Emergence of human cognition from spatial and temporal structures

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Zhicheng Lin

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Professors Sheng He and Daniel J. Kersten

September 2012

Acknowledgements

Although maybe not reflected in the thesis, I must admit that I spent lots of energy chasing bad ideas, producing half-completed experiments, and writing unpublishable manuscripts. Such is perhaps the reality of every thesis, but still I love doing what I have been doing, unable to resist the thrill of new ideas and findings. Looking back, I have learned to think for myself, take steps backward, and design informative experiments. For all these and others, I thank my primary advisor Sheng He and committee members Dan Kersten, Steve Engel, and Bin He. I also benefited from discussions with people in the vision labs, especially Peng Zhang and Min Bao.

Finally I thank my wife Lifang for making my life better and reminding me of the fun in living.

Thanks to the University of Minnesota for financial support.

Minneapolis, Minnesota Zhicheng Lin

To my parents

Abstract

Like solving a jigsaw puzzle by fitting pieces together bit by bit, the human mind makes sense of the world by transforming inputs into outputs throughout the nervous system, with the emergence of visual descriptions during such transformations relying heavily on perceptual integration. This thesis distinguishes emergent perceptual integration from non-emergent perceptual integration, with the former but not the latter resulting in visual descriptions unique to the whole and cannot be found in the parts; this distinction provides clues to the nature of unconscious processing (Chapter 1). Indeed, emergent properties, elusive as they sound, are tractable as revealed by emergent filling-in and its unique adaptation aftereffect (Chapter 2).

To further understand the principles that constrain the emergence of visual descriptions, the thesis considers how spatial and temporal structures, two main types of contextual effects, regulate the emergence process. Rich contextual regularities are embedded in spatial configurations; a moving frame technique reveals that the coupling of objects to the contextual frames is pervasive and relatively automatic (Chapter 3), resulting in automatic deployments of exogenous visual attention to non-retinotopic, frame-centered locations (Chapter 4). Statistical regularities also abound in the vast associative knowledge one learns in the life time. Indeed, associated links can be established rapidly and without rewards, resulting in subsequent inventory attentional capture to the associated colors when the location of the target is uncertain (Chapter 5), and subsequent attentional inhibition of the associated colors when the location of the target is fixed (Chapter 6). These results thus reveal the importance in distinguishing emergent from non-emergent perceptual integration, and show that the emergence of visual descriptions is strongly constrained by spatial and temporal structures.

Table of Contents

List of Tables vi
List of Figures vii
1 Introduction and Synthesis: Emergence of human cognition from spatial and temporal structures 1
1.1 Perceptual integration in the emergence of visual description: emergent vs. non-emergent properties 3
1.2 Unconscious binding: emergent vs. non-emergent perceptual integration 8
1.3 Spatial configuration and the emergence of visual description 12
1.4 Associative knowledge and the emergence of visual description 16
1.5 Concluding remarks 17
2 Emergent filling-in induced by motion integration reveals a high level mechanism in filling-in 19
2.1 Introduction 19
2.2 Method 23
2.3 Results 25
2.4 Discussion 30
3 Frame-centered object representation and integration revealed by non-retinotopic priming and masking 34
3.1 Introduction 34
3.2 Methods 37
3.3 Results 39
3.4 Discussion 52
4 Frame-centered object representation supports flexible exogenous visual attention across translation and reflection 58
4.1 Introduction 58
4.2 Methods 60
4.3 Results 63
4.4 Discussion 70
5 Association-driven attentional capture 73
5.1 Introduction 73

5.2 Methods 75
5.3 Results 78
5.4 Discussion 84
6 Visual familiarity induces inhibition through attentional inertia 88
6.1 Introduction 88
6.2 Methods 91
6.3 Results 93
6.4 Discussion 99
7 References 102

List of Tables

Table 2-1. Magnitude of motion aftereffect as measured by the difference in the percentage of rightward responses between the leftward and rightward adaptation 30

Table 6-1. Mean reaction time (in milliseconds) and error rate (in %) in the test phase (the inside target shape and the outside distractor shape were presented simultaneously; the target color and the distractor color were the familiar color and the unfamiliar color, respectively, or the reverse, randomized across trials) of Experiment 1 (same shapes across familiarization and test) and Experiment 2 (different shapes across familiarization and test) 95

Table 6-2. Mean reaction time (in milliseconds) and error rate (in %) in Experiment 3 (either the inside shape or the outside shape was presented, not both, the color of which was either the familiar or the unfamiliar color). 99

