported bottom-up theory more strongly. The multistable figure they used was a rotating Necker cube. When viewing this stimulus, the direction of rotation can be perceived as to the left or to the right. They divided one trial into a two-minute adaptation period and a two-minute test period. They found that if the rotating Necker cube’s location or size presented in the adaptation period was the same as in the test period, the reversal rate was high during the test period, reflecting neural-channel fatigue. On the contrary, the reversal rate was low (as a no-adaptation control condition) when the rotating cube’s location or size presented in the two periods was different. This is because the different location or size of stimulus in the two periods is related to the processes of different neural channels. Thus, neural-channel fatigue did not occur. A more striking piece of evidence was as follows: when the stimulus presented in the test period changed into two side-by-side rotating cubes, the reversing rate of the two cubes was different. The cube at the same location in the adaptation period reversed more rapidly than the cube at a different location in the adaptation period. These localization characteristics of fatigue are consistent with bottom-up theory and incompatible with top-down theory.

Another study supporting bottom-up theory is Kornmeier and Bach's (2005) (2005). They recorded event-related potentials (ERPs) when presenting Necker cubes intermittently with a temporal regime (800ms on, 400ms off, shown in Figure 2). Participants have to report whether their perception reversed or not by comparing any given stimulus with the preceding one. According to their report, the ERP data can be divided into two conditions for comparison—perceptual reversal and non-reversal. They found that the earliest significant ERP difference associated with perceptual reversal peaks around 120 ms after stimulus onset, and it is most prominent at the occipital cortex. This result shows that perceptual reversals can be initiated during very early visual processing stages, implicating bottom-up processing.

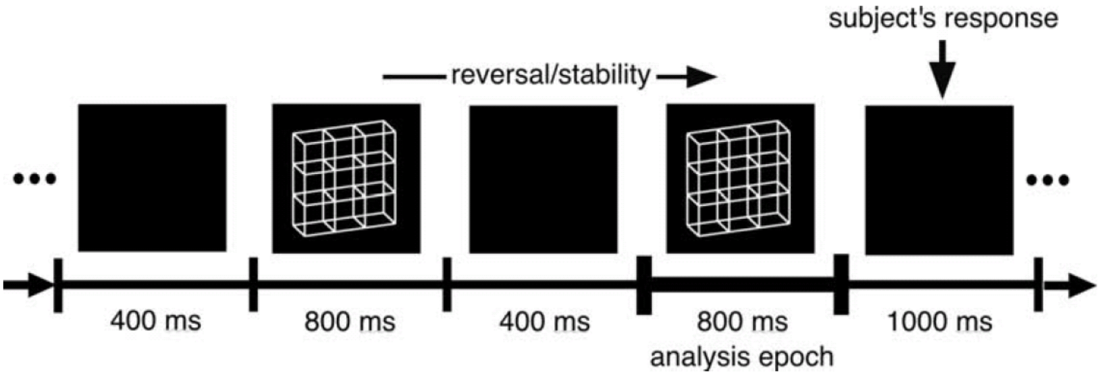


Figure 2. The experimental paradigm of Kornmeier and Bach (2005).

However, there still are some findings that cannot be explained well by bottom-up theory. For example, Long et al. (1983) separated the viewing sessions of multistable figures across four weeks and found that the reversal rate increased across the four weeks. This result is difficult to explain using the bottom-up fatigue and recovery mechanism because neural-channel fatigue seems unlikely to last for weeks. Instead, this can be explained by some higher-order processes such as learning (Toppino & Long, 2005). Many researchers have used this viewpoint to investigate the mechanism underlying multistable figure perception, and they are called top-down theorists.

Top-down theory

According to top-down theory, multistability is explained in terms of higher-level cognitive processes that exert a top-down influence on the interpretation of ambiguous retinal stimulation. Top-down theory conceptualizes perception as requiring problem-solving or decisional processes that are restrained by limited attentional resources and influenced by experience (Long & Toppino, 2004). They attribute reversals to fluctuations of attention (e.g., Kawabata & Mori, 1992), alternating decision processes (e.g., Rock, 1975), or successive testing of perceptual hypotheses (e.g., Gregory, 1974). In summary, top-down theory attributes the reversal of multistable figure perception and the property of exclusivity to higher-order cognitive processes, such as learning, expectation, experience, the viewer's intentions, cognitive state, and rival demands on attentional resources (Kornmeier & Bach, 2005; Long & Toppino, 2004).

Contrary to bottom-up theorists, top-down theorists attribute the increasing rate with increased viewing time to learning. That is, observers become more facile in accessing the internal representation of each representation of a multistable figure with practice (Toppino & Long, 2005). Furthermore, many researchers have found that observers can exert intentional control (or volitional control) during multistable figure perception (Suzuki & Peterson, 2000; Kornmeier, Hein, & Bach, 2009). For instance, observers can hold a particular percept of the figure or switch between each percept as rapidly as possible (Peterson & Gibson, 1991; Struber & Stadler, 1999). This kind of intentional control is assumed to be produced by top-down priming or activation of the desired representation (Toppino, 2003; Toppino & Long, 2005; Suzuki & Peterson, 2000).

Shifting in the locus of spatial attention is also a top-down influence on multistable figure perception (Tsal & Kolbet, 1985). Slotnick and Yantis (2005) used event-related functional MRI to compare neural activities during voluntary shifts of spatial attention and voluntary shifts of the perception of the Necker cube. They found this two brain processing is associated with common activity in the posterior parietal cortex, implying that voluntary shifts of spatial attention are usually accompanied by shifts of the perception. This result implies that spatial attention, a top-down influence, can mediate perceptual multistability.

Knowledge of multistability is also an important factor that can influence multistable figure perception. For example, if the observer does not know that the figure has more than one interpretation and perceives it just in one way, few reversals will be reported (Long & Toppino, 2004). This evidence supports top-down theory.

A hybrid model

The best interpretation of the findings of the two approaches is that both bottom-up and top-down processes have influence on multistable figure perception. One of the most striking pieces of evidence of this contention is the effects of pre-exposure to an unambiguous figure before the multistable figure presentation (Toppino & Long, 2005). Long et al. (1992) manipulated the duration of pre-exposure to an unambiguous figure (such as a rotating cube) from zero to 150 seconds. Then a 30-second test period followed immediately, and observers had to report their perception of the ambiguous (multistable) figure (such as a rotating Necker cube) that corresponded to the pre-exposure stimulus. They found the same-bias effect (i.e., initial perception of the ambiguous figure tended toward the same configuration as the prior unambiguous figure) in short pre-exposure duration trials (less than 5 seconds) and a reverse-bias effect (i.e., initial perception tended toward the opposite configuration as the prior unambiguous figure) in long pre-exposure duration trials (more than two minutes). This is because the short pre-exposure duration of the unambiguous figure induces a top-down priming effect toward the same representation. Thus, the followed ambiguous figure is perceived as a bias toward the same representation. On the contrary, a long pre-exposure duration of an unambiguous figure induces a bottom-up adaptation and fatigue effect. That is, the neural channels that were not adapted by the unambiguous figure presented previously became more dominant than the adapted channels, causing the followed ambiguous figure to be perceived the other way (Toppino & Long, 2005).

Consequently, many researchers (e.g., Hochberg, 1968; Long et al., 1992;

Palmer & Bucher, 1981; Toppino & Long, 1987) have adopted the opinion that multistable figure perception is influenced by both bottom-up and top-down processes. This opinion is called hybrid theory. In the hybrid theoretical framework, one of the most important issues is clarifying how the bottom-up and top-down processes interrelate, contributing to the multistable figure perception together (Toppino & Long, 2005; Pitts, Gavin, & Nerger, 2008; Kornmeier, Hein, & Bach, 2009).

For example, Toppino (2003) used the Necker cube as a stimulus to investigate whether intentional control (top-down) over perception could be explained in terms of intentionally selecting appropriate focal features within the stimulus for primary processing, which is called “focal-feature processing” (bottom-up). He mainly manipulated two independent variables—observers’ intention (to hold down-to-the-left or up-to-the-right orientation of the cube) and fixation (fixating at the bottom-left or top-right corner of the cube) when viewing the Necker cube. He found that these two factors have additive effects on perception. Even when the cube was too small to exhibit the fixation effect, intention still can bias perception of the cube. Another experiment by Meng and Tong (2004) also used the Necker cube as a stimulus and saw similar findings, with the manipulation of intention called “selective attention.” Hence, it is likely that intentional control is independent of focal-feature processing—perhaps through top-down activation or priming of perceptual representations—as previously described. The effect of focal-feature processing will be discussed below.

1.3 Effect of Fixation and Spatial Attention

Effect of fixation

Many studies have found that fixating on different locations within a Necker cube tends to favor one or the other perceptual response. There is a tendency to perceive the vertex of the Necker cube that is nearest the visual fixation point as the frontal face (Inui, Tanaka, Okada, Nishizawa, Katayama, & Konishi, 2000; Kawabata, Yamagami, Noakl, 1978). For example, fixating at its right-up corner tends to favor the perception of the cube as front face down to the left, and fixating at the left-down corner tends to favor the front face up to the right perception (Kawabata, Yamagami, & Noakl, 1978; Long & Toppino, 2004), as showed in Figure 3. This is because “the fixated area tends to be seen as being relatively closer to the observer” (p. 1287) (Toppino, 2003). In Necker’s view, “the foveated portion of the figure was naturally supposed [by the observer] to be nearer and foremost” (p. 337) (Necker, 1832, 1964; Long & Toppino, 2004). Kawabata, Yamagami, and Noakl (1978) thought that the degree of clarity caused by the difference of visual resolution and attention is an important cue for the depth perception of the Necker cube. That is, areas near the visual fixation point are seen clearly, and the clear areas are seen as being in the front.

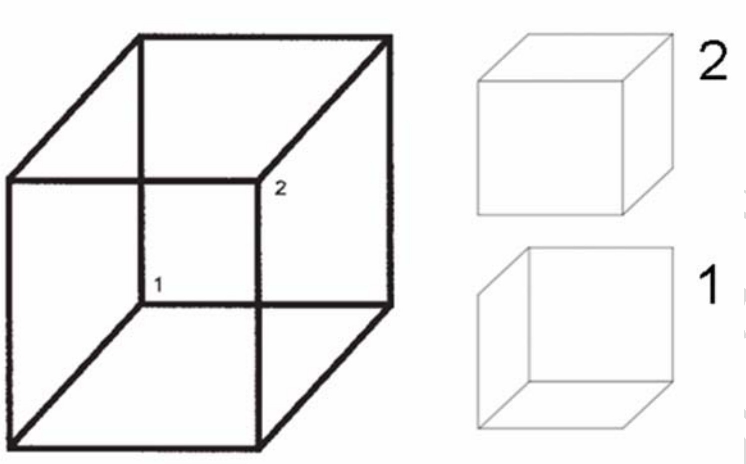


Figure 3. The two corners of the Necker cube and their correspondent perception when attending or fixating on them.

Garcia-Perez (1992) also assumed that perceptions of multistable figures are biased depending on where observer fixates in a figure. The explanation is that because fixations near a focal area for one of the percepts will bring out fine-detail (high spatial-frequency) information relevant to that percept, it is easier to build that perceptual interpretation. This opinion can also be explained by the notion that fixation areas are seen as being closer because the clearer the image of an object is, the nearer this object is perceived to be.

Toppino (2003) summarized these effects of fixation on multistable figure perception described above and proposed the focal-feature hypothesis. Its main assumption is that “different focal regions or areas within a multistable figure favor different global interpretations” (p. 1286). Furthermore, “which interpretation is perceived is assumed to depend on which focal area is selected for enhanced processing, regardless of whether that selectivity reflects the locus of fixation or the focus of attention alone” (p. 1286).

We should notice that the locus of fixation is not the same as the focus of attention. Even though these two coincide with each other most of the time, attention can be shifted independently of eye movements (Posner, 1980; Palmer, 1999; Tsal & Kolbet, 1985). For example, we can fixate on one place while attending to another place; this is also called covert attentional shift (Suzuki & Peterson, 2000; Theeuwes, 1992; Leopold & Logothetis, 1999). Therefore, the effect of fixation and the effect of attention should be differentiated.

In addition, there is another issue in the focal-feature hypothesis that needs to be clarified: what is the “enhanced processing” for the selected focal area that biases the interpretation of multistable figures? Toppino (2003) did not clearly point out why “different focal areas within a multistable figure favor different global interpretations” (p. 1286). One of the most likely reasons is that focal areas tend to be seen as closer or nearer to the observer, as Toppino and Necker (1832, 1964) assumed.

Effect of spatial attention

Kawabata (1986, 1987) directed observers' attention to an angle at a vertex of a briefly presented (500 ms) Necker cube. He found that the attended angle was perceived as the front part of the cube and other parts were interpreted to match this interpretation. However, this study had two main disadvantages with respect to directing observers' attention. First, in every trial, the vertex that needed attending was always presented at the fixation point. This design can allow observers to attend to the vertex easily and spontaneously, but the effect of the attention would be confounded with the effect of fixation. Second, the author used two heavy lines to indicate the angle (one of the three angles of the attended vertex) to which observers should give attention. However, it may change the nature of the stimulus itself. Xu and Franconeri's (2010) study had similar findings and challenges. They found that people are more likely perceiving the cued side of the Necker cube as the closer side. The problem is that Xu and Franconeri used bright lines on the corners of the cube as a cue to direct people's exogenous attention. However, showing the cue may affect perception. In summary, the effect of attention found in these studies may be confounded with the effect of fixation or the effect of the cue itself.

Peterson and Gibson (1991) used the Necker cube to investigate whether people can direct spatial attention into different sub-regions of an object while ignoring other sub-regions. They manipulated observers' fixation location, spatial attention location (on the biased region or unbiased region, as showed in Figure 4), and intention (to hold either the horizontal or the vertical line in front at the attended intersection) and recorded their perception (whether the horizontal or the vertical line look forward) in 30-second trials. They found that the effect of the bias region shows only when the bias region is attended regardless of fixation location. They propose that the processing of the stimulus is facilitated in the attended location, and it is attenuated in the unattended location. Thus, when attention is directed to the biased region, the processing of the depth information—occlusion and shading—is facilitated, and it strongly affects the perception of the stimulus. On the other hand, when attention is directed to the unbiased region, the processing of the depth information of the biased region is attenuated. At this time, intention can exert its effect through top-down activation of the desired representation. However, due to spatial attention locations mentioned here that are different from the attended corners mentioned previously (such as in Figure 3), this study provides no evidence about whether attended locations look closer.

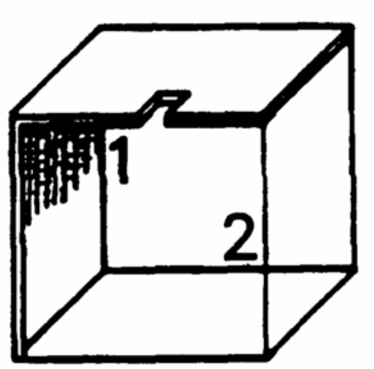


Figure 4. One of the partially biased Necker cubes used by Peterson Gibson (1991). 1 is the biased region on which occlusion and shading define the horizontal line as being in front of the vertical line. 2 is the unbiased region.