List of Figures

Figure 1-1. Emergence of visual description from perceptual grouping and during filling-in (perceptual completion) 7
Figure 1-2. A critical role of awareness in emergent perceptual integration but not in non-emergent perceptual integration 11
Figure 1-3. Spatial configuration drives the emergence of visual description 13
Figure 1-4. Robust and automatic coding of objects to the contextual frames as revealed by frame-centered priming and masking effects 15
Figure 2-1. Filling-in (perceptual completion) illustrated 22
Figure 2-2. Results from the passive adaptation experiments 26
Figure 2-S1. Contributions of both contour filling-in and surface filling-in in the emergent filling-in motion aftereffect 27
Figure 2-S2. Does a balanced configuration (i.e. two up and two down motions across the four apertures, as in the integrated condition) produce a similar adaptation effect as a sequentially balanced configuration (i.e. one up and three down motions sequentially alternate with three up and one down motions across the four apertures, as in the nonintegrated condition)? 28
Figure 3-1. Procedure, stimuli, and results from Experiment 1 41
Figure 3-2. Stimuli and results from Experiment 2 43
Figure 3-3. Stimuli and results from Experiment 3 45
Figure 3-4. Procedure and results from Experiment 4A and 4B 47
Figure 3-5. Procedure, stimuli, and results from Experiment 5 49
Figure 3-6. Procedure and results from Experiment 6 and 7 51
Figure 4-1. Flexible exogenous attention across translation 64
Figure 4-2. Flexible exogenous attention across translation through an implicit frame 66
Figure 4-3. Flexible exogenous attention across mirror reflection 69
Figure 5-1. Attentional capture by uninformative cues after associative learning 79
Figure 5-2. Attentional capture by invalid cues 81

Figure 5-3. Attentional capture by uninformative color cues even when attention is guided by a clear top-down goal (search-for-a-square) and its generalization to a different task 83
Figure 6-1. Familiarity induces inhibition in visual attention through attentional inertia 96

Chapter 1

Introduction and Synthesis: Emergence of human cognition from spatial and temporal structures[1]

Human cognition is complex, precluding a straightforward understanding by a single equation or theory. At least in vision, it is now well recognized that each problem needs explanations at multiple levels, such as the levels of computation (what and why), algorithm and representation (how), and physical implementation (Marr, 1982). These levels of analysis, though constituting the whole of the system, can be relatively independent, and the choice of the most useful level of analysis usually depends on the problem at hand—so that, for instance, accordingly to Marr, afterimages are best explained by the physical implementation of vision, whereas the bistable Necker cube illusion can be attributed to the shifting of representations. Such an information processing perspective of human cognition emphasizes the goal of vision as providing a useful description for the task at hand (Marr, 1976) but shies away from the phenomenal aspect of human cognition. This perspective in no doubt has fueled research in the past several decades, but it leaves unanswered an important question: Exactly how does the visual system arrive at the useful description, and in doing so, how do phenomenal experiences, so central to and characteristics of human condition, emerge? Marr touches upon the first issue—emergence of description—in shape recognition (from primal sketch, to 2½D sketch, to 3-D representation), but accumulating research indicates that this issue is a general one in human cognition, and the emergence of visual description is tightly connected to the emergence of conscious awareness, with subjective interpretation of visual inputs based on structural regularities being a cornerstone of human experiences. This view is a central theme in this dissertation. In this general introduction and conceptual synthesis, we begin with a general argument that emergence of visual

[1] Parts of this chapter were published as Lin, Z, & He, S. (2009). Seeing the invisible: the scope and limits of unconscious processing in binocular rivalry. *Progress in Neurobiology*.

descriptions is a general computation in human cognition and that emergent visual description found in the whole but not in the parts may be a strong marker of visual consciousness (see **Glossary**). We then outline how spatial and temporal contexts shape the emergence of visual descriptions.

Glossary

Associative learning: the process by which an association between two events, such as stimuli and behaviors, is learned.

Binding: the dynamic integration of multiple elements leading to the perception of a coherent, unified whole, such as an object, a surface, and a scene.

Binocular rivalry: when conflicting images are presented to each of the two eyes, the two images rival for visibility, with one temporarily dominating perception only to be replaced in dominance by the other in turn.

Continuous flash suppression (CFS): when a series of different, contour-rich, high-contrast patterns are continuously flashed to one eye at about 10 Hz (i.e. 100 ms per frame), these rich patterns usually can suppress information presented to the other eye for a relatively long viewing period, sometimes longer than 3 minutes.

Filling in/perceptual completion: the various phenomena in which the visual system recovers a coherent percept from incomplete, fragmented pieces.

Amodal completion: a type of perceptual completion in which the completed part is occluded and thus lacks visible attributes.

Modal completion: a type of perceptual completion in which the completed part shares the same "mode" (e.g. brightness) as the rest of the figure.

Emergent: properties of a whole that are unique to the whole and cannot be found in the parts.

Adaptation aftereffects: prolonged exposure to a visual stimulus (i.e. adaptor) alters the visual system's sensitivity to, or the appearance of a subsequent related stimulus (i.e. test), with the altered appearance called visual aftereffect.

Predictive coding: in digital image processing, to increase efficiency, information is usually coded by reducing the signal amplitude through the removal of predictable and thus redundant components; in neural processing, to exploit stimulus correlation across space and time, neurons are said to code the difference between the predicted and actual values rather than the raw values, so that neural activity is higher when a stimulus is not well-predicted or explained by a generative neural model.

Priming: the facilitation effect on processing of information from previously presented information.