Tsal and Kolbet (1985) used meaning-ambiguous figures, such as the duck/rabbit figure, to investigate whether the perception of an ambiguous figure results from focusing attention on a focal area that contains features significant for this percept. In their Experiment 1 (see Figure 5, left), observers were instructed to maintain a given interpretation through the block. After the meaning-ambiguous figure was presented briefly, a letter may have been presented at one of the two focal areas. Observers had to respond if they saw a letter only when they maintained the given interpretation successfully. They found that the letter detection was faster when the letter appeared in the focal area of the perceived interpretation versus the focal area of the alternative one. In their Experiment 2 (see Figure 5, right), the observers' attention was directed by a letter shortly before the presentation of the figure. They found that the perception of the figure was more frequently associated with the letter-presented focal area than with the alternative focal area. They suggested that maintaining different interpretations of the same ambiguous figure is mediated by focusing attention on different focal areas of the figure. Because different features of the figure support different overall interpretations (for example, the right part of the duck/rabbit figure in Figure 5 looks more like a beak than a pair of ears), attending to a feature may cause the correspondent interpretation to become dominant (Tsal & Kolbet, 1985). Nevertheless, this explanation may be appropriate only for meaning-ambiguous figures. For depth-reversible figures such as the Necker cube, the right-up corner looks just like the left-down corner; they have similar features (including a horizontal line and a vertical line), so it is difficult for them to support different overall interpretations. There may be other reasons that cause spatial attention to affect depth-reversible figure perception. One of the most likely reasons is that spatial attention can affect depth perception, making the attended areas look closer, like the effect of fixation mentioned above. Thus, in the current study, experiments are designed to test this notion. Before discussing this, some relative studies need to be reviewed in order to evaluate the possibility that attention can affect depth perception, making the attended areas look closer.

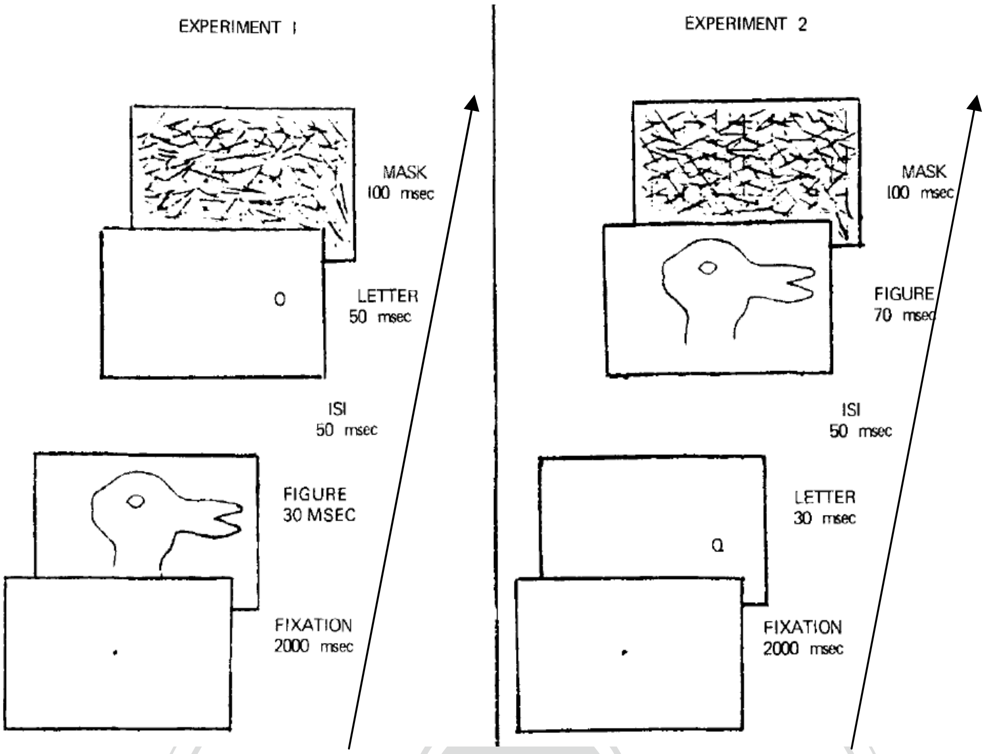


Figure 5. Procedures of the two experiments in Tsal and Kolbet's (1985) study.

1.4 Aspects of Spatial Attention that may Affect Depth Perception

The evidence indicating that spatial attention may affect depth perception can be classified into two categories. The first category comprises the evidence that attention can modulate very early visual processes. For example, Tootell, Hadjikhani, Hall, Marrett, Vanduffel, Vaughan et al. (1998) used functional magnetic resonance imaging (fMRI) to label cortical activity due to visual spatial attention. They found that directing attention to a stimulus location results in increased activity not only in visual areas V4, V3, V3a, and V2 (robust attentional modulations) but also in V1 (smaller attentional modulations). Similarly, McAdams and Maunsell (1999) trained two monkeys attending to a stimulus inside or outside the receptive fields of recorded neurons in V4 and V1. They found that attention enhanced not only the responses of V4 neurons (median 26 % increase) but also the selectivity of V1 neurons (median 8 % increase).

Event-related potential (ERPs) studies have found that modulations of spatial attention first occur in the extrastriate cortex at a latency of 80-100ms after stimulus onset. In addition, a delayed attentional modulation of V1 begins around 130-160ms, indicating a mechanism that delays feedback from higher areas in modulating neural activity in V1 (Martinez et al., 2001; Noesselt et al., 2002; Di Russo, Martinez, and Hillyard, 2003). Lamme and Spekreijse (2000) proposed that the delayed feedback activity in V1 may improve figure-ground segregation and enhance the stimulus salience in attended areas (Hopfinger, Luck, & Hillyard, 2004).

Vecera, Flevaris, and Filapek (2003, 2004) directed participants’ attention by placing an unpredictable peripheral cue inside a figure-ground display so that exogenous attention was directed to one of the possible figural regions. Participants then had to indicate which of two probe stimuli matched one of the regions that appeared in the figure-ground display. They found that exogenous attention can influence figure-ground assignment: participants are faster to match the cued region from memory than they are the uncued region. On the other hand, in general, participants are faster to match the convex region than the concave region from memory. Convexity is a Gestalt figure-ground cue, so a convex region is more likely to be seen as figure than the concave region. However, this difference of reaction time (RT) between the convex and concave regions is reduced when attention is directed to the concave region. These results suggest that figure-ground processes are not entirely completed prior to the operation of spatial attention.

To sum up the evidence in the first category, attention can modulate very early visual processes, including neural activity in V1 and figure-ground processes. Thus, it is possible for attention to modulate follow-up visual processes, such as depth perception.

The second category comprises the evidence that spatial attention can alter subjective perception of a stimulus appearance. One example is that both exogenous/involuntary/transient spatial attention and endogenous/voluntary/sustained spatial attention can increase perceived stimulus contrast, causing the apparent contrast of cued stimulus to be higher than that of uncued stimulus (Carrasco, 2006; Carrasco, Ling, & Read, 2004; Carrasco, Fuller, & Ling, 2008; Liu, Abrams, & Carrasco, 2009; Read & Carrasco, 2003; Treue, 2004; Ling, Carrasco, Lipson, Roche, Little, & Jones, 2007). Another example is that both exogenous and endogenous spatial attention can increase perceived spatial resolution, causing visual acuity to increase at cued locations in the Landolt gap resolution task (Carrasco, Williams, & Yeshurun, 2002; Montagna, Pestilli, & Carrasco, 2009; Yeshurun & Carrasco, 1999). Still another example is that exogenous spatial attention can increase perceived spatial frequency, causing spatial frequency in peripheral cued locations to be higher than in neutral cue conditions (Gobell & Carrasco, 2005). Exogenous spatial attention can also increase the perceived gap size of a Landolt square, the size of the moving random dot patterns (Gobell & Carrasco, 2005; Anton-Erxleben, Henrich, & Treue, 2007), as well as increasing color saturation (Fuller & Carrasco, 2006; Fuller, Ling, & Carrasco, 2004).

To sum up the evidence in the second category, attention can increase perceived stimulus contrast, spatial resolution, spatial frequency, and size. It is worth noting that these properties are relative to depth perception. For instance, if a stimulus is nearer to us, the stimulus will be more salient and acute (due to increased contrast, spatial resolution, and spatial frequency), and the size will be larger. Hence, it is possible for spatial attention to influence depth perception, making attended areas look closer. In order to test this hypothesis, the property of the selected stimulus must be able to reflect the effect of depth perception on it. Therefore, the diamond stimulus is chosen.

1.5 Experimental Stimulus

The multistable stimulus used in this study is the diamond stimulus, introduced by Lorenceau and Shiffrar (1992). The diamond stimulus contains a diamond outline moved in a circular trajectory with the four corners hidden by four occluders, as shown in Figure 6. When looking at this stimulus, observers can perceive the single coherent motion behind the occluders (Figure 6a) or the separate motions of the line segments (i.e., each of the four line segments moves sinusoidally in the direction shown in Figure 6b) (McDermott & Adelson, 2001).

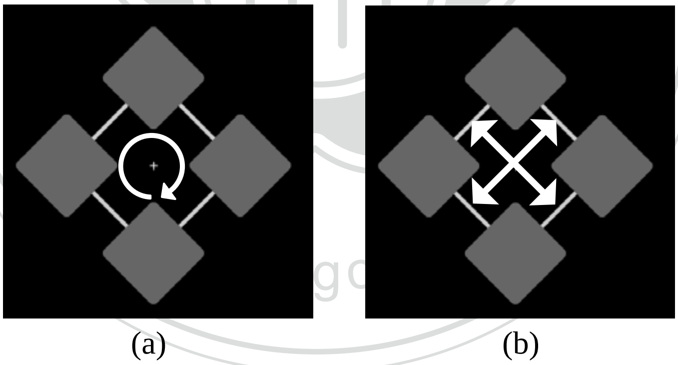
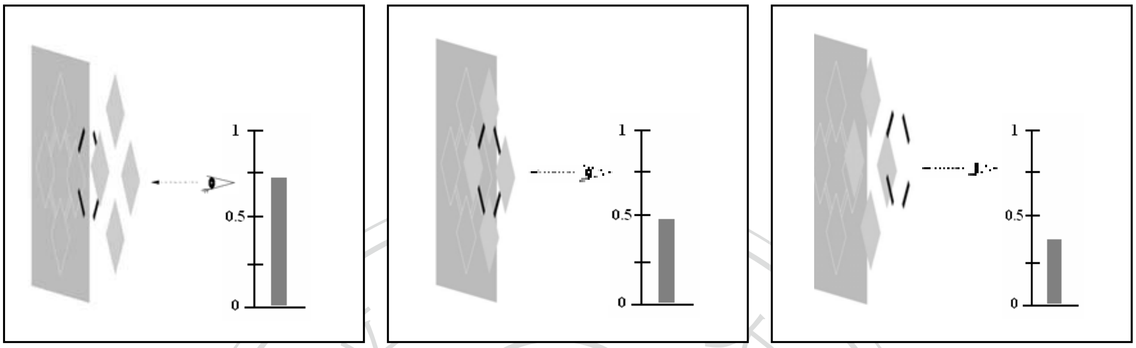

(a) (b)

Figure 6. The diamond stimulus, which can be perceived as (a) coherent motion or (b) separate motion.

The diamond stimulus is appropriate for this study for two main reasons. First, under Garcia-Perez's classification framework of multistable figures, the diamond stimulus is a figure-ground reversible figure. That is, when the four moving lines are seen as the ground, the moving lines are seen as connecting to each other, moving coherently behind the occluders. However, when they are seen as a figure, they are *not* seen as being behind the occluders and connected to each other. Thus, separate motion will be perceived. Most of the previous studies that found the effect of spatial attention or fixation used the Necker cube as the stimulus, which belongs to perspective-reversible figure category. Therefore, it is unclear whether spatial attention can bias the perception of figure-ground reversible figures. However, if spatial attention is exerted through the depth mechanism, the effect of attention found in perspective-reversible figures should also present in figure-ground reversible figures because they all belong to the depth-reversible figure category.

The second reason for using the diamond stimulus is that the depth relationship between occluders and moving lines can be manipulated to investigate the depth effect of spatial attention. Kuo and Huang (2005) have found that the depth relationship between the occluders and moving lines can influence the perception of its motion. For example, if the occluders are in front of the moving lines, the percentage of coherent motion perception in one minute is higher than that under the conditions of occluders being perceived as behind or on the same plane as the moving lines. Conversely, if the moving lines are in front of the occluders, the percentage of coherent motion perception is lower than that under the conditions of occluders being perceived as behind or on the same plane as the moving lines. The results are shown in Figure 7.



(a) (b) (c)

Figure 7. Percentage of coherent motion perception in one-minute trials under different kinds of binocular disparity of the occluders and moving lines. Coherent motion perception was greater when the occluders were in front of the moving lines (a) and less when the moving lines were more in front of the occluders (c). When the occluders and the moving lines were on the same plane, Coherent motion perception fell between the other two conditions (b) (Kuo & Huang, 2005).

Since the motion perception of the diamond stimulus can be affected by the depth relationship between the occluders and the moving lines, it is suitable to be used to investigate whether spatial attention can bias multistable motion perception by making the attended areas look closer in depth. If the answer is yes, motion perception should be different under the two attention conditions. That is, when observers attend to the four occluders, coherent motion perception should be reported more than the attending-to-lines condition. This is because the attended occluders are seen as closer and the moving lines are seen as connected to each other behind the occluders. In contrast, when observers attended to the four moving lines, separate motion perception should be reported more than under the attending-to-occluders condition because the moving lines are seen as closer, not occluded, and moving separately.

1.6 Purpose, Questions, and Hypothesis of the Current Study

Based on previous studies, directing fixation or spatial attention to different regions of a Necker cube can bias the perception of it. Many researchers believe this is because fixated or attended regions look closer. However, no study has examined this assumption directly. In this study, the multistable stimulus used to examine this assumption is the diamond stimulus, which belongs to the depth-reversible figure category, which includes the Necker cube. In addition, the mechanisms of fixation and spatial attention are different and should be differentiated.

In the research concerned with attention, few studies have mentioned whether spatial attention can affect depth perception. Many studies have found that spatial attention can influence many properties of stimulus perception, which may relate to depth perception. Thus, it is possible for spatial attention to affect multistable perception through the depth mechanism.

The purpose of this study is to investigate the effect of spatial attention on depth-reversible figures perception and the possible underlying mechanism of this effect. The hypothesis is that spatial attention can bias the multistable motion perception of the diamond stimulus by making the attended areas look closer in depth, as shown in Figure 8.

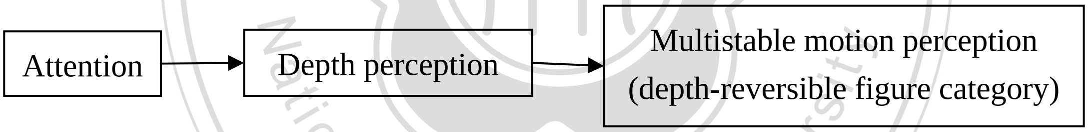


Figure 8. The diagram of the hypothesis. Attention affects multistable motion perception by making the attended areas look closer in depth.

In order to test this hypothesis, four main questions need to be considered serially. The first is whether spatial attention can bias multistable motion perception. If this is supported, we can further check to determine whether the direction of bias coincides with the prediction that the attended areas look closer. These two questions will be verified in Experiment 1.

However, even if the results of Experiment 1 are consistent with the prediction of our hypothesis, this will not directly confirm that spatial attention is biasing multistable motion perception by affecting depth perception. Accordingly, in Experiment 2, this question will be verified directly. If the results of Experiment 2 support the notion that attention can affect depth perception to make attended areas look closer, it is more likely that the effect of spatial attention on multistable motion perception found in Experiment 1 occurs through the mechanism of making the attended areas look closer in depth.

Experiments 3 and 4 are designed to test the hypothesis from another perspective. Binocular and monocular depth cues are manipulated in Experiments 3 and 4, respectively. The question to be clarified is whether the effect of spatial attention on biasing multistable motion perception will decrease or disappear when depth cues of the stimulus are well-defined. The reasoning is that if attention can bias multistable motion perception by making the attended areas look closer in depth, then, if there are clear, well-defined depth cues (regardless of binocular or monocular cues) to define the depth relationship between occluders and moving lines, the effect of attention should be eliminated or overridden. In other words, the effect of attention on multistable motion perception will decrease or disappear.

In addition, we should note that intention is also a key factor that can influence multistable figure perception greatly, as mentioned previously. For example, observers can hold or switch to a particular percept of a multistable figure (Toppino & Long, 2005). Suzuki and Peterson (2000) suggested that although observers can view a multistable display passively without actively intending to see one or the other alternative perception, uncontrolled intentions could seriously confound experimental results. In particular, observers' expectations or supposition of the experimental purpose may affect their intention. Therefore, it would be advantageous to control the observers' intention, or the effect of attention may not be revealed. Therefore, the variable of intention will be controlled in all experiments in the current study.

2. Experiment 1: Can Spatial Attention Bias Multistable Motion Perception?

Experiment 1 is designed to examine whether spatial attention can bias multistable motion perception. Participants' attention was manipulated by requesting they attend to the four occluders or the four moving lines while simultaneously remaining fixated on the center cross. Using this method, the effect of fixation can be eliminated and the pure effect of spatial attention can be revealed. If attention can bias multistable perception by making the attended areas look closer, then, when participants attend to the four occluders, the occluders will look closer, as if they are in front of the moving lines. Moreover, previous research has found that when the occluders are in front of the moving lines, coherent motion perception increases (Kuo & Huang, 2005). Therefore, coherent motion perception should increase under the attending-to-occluders condition. Similarly, when participants attend to the four moving lines, the four moving lines will look closer, as if they are in front of the occluders. Previous research has found that when the moving lines are in front of the occluders, coherent motion perception decreases. Therefore, coherent motion perception should decrease under the attending-to-moving-lines condition.

In Experiment 1a, participants' intention was also manipulated as an independent variable for controlling purposes by requesting that participants attempt to hold the perception of coherent or separate motion. The prediction is that participants can exert intention control to hold coherent motion or separate motion, so the percentage of time perceiving coherent motion should be higher in “hold coherent” condition than “hold separate” condition.

On the other hand, investigating the mechanisms between attention and intention is also explores in Experiment1a. Toppino (2003) found that observers’ intention (to hold the down-to-the-left or up-to-the-right orientation of the Necker cube) and fixation (fixating at the bottom-left or top-right corner of the cube) are two independent mechanisms that have additive effects on perception. Hence, if the mechanism of attention is similar to fixation, which works with intention independently, then there should be no interaction effect between attention and intention. Moreover, in this study’s further experiments, the control of intention can remain constant (for example, all hold coherent motion) if the effect of attention under the two intention condition is similar.

2.1 Experiment 1a: The Relationship between Spatial Attention and Intention: Dependent or Independent Mechanisms?

2.1.1 Methods

Participants

Twelve participants who had normal or adjusted to normal vision were recruited in this experiment. Their age range was from 18 to 24 years. After finishing this experiment, they received a 0.5 bonus point in the “Methods of Psychological Experiment” course or 50 NT dollars.

Design

The experiment is a 2 (intention) × 2 (attention) mixed factorial design. The first factor is a between-participant factor and the second is a within-participant factor. Participants’ attention was manipulated by instructing them to attend to the four occluders in one block and the four moving lines in the other block. The order of the two attention conditions was counterbalanced across participants. At the same time, their intention was also manipulated through the experiment by instructing half of the participants to hold coherent motion perception and the other half to hold separate motion perception. Twelve participants were randomly assigned to each of the two intention conditions.

Participants had to report their motion perception (coherent or separate) by key-pressing during two one-minute trials in each block. The average percentage of time perceiving coherent motion was measured as a dependent variable.

Materials

The visual stimulus used here was the diamond stimulus mentioned above, containing four gray occluders, four white moving lines, and a fixation on the center. It subtended 7.43º × 7.51º of the visual angle on the monitor programmed by using Macromedia Flash MX 2004.

Apparatus

A View Sonic 17-inch CRT monitor with a refresh rate 75 Hz and screen resolution of 1024 × 768 pixels was used to present the stimulus. Participants' heads were restrained with a chinrest so their observation distance could be kept at 100 cm to the middle of screen. In order to isolate other light sources in the environment, participants viewed the stimuli through an observation box. Two numerical keyboards were used for key-pressing at each hand to start each trial and record responses.

Procedures

At the beginning of the experiment, the nature of the diamond stimulus was explained to the participants. The experiment could proceed only after it is confirmed that each of the two motion interpretations of the diamond stimulus can be perceived. Each block contained two one-minute-trials. Before the beginning of each block, there were two 30-second practice trials to let participants practice the procedures of their particular intention and attention conditions.

In each trial, the diamond stimulus mentioned previously was presented for one minute on a black background. There was a fixation cross (+) located on the center of the stimulus for participants to keep their fixation on. Participants had to report their motion perception by key-pressing during one-minute trials. Pressing the numeral “1” key indicated that they perceived coherent motion perception, and pressing the numeral “3” key indicated separate motion perception. The average percentage of time perceiving coherent motion was measured as the dependent variable.

2.1.2 Results

Mean percentage of time perceiving coherent motion under the 2 (attention) × 2 (intention) conditions is plotted in Figure 9. Two-way ANOVA show that both the main effect of attention and intention are significant. The percentage of time perceiving coherent motion is higher in the attending-to-occluders condition (47.15%) than in the attending-to-moving-lines condition (41.26%) ($F(1,10) = 5.49$, $p < 0.05$, partial $\eta^2 = 0.354$), which is consistent with the hypothesis that attention can bias multistable motion perception. Further, the percentage of time perceiving coherent motion in the "hold coherent" condition (52.84%) is higher than that in the "hold separate" condition (35.57%) ($F(1,10) = 11.31$, $p < 0.01$, partial $\eta^2 = 0.531$), which is consistent with the intention control effect found previously. There is no significant interaction between attention and intention ($F(1,10) = 0.17$, $p = 0.689$). The effects of attention and intention are additive, implicating that they are independent mechanisms on multistable motion perception. The effect of attention is similar to the effect of fixation found previously (e.g., Toppino, 2003) in two aspects. First, all have no interaction with the effect of intention. This implies that the mechanisms of attention and intention are independent. Second, the effect of intention on affecting multistable perception seems more powerful than the effects of fixation and attention (The mean difference of the percentage of time perceiving coherent motion under the two attention conditions is 5.89%; the mean difference between the two intention conditions is 17.27%).

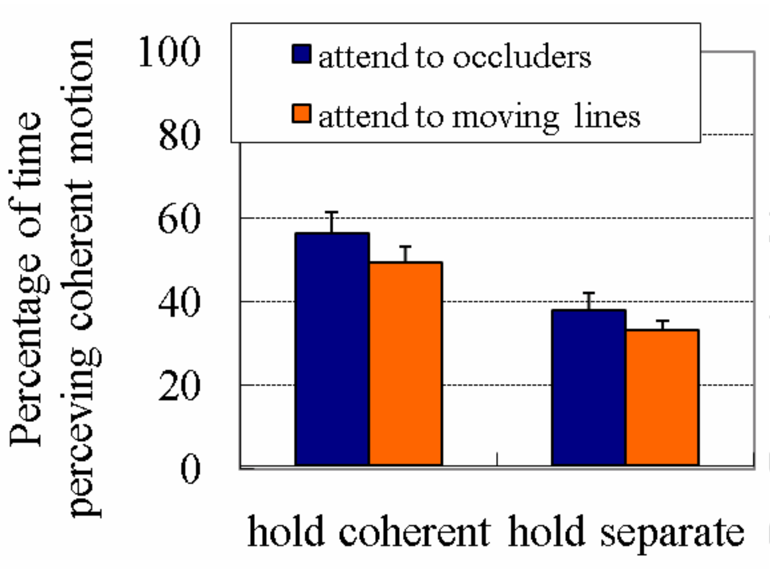


Figure 9. Results of Experiment 1a. Means and 1 standard errors of the percentage of time that the diamond stimulus was perceived coherent motion are plotted under the 2 (attention) × 2 (intention) conditions.

2.1.3 Discussion

Results of this experiment show that spatial attention can bias multistable motion perception, and the type of bias is consistent with the prediction of the hypothesis. That is, the percentage of time perceiving coherent motion is higher in the attending-to-occluders condition than the attending-to-moving-lines condition. This result is consistent with that of attention biased multistable motion perception by making attended areas look closer, but it does not directly support this notion. This notion will be directly investigated in Experiment 2.

The effect of attention is comparatively smaller than the effect of intention. One reason might be that the trial duration (1 minute) was too long to keep the participant's attention on demanded areas, weakening the effect of attention. Thus, in Experiment 1b, the trial duration will be reduced, and a manipulation check task of attention will be executed to examine whether participants allocate their attention on demanded areas. In addition, there was no interaction between attention and intention (*power* = 0.725, which is estimated by Cohen's (1988) medium effect size *f* = 0.25), implicating that attentional control and intentional control are independent mechanisms in influencing multistable motion perception. Therefore, the control of intention can be held constant—all coherent motion—in further experiments of this study because the effects of attention under the two intention conditions are similar.

2.2 Experiment 1b: Adding a Manipulation-Check Task to Spatial Attention

In Experiment 1a, participants’ attention was manipulated by instructing them to attend to the four occluders or attend to the four moving lines. However, it is still not clear whether participants obeyed these instructions. Hence, in Experiment 1b, a manipulation-check task of spatial attention was added. Participants had to respond to a probe (the lightening of an occluder or a moving line) presented at the end of each trial as fast as possible. The reaction time to the probe was analyzed to verify whether the manipulation of spatial attention was valid. The prediction is that if participants allocated their attention according to the instructions, RT to the lightening of the occluder and line should be different. The RT should be shorter to for the lightening of attended areas than for the unattended areas.

2.2.1 Methods

Participants

Another nine participants were recruited with the same standards as described for Experiment 1a.

Design

This experiment is a one factor (attention) within-participant design. Participants' attention was manipulated by instructing them to attend to the four occluders and attend to the four moving lines in different blocks. Simultaneously, all participants were instructed to hold coherent motion perception through the experiment. In addition, at the end of each trial, participants had to respond to a probe—the lightening of an occluder or a moving line—by pressing a key as fast as possible. The response time (RT) to the probe was measured as a dependent variable. In addition, participants had to report their motion perception (coherent or separate) by key-pressing until a probe presented. The average percentage of time in perceiving coherent motion was measured as another dependent variable.

Materials

The stimuli used in this experiment were similar to that in Experiment 1a, except that the four gray occluders were changed into hollow occluders to make the two probe conditions similar (that is, the light of an occluder outline or the light of a line).

Apparatus

The apparatus used here was identical to that in Experiment 1a.

Procedures

The procedure was similar to Experiment 1a, except that participants had to additionally respond to a probe that was shown 3, 4, 5, 6, 7, or 8 seconds after the trial began by pressing the numeral “4” key with the left hand. The six probe-showing times were the same in all blocks, but the sequence was different (by randomizing) so participants could not predict when the probe would present itself. After responding to the probe, the trial ended and a new trial began. Since the recording of motion perception was terminated during probe presenting, the lightening of the occluder or line would not influence the multistable motion perception process.

This experiment had four blocks—two attending-to-occluders blocks and two attending-to-moving lines blocks. The order of the four blocks was counterbalanced within participants. Each block had six trials corresponding to the six probe-showing times. In total, each participant has to do 4 × 6 = 24 experimental trials.

2.2.2 Results

The trials in which the RT to the probe was greater than 1500 ms were deleted and not included in further analysis. One trial was deleted for two participants, and another participant's entire data were deleted because the deleted trials were too much (six trials of total 24 trials). Thus, the data for further analysis contained only eight participants' non-deleted data.

The mean percentage of time perceiving coherent motion and RT to the lightening of the occluder or line under the two attention conditions are plotted in Figure 10. One-way ANOVA shows that the effect of attention is significant. The percentage of time perceiving coherent motion is higher in the attending-to-occluders condition (67.28%) than in the attending-to-moving lines (49.17%) ($F(1,7) = 24.87$, $p < 0.01$, partial $\eta^2 = 0.780$), which is consistent with the hypothesis that attention can bias multistable motion perception. The type of bias was also consistent with the prediction that the attended areas look closer. On the other hand, a two-way ANOVA of RT shows that the interaction of the RT to the probe under the two attention conditions and the two probe conditions is significant ($F(1,7) = 6.35$, $p < 0.05$, partial $\eta^2 = 0.476$), and the pattern of interaction in Figure 10 implies that the participants allocated their attention on demanded areas. That is, the value of subtracting the RT of lightening line from the RT of occluder lightening under the attending-to-lines condition is larger than that under the attending-to-occluders condition. The main effect of the two attention conditions (*F*(1,7) = 0.91 , *p* = 0.371) and the two probe conditions (*F*(1,7) = 1.38 , *p* = 0.278) are not significant.

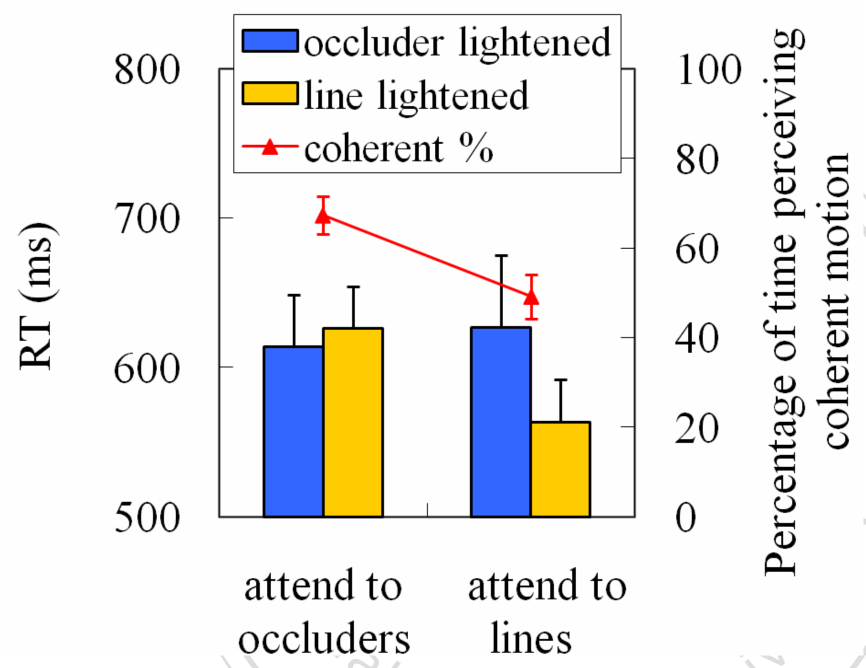


Figure 10. Results of Experiment 1b. Mean percentage of time perceiving coherent motion and RT to the lightening of an occluder or a line under the two attention conditions.

In a further analysis of the data of RT to the probe, the eight participants can be divided into two groups depending on whether their individual patterns of interaction imply that the participants allocated their attention on demanded areas, which is defined by “the index of attention.” That is, subtract the RT of line lighting from the RT of occluder lighting under the attending-to-lines condition and under the attending-to-occluders condition, then subtract the later from the former; this value is the index of attention. If the index of attention is positive, the participant is assigned to the "positive interaction" group, implying he/she does allocate his/her attention on demanded areas. Conversely, participants are assigned to the "negative interaction" group if the index of attention is negative.

The mean percentage of time perceiving coherent motion under the 2 (group) × 2 (attention) conditions are plotted in Figure 11. There are seven participants in the “positive interaction” group and only one participant in the “negative interaction” group. Two-way ANOVA shows that the effect of attention ($F(1,6) = 6.26$, $p < 0.05$, partial $\eta^2 = 0.510$) and interaction is significant ($F(1,6) = 6.67$, $p < 0.05$, partial $\eta^2 = 0.527$). Simple main effect shows that the effect of attention is significant only in the “positive interaction” group ($F(1,6) = 51.72$, $p < 0.001$) but not significant at the “negative interaction” group ($F(1,6) = 0$, $p = 0.965$). This evidence implies that the effect of attention will show, as predicted, only when participants allocate their attention on demanded areas (in the “positive interaction” group). Hence, the index of attention can be used in later experiments to filter out the data that does not meet the criterion of allocating attention on demanded areas.

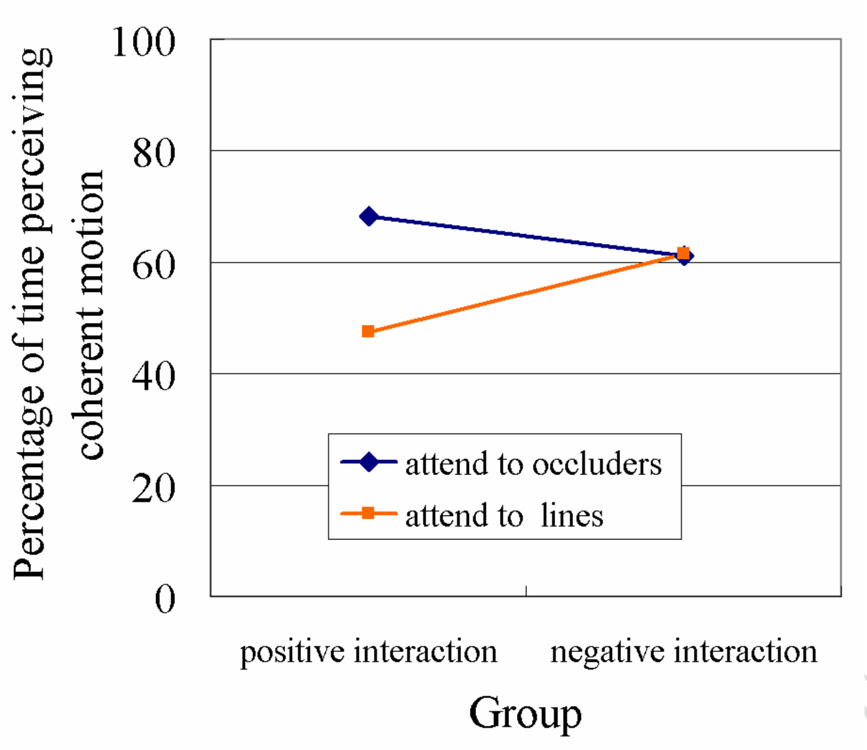


Figure 11. The mean percentage of time perceiving coherent motion under the 2 (group) × 2 (attention) conditions.

2.2.3 Discussion

This experiment repetitively verified the main result of Experiment 1a that the percentage of time for perceiving coherent motion was higher in the attending-to-occluders than in the attending-to-moving-lines condition. This result is consistent with the prediction that spatial attention can bias multistable motion perception by making attended areas look closer. Furthermore, after reducing Experiment 1a's trial duration from 1 minute to 3 to 8 seconds, the effect of attention increased dramatically. The difference of the percentage of time perceiving coherent motion under the two attention conditions increased from 5.89% to 18.11%. This implies that participants can keep their attention on demanded areas better during a shorter duration trial than in a longer one. The manipulation check task of attention also showed that most participants allocated their attention according to the instructions. On the other hand, the increasing effect of attention brings another interesting issue worth investigating in the future: is the effect of intention still more powerful than the effect of attention after reducing the trial duration? Because intention was not manipulated in this experiment, this question cannot be answered here.

However, even if the results of Experiments 1a and 1b were consistent with the hypothesis that attention can bias multistable motion perception, it is still not certain whether the effect of attention is actually affected by the depth perception mechanism, making attended areas look closer. Accordingly, in Experiment 2, this question will be verified directly.

3. Experiment 2: The Effect of Attention: Can Spatial Attention Alter Perceived Depth?

Experiment 2 is a depth-judgment task designed to investigate whether spatial attention can affect depth perception, making attended areas look closer. This experiment has nothing to do with multistable motion perception, but the diamond stimulus is still used here in order to determine whether spatial attention affects the stimulus used in Experiment 1b. In other words, the test will indicate whether attending to the four occluders can make the occluders look closer and whether attending to the four lines can make the lines look closer.

In the depth-judgment task, nine different levels of binocular disparity of the occluders and the lines are manipulated. Participants have to judge which one is farther in depth, the four occluders or the four lines, after the diamond stimulus is briefly shown. Participants' attention is also manipulated in different blocks, as in Experiment 1b. The point of subjective equality (PSE, on which the participants have a 50% probability of reporting that the lines are more behind the occluders) will be measured as the dependent variable, and it is predicted that the PSE under the attending-to-occluders condition will be different from that under the attending-to-lines condition. If attention can make the attended areas look closer, then when participants attend to the four occluders, the occluders will look closer. Thus, participants will perceive occluders and lines in the same plane only when the lines are defined by binocular disparity as in front of occluders, so the PSE should be smaller than zero. In the same way, the four lines will look closer when attending to the four lines. Thus, participants will perceive occluders and lines in the same plane only when the occluders are defined by binocular disparity as in front of lines, so the PSE should be larger than zero. Therefore, the PSE under the condition of attending to the four lines will be larger than under the condition of attending to the four occluders. A manipulation-check task of spatial attention (as in Experiment 1b) will also be executed to verify whether the manipulation of attention is valid.

3.1 Methods

Participants

Another 14 participants were recruited with similar standards as described for Experiment 1a. After finishing this experiment, they received either one bonus point in the "Methods of Psychological Experiment" course or 100 NT dollars.

Design

The experiment is a 2 (attention) × 9 (binocular disparity) completely within-participant design. Participants' attention was manipulated by instructing them to attend to the four occluders or to the four lines in different blocks, as in Experiment 1b. Binocular disparity was manipulated using a stereoscope at nine different levels (-4, -3, -2, -1, 0, 1, 2, 3, and 4 pixels). A greater positive disparity means that the lines were more behind the occluders; the greater negative disparity means that the lines were more in front of the occluders. Zero disparity means that the occluders and the lines were at the same depth. Participants had to judge which were behind (rather than in front), the occluders or the lines. Thus, the possibility that participants' judgments were biased toward the instructions of attention can be ruled out. The point of subjective equality (PSE, on which the participants had a 50% probability of reporting that the lines were behind the occluders) was calculated as a dependent variable.

The experiment contained eight blocks split over two days; each day contained four blocks: two attending-to-occluders blocks and two attending-to-lines blocks. The order of the two attention conditions was counterbalanced within participants. Each block contained 22 trials in which the nine levels of binocular disparity were repeated two times, respectively, in a random sequence. The main task was to judge the depth relationship between occluders and lines. The remaining four trials were probe-detecting trials, including two occluder-lightening trials and two line-lightening trials (all were in zero disparity). In these trials, participants had to respond to the probe as fast as possible and did not need to do depth judgment. In total, each participant had to complete 2 × 4 × 22 = 176 experimental trials in two days. The PSE was calculated as a dependent variable.

Materials

The diamond stimulus used here was identical to that in Experiment 1b, except that it was static (the four lines were not moving).

Apparatus

The apparatus used here was similar to that described for Experiment 1b, except that the View Sonic CRT computer monitor was changed from a 17-inch to a 19-inch and a stereoscope was added. With the help of the stereoscope, participants can easily fuse a stereogram with binocular disparity to form the perception of the three-dimensional diamond stimulus.

Procedures

The procedures were similar to that described for Experiment 1b, except that participants did not have to respond to their motion perception of the diamond stimulus because it was static. Instead, they had to judge the depth relationship between occluders and lines.

At the beginning of each block, participants were instructed to allocate their attention on the four occluders or the four lines throughout the whole block. Participants pressed the numeral key “4” with their left hand to start each trial by themselves when they were well prepared (fixated on the central cross and concentrating). After 150 ms, 300 ms, or 450 ms latency—applied at random to eliminate the expectancy effect, increasing the level of attention—the static diamond stimulus appeared for 600 ms and then was covered by a random-dot mask to prevent the afterimage from interfering. Participants had to report their depth judgment—occluders or lines were behind—by key-pressing after the diamond stimulus was shown. Pressing the numeral key “1” with the right hand indicated that they perceived the occluders behind the lines, and pressing the right-hand numeral key “3” indicated that the lines were seen as behind the occluders. In probe-detecting trials, the lightening of an occluder or a line was presented using 150 ms, 200 ms, or 250 ms latency after the diamond stimulus was shown. Participants had to press the left-hand numeral key “6” as fast as possible when seeing the probe. The next trial would not begin until participants pressed the key to report their depth judgment or probe detection. The flow chart of the procedure in Experiment 2 is shown in Figure 12.

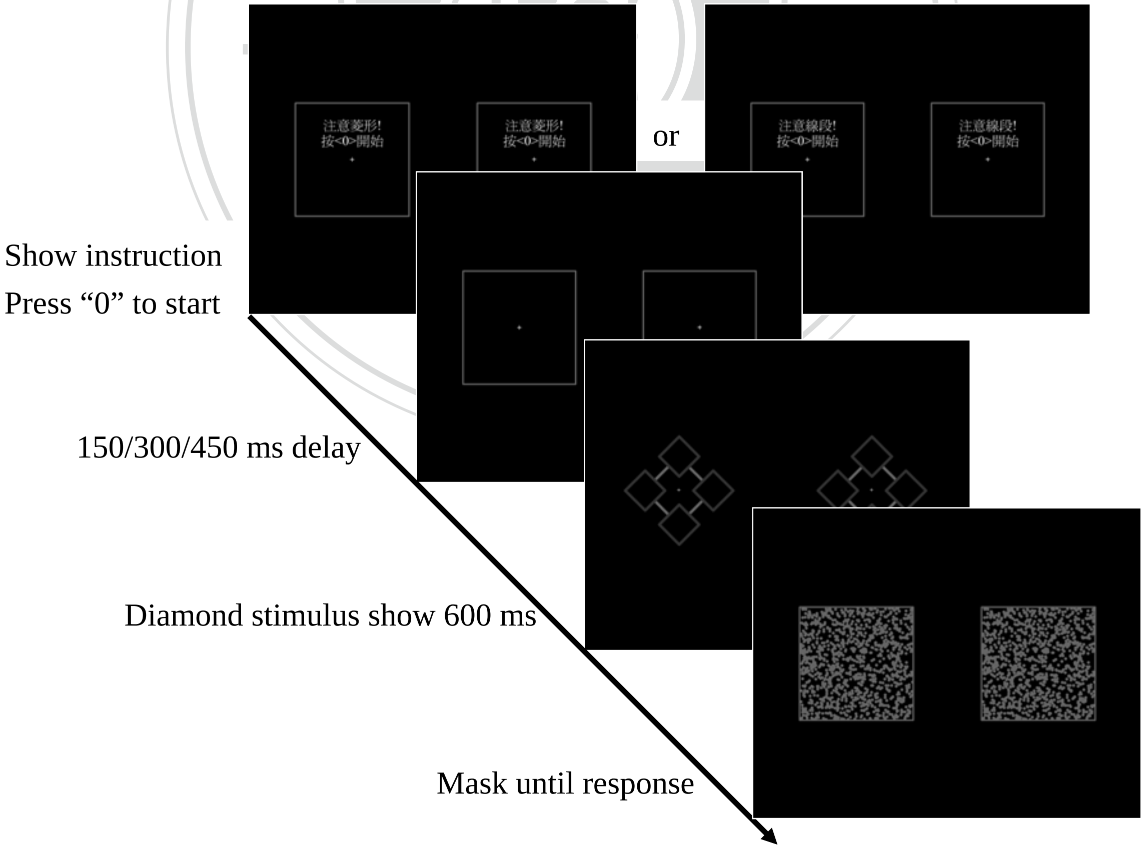


Figure 12. The flow chart of the depth-judgment trials in Experiment 2. By using a stereoscope, participants' left eye and right eye see a left picture and right picture, respectively, which is then fused into a three-dimensional diamond stimulus.

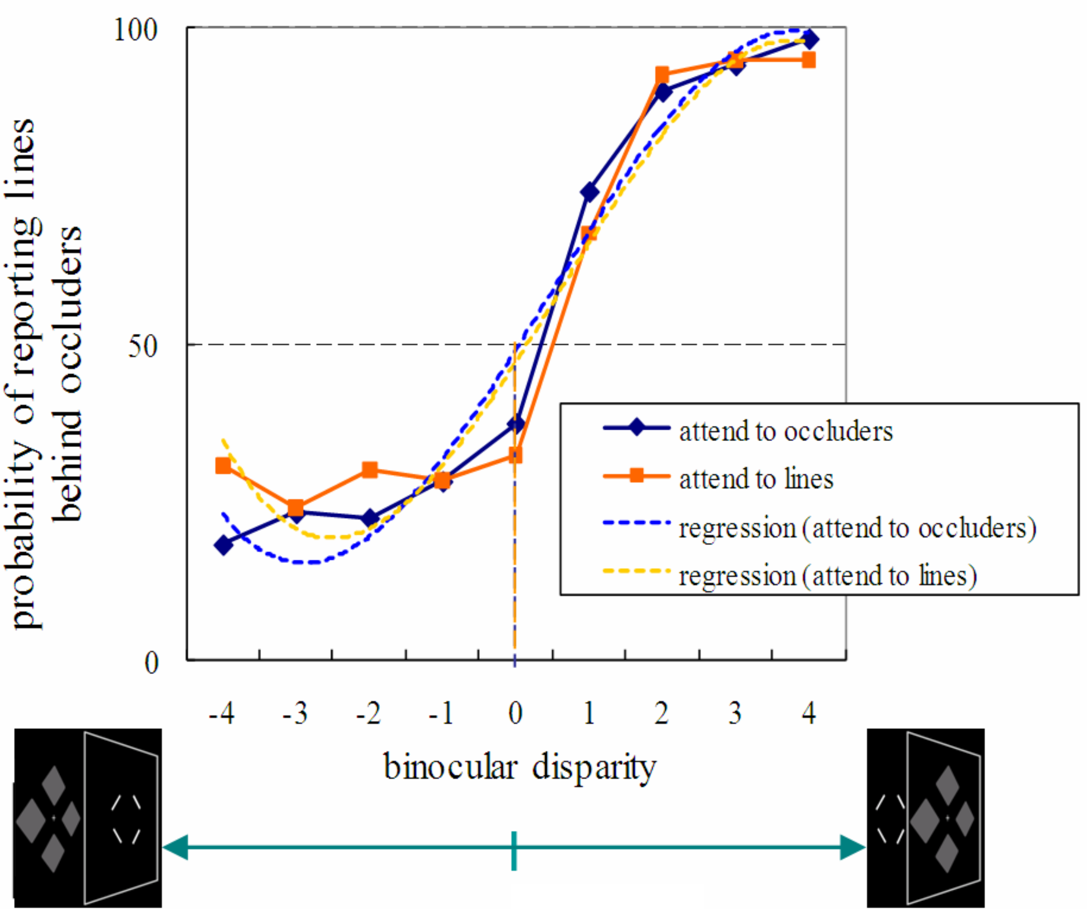
3.2 Results

The mean probability of reporting that the lines were behind the occluders under 2 (attention) × 9 (binocular disparity) conditions in eight blocks were counted and charted in Figure 13a. The t-test shows that the PSE under the attending-to-occluders condition (PSE = 0.087) and attending-to-lines condition (PSE = -0.019) is not different ($t(13) = 0.413$, $p = 0.686$), which is not consistent with the prediction. In addition, the two-way ANOVA of RT to the probe in eight blocks also shows that the interaction is not significant as predicted ($F(1,13) = 2.53$, $p = 0.136$), implicating that participants did not totally follow the instructions to allocate their attention. This might be why the results of the PSE were not consistent with the prediction. Next, only the data in which participants followed instructions to allocate their attention were selected for further analysis. The method of selection is described below.

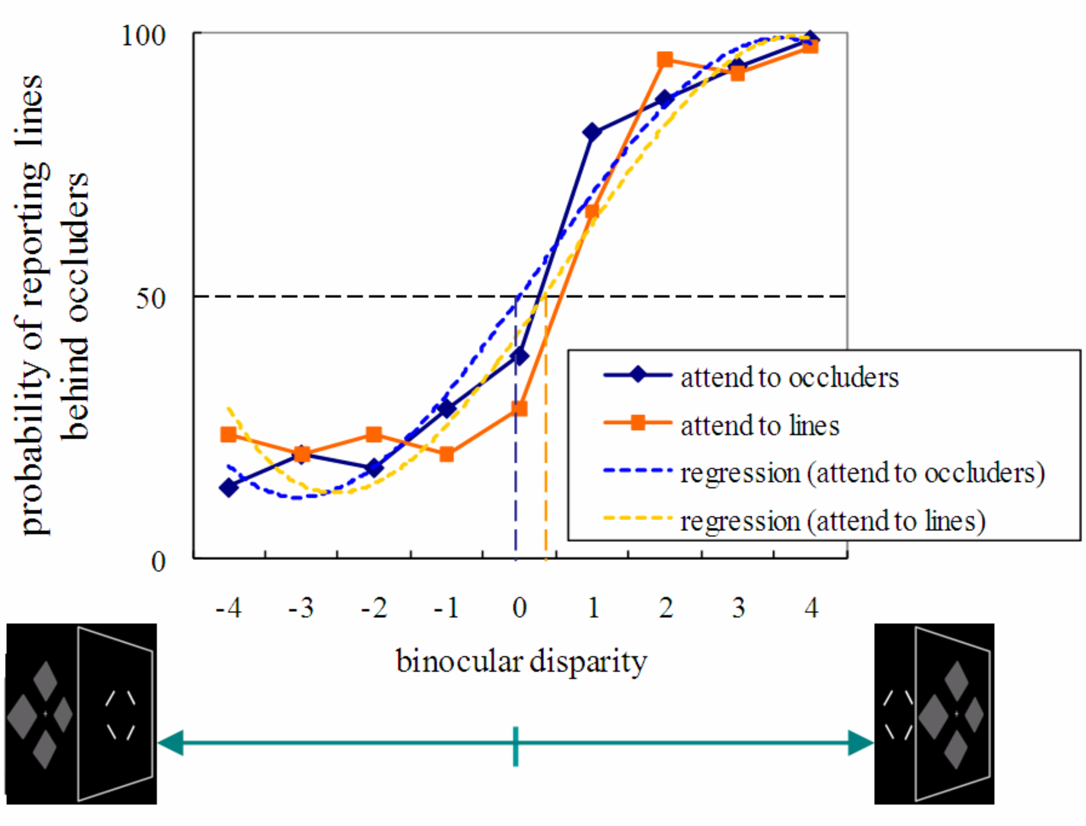
The eight blocks were divided into four sections (Section 1: Blocks 1 and 2; Section 2: Blocks 3 and 4, and so on). Each section contained an attending-to-occluders block and an attending-to-lines block to check whether the index of attention (as described in Experiment 1b) in each section is positive. Only the sections in which the index of attention is positive—implying that participants allocated their attention according to the instructions—are selected for further analysis. There were 26 sections selected (total: 4 sections × 14 participants = 56 sections; three participants' data were not selected), which contained 12 attending-to-occluders-first and attending-to-lines-next sections and 14 attending-to-lines-first and attending-to-occluders-next sections. Thus, the confounding of order effect would be very small. Two-way ANOVA of RT to the probe shows that the interaction is significant ($F(1,9) = 9.76$, $p < 0.05$, partial $\eta^2 = 0.520$), implicating that participants allocated their attention according to the instructions in the selected sections. Scheffe's post hoc test shows that participants' response to the lightening of lines was shorter than to the lightening of occluders at attending-to-lines condition ($F = 33.92 > F' = 5.12$) but not at the attending-to-occluders condition ($F = 1.97$).

Mean probability of reporting that the lines were behind the occluders under 2 (attention) × 9 (binocular disparity) conditions in the selected sections were counted and are charted in Figure 13b. Two-way ANOVA shows that the main effect of binocular disparity is significant ($F(8,72) = 47.28$, $p < 0.001$, partial $\eta^2 = 0.840$), implicating that the manipulation of binocular disparity is valid. However, neither the main effect of attention nor interaction are significant ($F(1,9) = 0.44$, $p = 0.523$; $F(8,72) = 1.61$, $p = 0.138$).

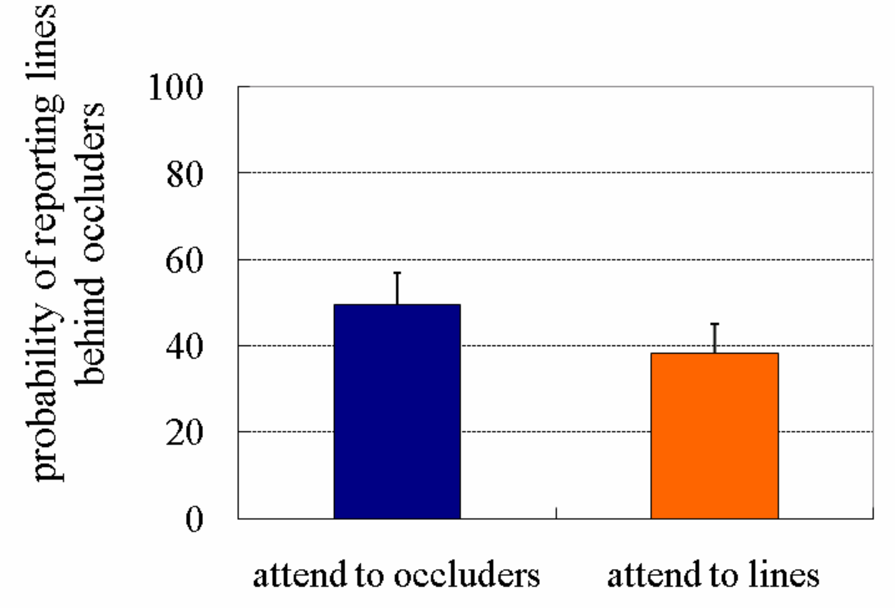
The PSE under the attending-to-lines condition (PSE = 0.179) is larger than under the attending-to-occluders condition (PSE = -0.045) and consistent with the prediction, but the t-test of PSE is not significant ($t(9) = 1.272$, $p = 0.235$). However, the mean probability of reporting lines behind occluders in small binocular disparities (-1, 0, and 1) is significantly higher under the attending-to-occluders condition (49.58%) than that under the attending-to-lines condition (38.33%) ($t(9) = 4.045$, $p < 0.01$) (see Figure 13c). This evidence supports the hypothesis that attention can affect depth perception, making attended areas look closer.



(a)


probability of reporting lines behind occluders
100
50
0
-4
-3
-2
-1
0
1
2
3
4
binocular disparity
attend to occluders
attend to lines
regression (attend to occluders)
regression (attend to lines)

(b)


probability of reporting lines behind occluders
100
80
60
40
20
0
attend to occluders
attend to lines

(c)

Figure 13. Results of Experiment 2. (a) Mean probability of reporting that the lines behind the occluders were plotted under 2 (attention) × 9 (binocular disparity) conditions in eight blocks. (b) Mean probability of reporting that the lines were behind the occluders in the selected sections were plotted under 2 (attention) × 9 (binocular disparity) conditions. (c) The mean probability of reporting lines behind occluders in small binocular disparities (-1, 0, and 1) under the two attention conditions in the selected sections.

3.3 Discussion

The mean probability of reporting lines behind occluders in small binocular disparities (-1, 0, and 1) is significantly higher under the attending-to-occluders condition than that under the attending-to-lines condition, consistent with the hypothesis that attention can affect depth perception, making attended areas look closer. The resulting pattern also shows that PSE under the attending-to-lines condition is larger than under the four occluders condition (although not significant) and is consistent with the prediction.

There were some problems that may influence the results. One problem is that the probability of reporting lines behind occluders did not decrease with greater negative disparity. The reason might be the conflict between binocular and monocular depth cues at negative disparity. That is, due to the monocular depth cues, such as T-junctions, the lines are easily perceived as completed behind the occluders. However, the lines should be perceived in front of the occluders under negative disparity conditions. In fact, most participants reported that it was hard to perceive lines in front of occluders (in negative disparity conditions) while it was easy to perceive lines behind occluders (in positive disparity conditions). Notice that this problem may also affect the PSEs measured here because the probability of reporting lines behind occluders in the negative disparities can influence the function of the entire regression line.

Another problem is that in the zero-disparity condition, the probability of reporting lines behind occluders should be around 50%; however, it is below 50% in both attention conditions. This may be induced by a response strategy from the participants. As mentioned above, most participants reported that it was difficult to perceive lines in front of occluders, so they tended to respond “occluders behind lines” in the trials in which they had difficulties judging the depth. This can explain why participants showed less probability of reporting lines behind occluders in the zero-disparity condition.

4. Experiment 3: Can Binocular Disparity Affect Spatial Attention?

Experiments 3 and 4 are designed to test the hypothesis in different ways. Experiment 3 tests whether the effect of binocular disparity can override the effect of spatial attention. The hypothesis is that attention can bias multistable motion perception by making the attended areas look closer in depth. Accordingly, if there is a clear binocular depth cue (by manipulating binocular disparity) to define the depth relationship between occluders and moving lines, then the effect of spatial attention on biasing multistable perception should be eliminated or overridden. In particular, the larger the binocular disparity is manipulated, the more the effect of attention will be overridden. In other words, the effect of spatial attention on biasing multistable motion perception should decrease or disappear with increasing binocular disparity. The manipulation-check task of spatial attention used in Experiment 1b and Experiment 2 will also verify whether participants allocate their attention according to the instructions.

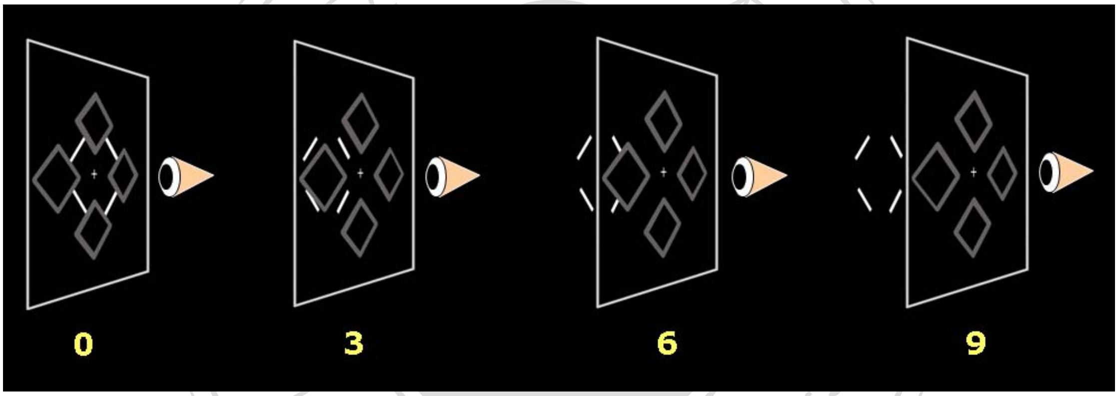
4.1 Methods

Participants

Another 20 participants were recruited using the same standards described for Experiment 1a. The number of participants increased in this experiment because Experiment 2 showed that much data (nearly half) would be deleted and not included in further analysis under the criterion of positive index of attention. Increasing participants can increase statistical power against data deleting.

Design

The experiment is a 2 (attention) × 4 (binocular disparity) completely within-participant design. Participants' attention was manipulated by instructing them to attend to the four occluders or attend to the four moving lines in different blocks, as in Experiments 1 and 2. The order of the two attention conditions was counterbalanced within participants. Binocular disparity was manipulated at four different levels (0, 3, 6, and 9 pixels, showed in Figure 14) with a stereoscope. The larger disparity means that the moving lines were seen as more behind the occluders. Zero disparity means that the occluders and the moving lines were at the same depth. In addition, participants' intention was controlled to be the same (by instructing them to hold the perception of coherent motion) throughout the experiment, as in Experiment 1b. Participants had to report their motion perception (coherent or separate) by key-pressing during 3000 ms to 6500 ms trials. The average percentage of time perceiving coherent motion was measured as a dependent variable. In addition, at the end of each trial, participants had to respond to a probe—the lighting of an occluder or a moving line—by pressing a key as fast as possible.



(a) (b) (c) (d)

Figure 14. The diagram of the binocular disparities manipulated in Experiment 3: (a) binocular disparity 0, (b) binocular disparity 3, (c) binocular disparity 6, and (d) binocular disparity 9.

Materials

The diamond stimulus used here was the same as in Experiment 2.

Apparatus

The apparatus used here was identical to that in Experiment 2, including a View Sonic 19-inch CRT computer monitor, a chinrest, an observation box, a stereoscope, and two numerical keyboards.

Procedures

At the beginning of the experiment, the experimenter instructed participants to adjust the stereoscope and a depth-judgment task was executed in order to determine whether the participants can see three-dimensional pictures correctly, as in experiment 2. Only the participants who pass this task can continue.

The nature of multistable figures was then explained to participants. The experiment can proceed only after confirming that each of the two motion interpretations of the diamond stimulus can be perceived, as in Experiment 1a.

A practice block was then executed. Participants were trained to allocate their attention on the four occluders or the four lines and respond to their motion perception repeatedly until they were well practiced and familiar with the entire procedure.

At the beginning of each block, the experimenter told participants to allocate their attention on the four occluders or the four lines throughout the block. Participants pressed the left-hand "0" key to start each trial when they were well prepared (fixate on the central cross and allocate their attention on demanded areas). Participants had to report their motion perception by key-pressing, as in Experiment 1b.

The manipulation-check task of spatial attention was also added in this experiment, as in Experiment 1b. A probe was shown at 3000 ms, 3500 ms, 4000 ms, 4500 ms, 5000 ms, 5500ms , 6000ms, or 6500ms after the trial began, and participants had to additionally respond by pressing the numeral “4” key with their left hands as fast as possible. The eight probe-showing times were the same in all blocks, but the sequence was different (by randomizing). After responding to the probe, the trial ended and a new trial began. Since the recording of motion perception was terminated while the probe was presenting, the lighting of the occluder or line would not influence the multistable motion-perception process.

The experiment contained eight blocks: four attending-to-occluders blocks and four attending-to-moving-lines blocks. The sequence of the eight blocks was counterbalanced between participants and split over two days. Each block had eight trials in which the four levels of binocular disparity were presented twice, respectively, in a random sequence. In total, each participant completed 8 × 8 = 64 experimental trials in two days.

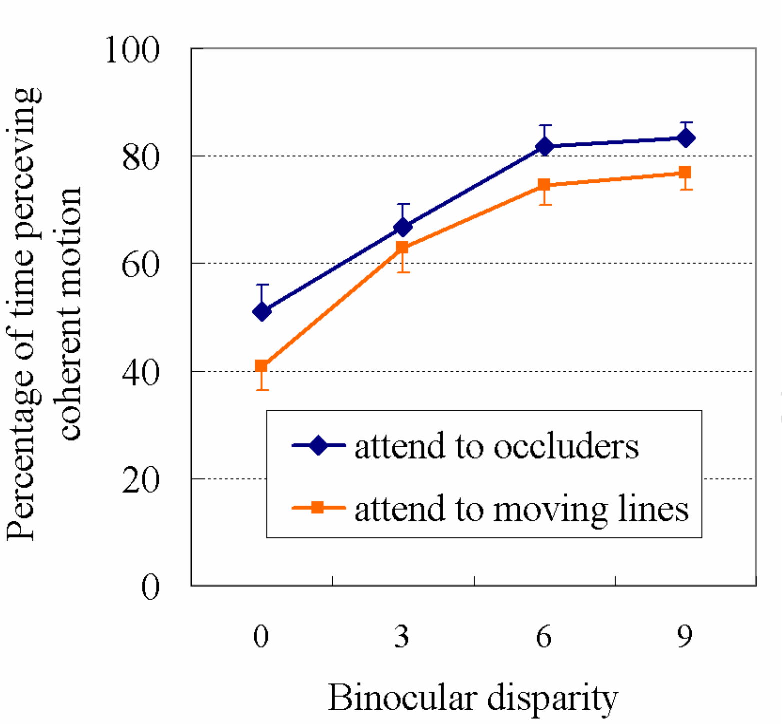
4.2 Results

Mean percentage of time perceiving coherent motion under the 2 (attention) × 4 (binocular disparity) conditions in eight blocks is plotted in Figure 15a. Two-way ANOVA shows that the main effect of attention is significant ($F(1,19) = 4.50$, $p < 0.05$, partial $\eta^2 = 0.191$). The percentage of time perceiving coherent motion is higher in the attending-to-occluders condition (70.76%) than in the attending-to-moving-lines condition (63.73%), which is consistent with the hypothesis. The main effect of binocular disparity is also significant ($F(3,57) = 51.42$, $p < 0.001$, partial $\eta^2 = 0.730$), implicating that the manipulation of binocular disparity is valid. However, the interaction of attention and binocular disparity did not reached a significant level ($F(3,57) = 0.72$, $p = 0.541$). In addition, the two-way ANOVA of RT to the probe in eight blocks also shows that the interaction is not significant, as predicted ($F(1,19) = 1.71$, $p = 0.207$), implicating that participants did not fully follow the instructions to allocate their attention. Next, only the data of participants who followed instructions and allocated their attention were selected for further analysis.

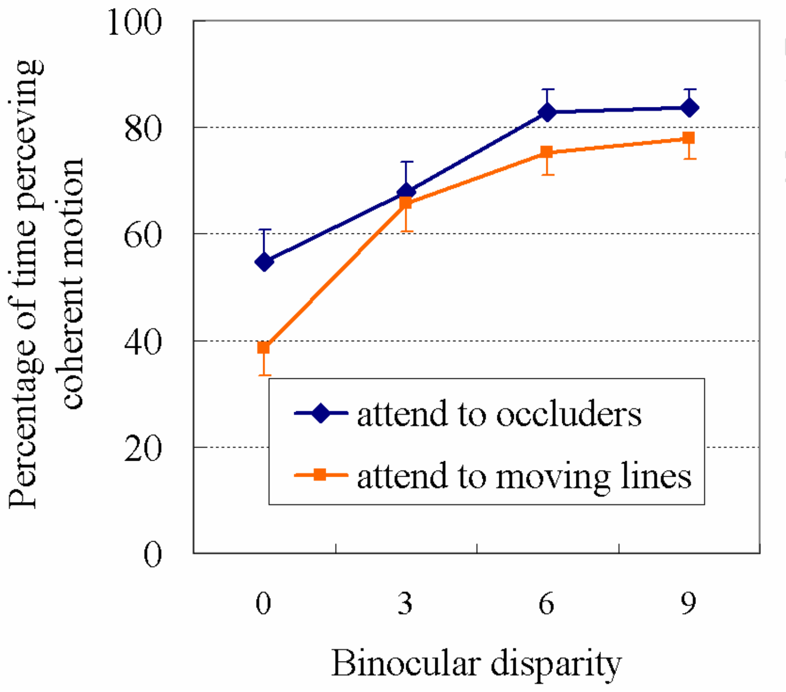
The method of selection was the same as in the Experiment 2: the eight blocks were divided into four sections (Section 1: Blocks 1 and 2; Section 2: Blocks 3 and 4, and so on). Each section contained an attending-to-occluders block and an attending-to-moving-lines block to check whether the index of attention (as described in Experiment 2) in each section is positive. Only the sections in which the index of attention is positive—implying that participants allocated their attention according to the instructions—were selected for further analysis.

There were 40 selected sections (total: 4 section × 20 participants = 80 sections), which contained 24 attending-to-occluders-first and attending-to-moving-lines-next sections and 16 attending-to-moving-lines-first and attending-to-occluders-next sections.

The mean percentage of time perceiving coherent motion under the 2 (attention) × 4 (binocular disparity) conditions in the selected sections is plotted in Figure 15b. Two-way ANOVA shows that only the main effect of binocular disparity is significant ($F(3,57) = 41.62$, $p < 0.001$, partial $\eta^2 = 0.687$) and consistent with the prediction. However, the main effect of attention ($F(1,19) = 3.16$, $p = 0.092$, partial $\eta^2 = 0.142$) and the effect of interaction ($F(3,57) = 1.59$, $p = 0.201$, partial $\eta^2 = 0.077$) are not significant. The simple main effect shows that the effect of attention is significant only at binocular disparity 0 ($F(1,19) = 4.76$, $p < 0.05$) and not significant at binocular disparity 3 ($F(1,19) = 0.1$, $p = 0.760$), 6 ($F(1,19) = 2.21$, $p = 0.154$), or 9 ($F(1,19) = 1.73$, $p = 0.204$). This evidence supports the prediction that the effect of attention decreases at non-zero binocular disparity due to the effect of binocular disparity weakening the effect of attention.



(a)



(b)

Figure 15. Results of Experiment 3. (a) Means and 1 standard errors of the percentage of time that the diamond stimulus was perceived as coherent motion under the 2 (attention) × 4 (binocular disparity) conditions in eight blocks. (b) Means and 1 standard errors of the percentage of time that the diamond stimulus was perceived as coherent motion under the 2 (attention) × 4 (binocular disparity) conditions in the selected sections.

4.3 Discussion

The resulting pattern of the selected sections shows that the effect of attention decreases with increasing binocular disparity (the difference of the percentage of time perceiving coherent motion under the two attention conditions decreased from 16.37% and 2.14% to 7.62% and 5.79% with increasing binocular disparity). The simple main effect shows that the effect of attention is significant only at binocular disparity 0 and not significant at the other three large binocular disparities (*power* = 0.865, which is estimated by Cohen’s (1988) medium effect size *f* = 0.25). This result is consistent with the hypothesis and implies that binocular disparity weakens the effect of attention.

The main problem of the results is that the effect of attention seems to be mostly blocked at binocular disparity 3 but still appears slightly at binocular disparities 6 and 9. A possible reason is that other factors interfered with the entire mechanism at binocular disparities 6 and 9. For example, in the attending-to-moving-lines condition, the occlusion effect (the tendency to see the lines as behind the occluders completed into a diamond and moving coherently) of the occluders may decrease with increasing binocular disparity. This is because it is easier to ignore the occluders and attend to lines alone when binocular disparity is large, so the occlusion effect decreases. Accordingly, in the attending-to-moving-lines condition, the percentage of time perceiving coherent motion perception at binocular disparity 6 or 9 is not as great as that in the attending-to-occluders condition as predicted. To solve this problem, the manipulation of binocular disparity will be changed into the manipulation of monocular depth cues in Experiment 4.

5. Experiment 4: Can Monocular Depth Cues Block the Effect of Spatial Attention?

Experiment 4 is designed to examine whether strong monocular depth cues can eliminate the effect of spatial attention. The reasoning is similar to that of Experiment 3: if attention can bias multistable perception by making the attended areas look closer, then giving a strong monocular depth cue to define the depth relationship between occluders and moving lines should override or eliminate the effect of spatial attention.

There are many types of monocular depth cues. For example, spatial frequency is a source of information about monocular depth. Klymenko and Weisstein (1986) examined the influence of spatial frequency on multistable figure perception. They found that the region containing higher spatial frequency sine wave grating had a higher percentage of time being seen as a figure than the region of lower spatial frequency grating. This is because the presence of higher spatial frequencies in the stimulus is often correlated with detail information, and the region that is more detailed tends to be seen as a figure. Another interpretation is that "the high and low spatial frequencies simulate high and low degrees of focus and therefore are interpreted as lying in different depth planes" (Klymenko & Weisstein, 1986, p. 329). Therefore, in Experiment 4, high spatial frequency (HSF) information of the occluders is filtered out (the blurred-occluders condition) or HSF information of the moving lines is filtered out (the blurred-moving-lines condition) in order to make them look far in depth. The main prediction is that the effect of attention should decrease or disappear under the blurred-moving-lines and blurred-occluders conditions due to the presence of clear depth cues, but the effect of attention should still be present under the normal condition.

5.1 Methods

Participants

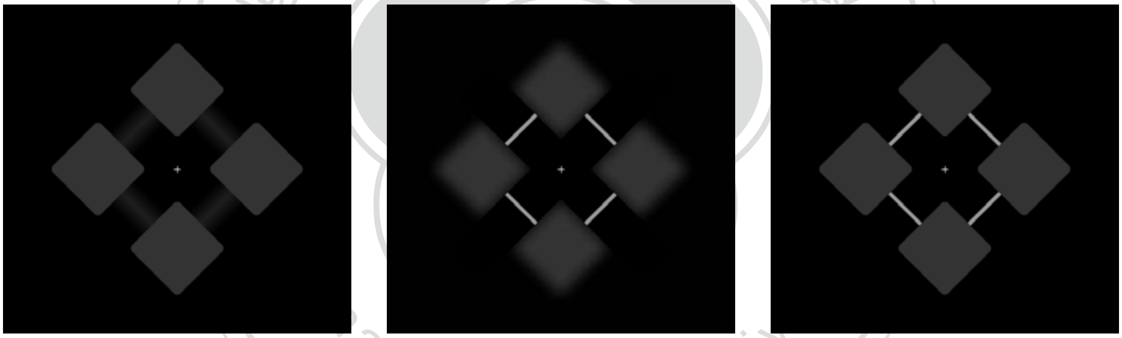
Another 11 participants were recruited with the same standards as described for Experiment 1a.

Design

The experiment is a 2 (attention) × 3 (monocular depth) completely within-participant design. The design is similar to that of Experiment 3, except that the four levels of the independent variable of binocular disparity in Experiment 3 were changed into three levels of monocular depth: blurred-moving-lines, blurred-occluders, and normal condition, which will be described in the Materials section.

Materials

Three types of diamond stimuli were used here, as shown in Figure 16. The normal condition used the same diamond stimulus from Experiment 1a. The blurred-moving-lines condition was made by the normal stimulus but filtered out the high spatial frequency information of the moving lines, making them look as if they were behind the occluders. The blurred-occluders condition was made by using the same method but the high spatial frequency information of the occluders was filtered out, making the occluders look as if they were behind the moving lines. The software Matlab 7 was used in the filtering procedure (using a linear filtering function). The length of each unoccluded moving line of the three monocular depth conditions was controlled to be the same. In addition, the sum of the brightness of each line or each occluder under the three conditions was also controlled to be similar.



(a) (b) (c)

Figure 16. Experimental stimuli used in Experiment 4: (a) is the blurred-moving-lines condition; (b) is the blurred-occluders condition; and (c) is the normal condition.

Apparatus

The apparatus used here was identical to that from Experiment 1a.

Procedures

The procedure was similar to Experiment 1b. Participants had to report their motion perception and respond to the probe that was shown at 3s, 4s, 5s, 6s, 7s, or 8s after the trial began by key-pressing. This experiment had eight blocks—four attending-to-occluders blocks and four attending-to-moving lines blocks. The order of the eight blocks was counterbalanced within participants. Each block had six trials in which the three levels of monocular depth conditions were presented twice, respectively, in a random sequence. In total, each participant completed 6 × 8 = 48 experimental trials.

5.2 Results

Mean percentage of time perceiving coherent motion under the 2 (attention) × 3 (monocular depth) conditions is plotted in Figure 17a. Two-way ANOVA shows that the main effect of attention is significant ($F(1,10) = 11.44$, $p < 0.01$, partial $\eta^2 = 0.534$). The percentage of time perceiving coherent motion is higher in the attending-to-occluders condition (76.36%) than in the attending-to-moving lines condition (60.62%), which is consistent with the hypothesis. The effect of monocular depth is also significant ($F(2,20) = 65.62$, $p < 0.001$, partial $\eta^2 = 0.868$). This is also consistent with the prediction, indicating that the manipulation of monocular depth is valid. The interaction of the two factors is also significant ($F(2,20) = 7.82$, $p < 0.01$, partial $\eta^2 = 0.439$). However, the simple main effects show that the effect of attention is not only significant at the normal condition ($F(1,10) = 9.62$, $p < 0.05$) but also significant at the blurred-occluders condition ($F(1,10) = 10.58$, $p < 0.01$) and marginally significant at the blurred-moving-lines condition ($F(1,10) = 4.20$, $p = 0.068$). These results conflict with the prediction that the effect of attention should disappear under the blurred-moving-lines and blurred-occluders conditions.

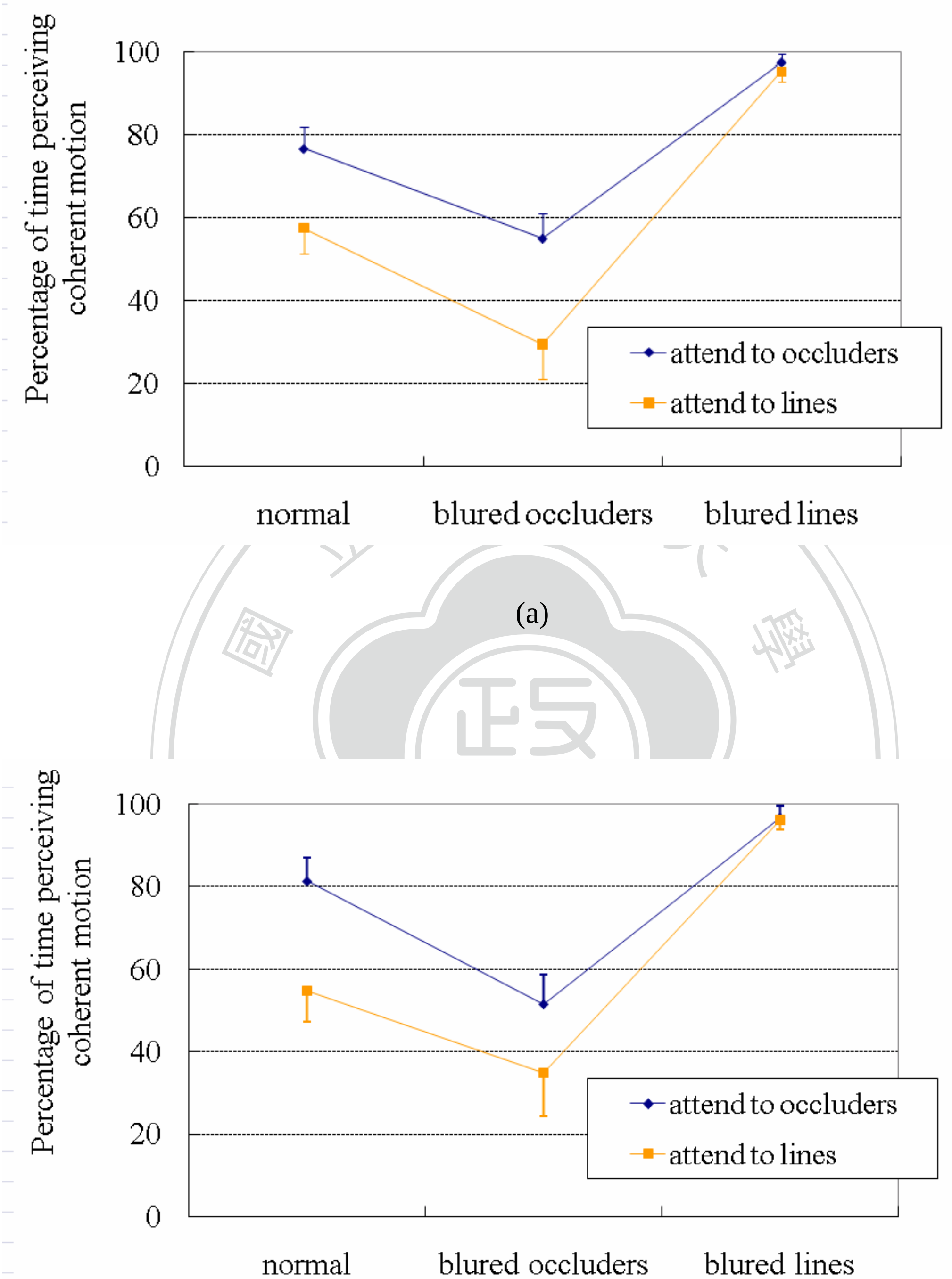
The two-way ANOVA of RT to the probe in eight blocks shows that the interaction is not significant as predicted ($F(1,10) = 0.24$, $p = 0.637$), implying that participants did not fully follow the instructions to allocate their attention. Therefore, only the data in which the participants followed the instructions to allocate their attention were selected for further analysis.

The selection method was the same as in the Experiment 2. The eight blocks were divided into four sections (Section 1: Blocks 1 and 2; Section 2: Blocks 3 and 4, and so on). Each section contained an attending-to-occluders block and an attending-to-moving-lines block to check whether the index of attention (as described in Experiment 1b) in each section is positive. Only the sections in which the index of attention is positive—implying that participants allocated their attention according to the instructions—are selected for further analysis.

There were 21 selected sections (total: 4 section × 11 participants = 44 section; the data of one participant was not all selected), which contained 11 attending-to-occluders-first and attending-to-moving-lines-next sections and 10 attending-to-moving-lines-first and attending-to-occluders-next sections. Thus, the confounding of the order effect would be very small. The mean percentage of time perceiving coherent motion under the 2 (attention) × 3 (monocular depth) conditions is plotted in Figure 17b. Two-way ANOVA shows that the main effect of attention is significant ($F(1,9) = 8.42$, $p < 0.05$, partial $\eta^2 = 0.483$). The percentage of time perceiving coherent motion is higher in the attending-to-occluders condition (76.52%)

than the attending-to-moving-lines condition (61.90%), which is consistent with the hypothesis. The effect of monocular depth is also significant ($F(2,18) = 40.50$, $p < 0.001$, partial $\eta^2 = 0.818$). Tukey's HSD post hoc test ($q.95(3,18) = 3.61$, $HSD = 21.39$) shows that all paired comparisons are significant. The percentage of time perceiving coherent motion is highest in the blurred-moving-lines condition (96.44%) and lowest in the blurred-occluders condition (43.15%), with the value of the normal condition between the two (68.04%). This is also consistent with the prediction, indicating that the manipulation of monocular depth is valid.

The most important aspect is that the interaction of the two factors is significant ($F(2,18) = 3.96$, $p < 0.05$, partial $\eta^2 = 0.306$). The simple main effects show that the effect of attention is only significant at normal conditions ($F(1,9) = 18.22$, $p < 0.01$) and not significant at the blurred-occluders condition ($F(1,9) = 2.23$, $p = 0.169$) or blurred-moving-lines condition ($F(1,9) = 0.82$, $p = 0.390$). These results are consistent with the prediction that the effect of attention should disappear under the blurred-moving-lines and blurred-occluders conditions and that the effect of attention should remain constant under normal conditions.



(a)

(b)

Figure 17. Results of Experiment 4. (a) Means and 1 standard errors of the percentage of time that the diamond stimulus was perceived as coherent motion under the 2

(attention) × 3 (monocular depth) conditions in eight blocks. (b) Means and 1 standard errors of the percentage of time that the diamond stimulus was perceived as coherent motion under the 2 (attention) × 3 (monocular depth) conditions in the selected sections.

5.3 Discussion

The results show that the monocular depth cue (by filtering out HSF information) can eliminate the effect of attention, which is consistent with the prediction in the hypothesis that attention can bias multistable motion perception through affecting the depth perception mechanism. By filtering out the HSF information of the four moving lines, the effect of attention almost completely disappears under the blurred-moving-lines condition (the difference of the percentage of time perceiving coherent motion under the two attention conditions is smaller than 0.70%). This implies that it is difficult for attention to influence multistable motion perception through affecting the depth perception mechanism because the monocular depth cue already defines the depth relationship of the occluders and lines. Since most participants indicated that they can attend to demanded areas even if they look blurred, the reason for the effect of attention disappearing under the blurred-moving-lines condition cannot be explained by participants being unable to attend to them.

Although there is still some effect of attention under the blurred-occluders condition (the difference of the percentage of time perceiving coherent motion under the two attention conditions is about 16.55%), it is not significant (*power* = 0.720, which is estimated by Cohen's (1988) medium effect size *f* = 0.25). Also, the effect of attention is much smaller than under normal conditions (26.59%). This implies that the effect of attention is weakened (but not completely blocked) due to the monocular depth cue. One reason might be that even in the blurred-occluders condition, there were still some monocular depth cues (such as T-junctions) and structures of the diamond stimulus to let the moving lines be perceived as completed behind the occluders, so the occluders cannot be seen as entirely behind the moving lines as expected.

An interesting question is that why the effect of attention showed in the blurred-occluders condition and blurred-moving-lines condition before deleting the negative “index of attention” sections, which should imply that participants did not allocate their attention on demanded areas. One reason may be that participants only attended to one moving line among the four. Accordingly, the “index of attention” in these sections became negative because the probe might show on other lines that they did not attend. This may cause the percentage of time perceiving coherent motion to decrease under the attending-to-lines condition because it is hard to perceive the four lines connecting to each other and moving coherently if only one line is attended. And this decrease may be the reason why the effect of attention showed in the blurred-occluders condition and blurred-moving-lines condition before deleting the negative “index of attention” sections.

6. General Discussion

In the multistable perception domain, many studies have found that directing fixation or spatial attention to different regions can bias the perception of the Necker cube (Toppino, 2003; Kawabata, 1986, 1987; Meng & Tong, 2004; Xu & Franconeri, 2010). The explanation in these studies is that the fixated or attended region would look closer, so the perception is biased. However, no study has examined this assumption directly. On the other hand, in research concerned with attention, few studies have mentioned whether spatial attention can affect depth perception. Some studies have found that spatial attention can influence many properties of stimulus perception, which may relate to depth perception (e.g., Carrasco, 2006; Carrasco, Williams, & Yeshurun, 2002; Gobell & Carrasco, 2005). Thus, it is possible for spatial attention to affect multistable perception through the depth mechanism.

The purpose of this study is to investigate the effect of spatial attention on multistable figure perception and the possible underlying mechanism of this effect. The diamond stimulus was used to test the hypothesis that spatial attention can bias multistable motion perception by making attended areas look closer. Four experiments are designed to examine this hypothesis and the main results will be briefly reviewed first.

Spatial attention can bias multistable perception

The effect of spatial attention was found to be very robust at normal conditions in Experiments 1a, 1b, 3, and 4, making the percentage of time perceiving coherent motion significantly higher in the attending-to-occluders condition than that in the attending-to-moving-lines condition. This result shows that spatial attention, like the effect of fixation, can bias multistable perception. However, this result still cannot directly support that spatial attention biases multistable motion perception by affecting depth perception. Accordingly, in Experiment 2, this question was verified directly.

Spatial attention can alter perceived depth slightly

Experiment 2 was a depth-judgment task. The same two attention conditions and nine different levels of binocular disparity between the occluders and the moving lines were manipulated. Results show that the mean probability of reporting lines behind occluders in small binocular disparities is significantly higher under the attending-to-occluders condition than that under the attending-to-lines condition. This result is consistent with the hypothesis that attention can affect depth perception, making attended areas look closer. Thus, it is more likely that the effect of spatial attention on multistable motion perception found in Experiments 1a and 1b occurs through the mechanism of making the attended areas look closer in depth.

Binocular and monocular depth cues can weaken the effect of spatial attention

Experiments 3 and 4 were designed to test whether the effect of spatial attention on biasing multistable motion perception can be blocked with binocular or monocular depth cues that define the depth relationship between occluders and moving lines. If the effect of spatial attention can be affected by the depth cues, it is more possible that attention is biasing multistable motion perception through depth mechanisms. The results of Experiment 3 show that the effect of spatial attention is only significant at binocular disparity 0 (occluders and lines were in the same depth) but not significant at the other three binocular disparity conditions (lines behind the occluders in three different levels). The results of Experiment 4 show that the effect of spatial attention is blocked when the monocular depth cues (by filtering out high spatial frequency information) define the occluders in front of the lines. In summary, these results are all consistent with the prediction of the hypothesis that spatial attention is biasing multistable motion perception through the depth perception mechanism. Further discussions will follow, including the integration of the results of this study and other relative studies as well as possible mechanisms of spatial attention on multistable perception.

Spatial attention versus intention

The relationship between the mechanisms of spatial attention and intention on multistable perception is an interesting issue worth investigating. Although these two mechanisms are all top-down processing under our cognitive control, they seem to act in different ways. The results of Experiment 1a show that the effects of spatial attention and intention on multistable motion perception are additive, implicating they are independent mechanisms. In addition, one of my other experiments, not included in this study, investigated whether participants would allocate their spatial attention on occluders or lines in order to maintain demanded perceptions. Intention was manipulated in a between-participant design, as in Experiment 1a, and the accuracy of continuous probe detection of occluder lightening or line lightening was measured. The results show that the difference of detecting occluder lightening under the two intention conditions is not significant ($F(1,10) = 0.117$, $p = 0.740$). Also, the difference of detecting line lightening under the two intention conditions was not significant ($F(1,10) = 0.177$, $p = 0.683$). On the other hand, the percentage of time perceiving coherent motion in "hold coherent" condition (63.29 %) was higher than that in the "hold separate" condition (37.07 %) ($F(1,10) = 8.23$, $p < 0.05$, partial $\eta^2 = 0.457$), which is consistent with the intention control effect found in Experiment 1a. These results imply that participants did not need to attend to particular areas to exert their intentional control. Instead, intentional control is more likely to be exerted through top-down priming of wanted representation, as Toppino (2003) mentions. However, in Experiment 1 of the aforementioned Tsal and Kolbet's study (1985), they found that observers tend to attend to the focal area of the interpretation that they had to maintain. The reason may be the short presentation time of the multistable figure (30 milliseconds). Directing their spatial attention to the focal area would be a more effective and faster way to perceive the demanded perception than just top-down priming of wanted representation.

Spatial attention versus depth cues (top-down vs. bottom-up mechanism)

In the four experiments of this study, the mechanisms of spatial attention and intention on multistable perception belongs to top-down processes, while binocular and monocular depth cues belong to bottom-up factors. The entire theoretical framework of the study is presented at Figure 18. Since all these factors are found having influence on multistable perception in this study, it consists with the hybrid model mentioned previously, which assumes both bottom-up and top-down processes together contribute to the multistable figure perception.

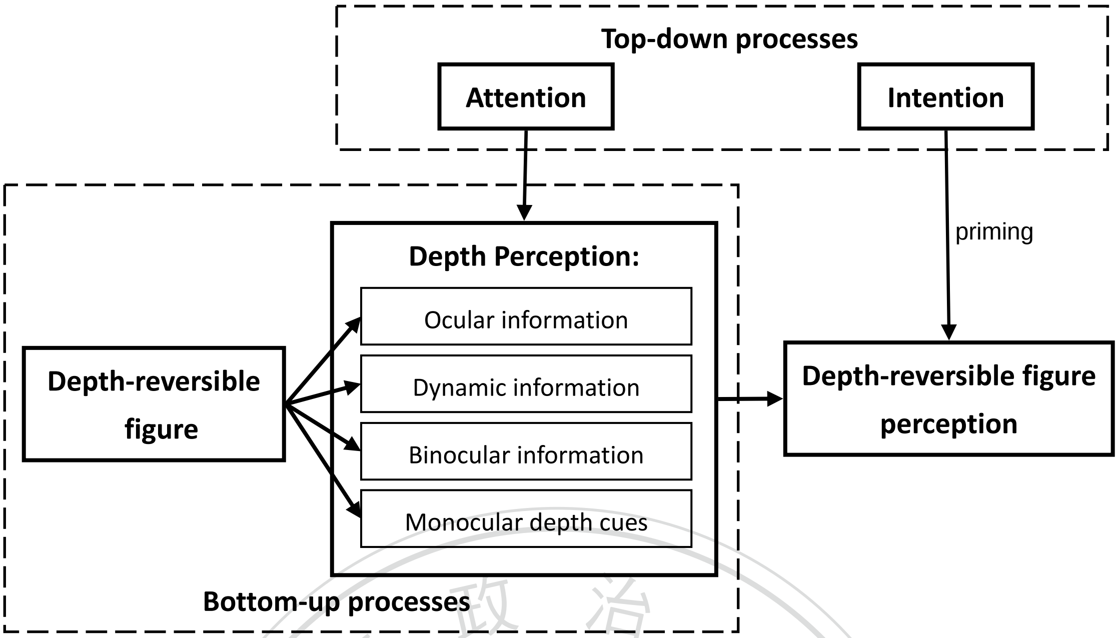


Figure 18. Theoretical framework of the study. According to Palmer (1999), depth perception is formed from integrating four kinds of depth information: ocular information (including accommodation and convergence), dynamic information (such as motion parallax and optical flow), binocular information (also called stereoscopic information), and monocular depth cues (also called pictorial information). The former two types of information are not discussed in this study. Both bottom-up processing (depth perception) and top-down processing (intention and attention) contribute to depth-reversible figure perception.

An important issue worth investigating is how these top-down and bottom-up factors act together on multistable perception. It seems that spatial attention can influence depth perception only when the depth cues of stimulus are not very clear.

For example, in Experiment 2, spatial attention has influence on perception—making attended areas look closer—only when the manipulated binocular disparities are small. In Experiment 3, the effect of spatial attention shows only at the binocular-disparity-0 condition, not at other large-binocular-disparity conditions. In Experiment 4, the manipulated monocular depth cues block the effect of spatial attention. These pieces of evidence all imply that spatial attention can exert its affection only when the depth cues of stimulus (such as the depth relationship of the occluders and the lines in the diamond stimulus) are not well-defined. A possible reason is that bottom-up depth cues are more dominant than attention in affecting depth perception. Thus, the effect of spatial attention is easily overridden by bottom-up depth cues.

Mechanisms of spatial attention on multistable perception

The mechanism of spatial attention on multistable perception is complicated. Some parts of the effect of spatial attention are like the effect of fixation found previously (e.g., Meng & Tong, 2004; Toppino, 2003). For example, the results of Experiment 1a show that the effect of attention is comparatively smaller than the effect of intention. In addition, it has no interaction with the effect of intention, implying that the effect of spatial attention and intention are independent mechanisms in multistable perception. This part of the effect of spatial attention may be involved in affecting bottom-up processing of monocular depth cues, perhaps by increasing perceived contrast, spatial frequency, and the spatial resolution of attended areas (e.g., Carrasco, Williams, & Yeshurun, 2002; Carrasco, 2006; Gobell & Carrasco, 2005). Through this, the attended areas look closer and affect multistable perception. Another piece of evidence supporting this notion is that monocular depth cues can almost block the effect of spatial attention in Experiment 4, implying that attention may act through this way. To summarize briefly, the effect of spatial attention on multistable perception is because it affects its bottom-up processing (depth perception), as shown in Figure 18.

However, it is still unclear whether there are other ways for attention to affect multistable perception through the depth perception mechanism. For example, attention may affect the integration of different sources of depth information, or attention itself may be a source of depth information. The main question comes from the uncertainty of how different sources of depth information are combined into a single coherent representation. For example, it is not clear whether different sources are independent processes, whether they interact, or whether their integration is additive or multiplicative (Palmer, 1999). Further studies are needed to discover a clearer and complete role of spatial attention in the depth perception mechanism.

Limitations and restrictions

One limitation in the manipulation-check task of spatial attention in Experiments 1b, 2, 3, and 4 is that the probe showed in every trial. Thus, it seems possible for participants to give the probe-detecting response in every trial, even though they did not see the probe. However, there are three reasons that can demolish this possibility. First, there is no trial of which the RT to the probe is shorter than 400 ms, and trials of which the RT to the probe are longer than 1500 ms are very rare. It implies that participants responded to the probe due to seeing it. The second reason is that the showing of the probe is so obvious that it is hard to create a false-alarm response or miss it. The third reason is that by the observation of the experimenter, participants did not select the probe-detecting response randomly or repetitively in each trial. Hence, for these reasons, it is unlikely the participants downhill choose the probe-detecting response when not seeing the probe.

Another limitation of this study is that nearly half the data are deleted due to the negative “index of attention” in Experiments 2, 3, and 4. The large abandoned data may affect the whole pattern of results and the interpretations of it. However, it should be noted that it is not easy to sustain spatial attention on four (rather than one) occluders or four moving lines for a period of time, not to mention that participants also have to hold coherent motion perception and report their perception at the same time in Experiments 3 and 4. It is necessary to use a strict criterion—the index of attention—to rule out the data of which participants did not allocate their attention on demanded areas. Thus, it is more convincing that the retained data were manipulated validly.

In summary, all four experiments in this study support the notion that spatial attention can bias multistable motion perception by affecting depth perception, making attended areas look closer. However, this notion is restricted only to depth-reversible figure perception. The effect of spatial attention on other multistable figure perception may involve mechanisms other than depth perception.

Contribution and future studies

How attention can influence or mediate brain processing of visual stimuli and the possible underlying mechanism driving it is an important issue that interests many researchers. In the multistable perception domain, the results of this study support the assumption that spatial attention can bias depth-reversible figure perception by making attended areas look closer. This assumption had been proposed in many studies of the Necker cube, but it is examined directly in this study. This study is also contributive in understanding how top-down and bottom-up factors act together on multistable perception. In future studies, an issue worth investigating is that whether the effect of spatial attention on depth-reversible figure perception is only via the depth mechanism or if there are other ways attention can influence depth-reversible figure perception. In addition, the effect of spatial attention on the multistable figures other than the depth-reversible figure is also an interesting topic. The effect of spatial attention and the underlying mechanisms of them can be compared with depth-reversible figures.

In the depth perception domain, few studies have mentioned whether or how attention affects depth perception, which is one of the most important pieces of information with respect to vision. This study also provides some evidence that spatial attention can affect depth perception by means of the direct, subjective depth-judgment task (Experiment 2) and the indirect multistable-perception task (Experiment 1, 3, and 4). In future studies, the degree to which spatial attention can affect depth perception should be examined more precisely by using stimuli other than multistable figures. Also, it is also important to clarify the relationship between spatial attention and other depth cues and how they are integrated into a final depth perception result in further studies.

7. References

Anton-Erxleben, K., Henrich, C., & Treue, S. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision, 7*(11), 1-9.

Bisiach, E., Ricci, R., Lai, E., De Tanti, A., & Inzaghi, M. (1999). Unilateral neglect and disambiguation of the Necker cube. *Brain, 122*(1), 131-140.

Carrasco, M. (2006). Covert attention increases contrast sensitivity: Psychophysical, neurophysiological, and neuroimaging studies. *Progress in Brain Research, 154*, 33-70.

Carrasco, M., Fuller, S., & Ling, S. (2008). Transient attention does increase perceived contrast of suprathreshold stimuli: A reply to Prinzmetal, Long, and Leonhardt (2008). *Perception & Psychophysics, 70*(7), 1151-1164.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*(3), 308-313.

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision, 2*(6), 467-479.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Science* (2nd ed.): Hillsdale, NJ: Lawrence Erlbaum Associates.

Fuller, S., & Carrasco, M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision Research, 46*(23), 4032-4047.

Fuller, S., Ling, S., & Carrasco, M. (2004). Attention increases perceived saturation. *Journal of Vision, 4*(8), 329-329.

Garcia-Perez, M. A. (1992). Eye movements and perceptual multistability. In E. Chekaluk & K. Llewellyn (Eds.), *The Role of Eye Movements in Perceptual Processes* (pp. 73-110): Elsevier Science Publishers B.V.

Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of apatial frequency and gap size. *Psychological Science, 16*(8), 644-651.

Hopfinger, J. B., Luck, S. J., & Hillyard, S. A. (2004). Selective Attention: Electrophysiololgical and neuromagnetic studies. In M. Gazzaniga (Ed.), *The cognitive neurosciences III*. Cambridge,MA: MIT Press.

Inui, T., Tanaka, S., Okada, T., Nishizawa, S., Katayama, M., & Konishi, J. (2000). Neural substrates for depth perception of the Necker cube; a functional magnetic resonance imaging study in human subjects. *Neuroscience letters, 282*(3), 145-148.

Kawabata, N. (1986). Attention and depth perception. *Perception, 15*(5), 563-572.

Kawabata, N. (1987). Interpretive process of depth in line drawing. *Systems and*

Computers in Japan, 18(7), 103-109.

Kawabata, N., Yamagami, K., & Noakl, M. (1978). Visual fixation points and depth perception. *Vision Research, 18*(7), 853-854.

Klymenko, V., & Weisstein, N. (1986). Spatial frequency differences can determine figure-ground organization. *Journal of Experimental Psychology: Human Perception and Performance, 12*(3), 324-330.

Kornmeier, J., & Bach, M. (2005). The Necker cube--an ambiguous figure disambiguated in early visual processing. *Vision Research, 45*(8), 955-960.

Kornmeier, J., Hein, C. M., & Bach, M. (2009). Multistable perception: When bottom-up and top-down coincide. *Brain and Cognition, 69*(1), 138-147.

Kuo, C.-F., & Huang, S.-L. (2005). *The formation of stereoscopic surfaces as occluders can affect the motion interpretation.* Paper presented at the Taiwanese psychology association 44th annual conference.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences, 3*(7), 254-264.

Ling, S., Carrasco, M., Lipson, Z., Roche, R., Little, A., & Jones, B. (2007). Transient covert attention does alter appearance: A reply to Schneider (2006). *Perception and psychophysics, 69*(6), 1051-1058.

Liu, T., Abrams, J., & Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychological science: a journal of the American Psychological Society/APS, 20*(3), 354-362.

Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychological Bulletin, 130*(5), 748-768.

Long, G. M., Toppino, T. C., & Mondin, G. (1992). Prime time: Fatigue and set effects in the perception of reversible figures. *Perception and psychophysics, 52*, 609-609.

Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research, 32*(2), 263-273.

McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, *19*(1), 431-441.

McDermott, J., Weiss, Y., & Adelson, E. (2001). Beyond junctions: nonlocal form constraints on motion interpretation. *Perception, 30*(8), 905-924.

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision, 4*(7), 539-551.

Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity.

Vision Research, 49(7), 735-745.

Palmer, S. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA.: MIT press

Peterson, M. A., & Gibson, B. S. (1991). Directing spatial attention within an object: Altering the functional equivalence of shape description. *Journal of Experimental Psychology: Human Perception and Performance. Vol 17(1), Feb 1991, 170-182., 17*(1), 170-182.

Pitts, M. A., Gavin, W. J., & Nerger, J. L. (2008). Early top-down influences on bistable perception revealed by event-related potentials. *Brain and Cognition, 67*(1), 11-24.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 32*(3), 3-25.

Read, R., Ling, R., & Carrasco, M. (2003). Covert attention alters visual appearance. *Journal of Vision, 3*(9), 329-329.

Slotnick, S. D., & Yantis, S. (2005). Common neural substrates for the control and effects of visual attention and perceptual bistability. *Cognitive Brain Research, 24*(1), 97-108.

Suzuki, S., & Peterson, M. A. (2000). Multiplicative effects of intention on the perception of bistable apparent motion. *Psychological Science, 11*(3), 202-209.

Theeuwes, J. (1992). Visual selective attention: A theoretical analysis. *Acta Psychologica, 83,* 93-154.

Tootell, R., Hadjikhani, N., Hall, E., Marrett, S., Vanduffel, W., Vaughan, J. et al. (1998). The retinotopy of visual spatial attention. *Neuron, 21,* 1409-1422.

Toppino, T. C. (2003). Reversible-figure perception: Mechanisms of intentional control. *Perception & Psychophysics, 65*(8), 1285-1295.

Toppino, T. C., & Long, G. M. (1987). Selective adaptation with reversible figures: Don't change that channel. *Perception & Psychophysics, 42*(1), 37-48.

Toppino, T. C., & Long, G. (2005). Top-down and bottom-up processes in the perception of reversible figures: toward a hybrid model *Dynamic Cognitive Processes* (pp. 37-58): Springer Tokyo.

Treue, S. (2004). Perceptual enhancement of contrast by attention. *Trends in Cognitive Sciences, 8*(10), 435-437.

Tsal, Y., & Kolbet, L. (1985). Disambiguating ambiguous figures by selective attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 37*(1), 25-37.

Vecera, S. P., Flevaris, A. V., & Filapek, J. C. (2003). Exogenous spatial attention influences figure-ground assignment. *Journal of Vision, 3*(9), 763-763.

Vecera, S. P., Flevaris, A. V., & Filapek, J. C. (2004). Exogenous spatial attention influences figure-ground assignment. *Psychological Science, 15*(1), 20-26.

Xu, Y., & Franconeri, S. (2010). *Change of object structure as a result of shifts of spatial attention*. Paper presented at the Vision Science Society 10th Annual Meeting.

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research, 39*(2), 293-306.


國立政治大學
National Chengchi University

8. Acknowledgment

First, I would like to thank my advisor Dr. Haung. After failing so many times in this study, you still encourage me and ask me to try again. With out your encouragements and support, I would already have given up. I also appreciate my oral committee members, Dr. Su-Ling Yeh and Dr. Man-Ying Wang. Thanks for pointing out the mistakes that I had not noticed, and your great comments are really important to improve this study.

Next, I want to thank all members of the Perception Lab of Chengchi University. Thanks for all your helps in this study. In particular, I would like to thank Yu-Chieh Chang and Hung-Wei Lee for your excellent suggestions.

I also want to thank all participants in this study. Thanks for your attendance, feedbacks, and "attention". I'm really happy to do experiments with you. Last but not least, I would like to thank my family for providing me a well environment and always trying to prompt me.

I would not get through with this study without any of you. Thank you again!