Masking: the detrimental effect on processing of information from previously (forward masking) or subsequently (backward masking) presented information.

1.1 Perceptual integration in the emergence of visual description: emergent vs. non-emergent properties

For a given computational unit, be it a neuron, a brain area, or the human brain on the whole, an input transforms into an output. How does the output emerge from the input? This issue of emergence represents a ubiqutous problem in human cognition—considering how the visual system arrives at vivid, rich percepts from impoverished lights impinging on the retina, one cannot but wonder what is the chain of computations working the magic, just like how the same, simple molecules transiting among distinct solid, liquid, and gaseous states. Understanding the mechanisms of emergence is central to understanding human cognition, including how the representation in the cortical hierarchy becomes more and more invariant to a host of factors such as variations in retinal position, scale, viewpoint, contrast, and lighting (DiCarlo & Cox, 2007; Graf, 2006). One can approach the problem of emergence from different scales ranging from biochemistry to whole system behaviors, but a simple distinction can be made by considering whether the emergence involves *emergent* properties, present in the computing whole but absent from the parts (McClelland, 2010). This perspective thus assigns a paramount role in human cognition to perceptual grouping and integration, determining whether emergence yields emergent or non- emergent properties. Consider the two-tone, so-called Mooney images in Figure 1-1A. When the percept of a dog or a man emerges from fragmented parts in the images, an impression of depth and specific form crystalizes because of a change in perceptual organization. The percept of a dog or a man in these examples is said to be emergent because it emerges from the integration of the parts yet cannot be explained by the parts alone, just like the properties of water are emergent because they emerge from the particular configuration that oxygen and hydrogen enter into when combined into molecules and from the particular interactions among these and other kinds of molecules, yet cannot be explained by either oxygen or hydrogen alone, nor by their additive or subtractive effects. Indeed, the effect is so dramatic and compelling that once the visual system recognizes the objects in the images, this recognition irresistibly and irreversibly alters our phenomenal experience of the

images—for instance, it is now hard not to perceive the dog and the face as the figures of the images (Irvin Rock, 1984).

Although emergent properties may seem elusive, we show in Chapter 2 that they are tractable and even can provide unique insights into our understanding of human cognition. Specifically, we studied perceptual completion or filling-in (Figure 1-1B); in our configuration (Figure 1-1C), integration of the local motion signals (uphill or downhill motion) generates a distinct filling-in motion percept (horizontal motion)—an emergent property, which we call emergent filling-in. Although it is debated whether filling-in is achieved through pointwise neural representation of visual features in retinotopic visual areas (Gerrits & Vendrik, 1970), or whether contour and surface interpolations take place at higher areas, based on the contrast information at the surface border (Gregory, 1972), emergent filling-in of motion highlights the synergetic interactions between local isomorphic processes (e.g. contour interpolation) and global symbolic processes (e.g. structure and global motion extraction). To study the function of this emergent motion property, we used motion aftereffect (MAE), wherein prolonged exposure to a moving stimulus renders a subsequent stationary test pattern to appear to move in the opposite direction (Addams, 1834; Mather, Verstraten, & Anstis, 1998). Even though there is no physical stimulation within the filling-in region in between the motion apertures, we observed a much stronger MAE in the integrated (with filling-in) condition than the nonintegrated (without filling-in) condition. The difference between the two conditions therefore reveals a MAE not confounded by transfers from nearby apertures (Bex, Metha, & Makous, 1999; Lalanne & Lorenceau, 2006; X. Meng, Mazzoni, & Qian, 2006; Price, Greenwood, & Ibbotson, 2004; Snowden & Milne, 1997; Weisstein, Maguire, & Berbaum, 1977), but due solely to adaptation to the emergent, perceived filling-in motion (Figure 1-1D). This filling-in MAE was observed in both the modal and amodal conditions, but, importantly, was reduced when the attention load of a fixation task was increased during adaptation, suggesting that the more symbolic processes in later areas might provide important feedback signals to guide more isomorphic process in earlier areas (Halgren, Mendola, Chong, & Dale, 2003), which critically depends on attention. This study thus provides a method for isolating the effect of emergent properties, which can quantify the functional significance by measuring

adaptation effect. The results highlight the importance of high-level interpolation processes in filling-in, and are consistent with the idea that, during emergent filling-in, the more cognitive/symbolic processes in later areas provide important feedback signals to guide more isomorphic process in earlier areas.

More generally, the emergence of visual description, be it an emergent property or not, marks a transition between two states, a state of unintegrated local elements ("A") and a state of the integrated global whole ("B"). This implies that when the brain is in state A, the neural correlates of state A (i.e., brain areas sensitive to local elements) rule, whereas when the brain is in state B, the neural correlates of state B (i.e., brain areas sensitive to the integrated whole) rule. This state transition principle applies to all the emergence of visual description, although it can be more pronounced in emergent perceptual integration. The principle is inspired by observations in bistable phenomena such as binocular rivalry; for instance, competition of a face image in one eye and a house image in the other eye results in alternative activations in face sensitive areas (e.g., fusiform face area, FFA) when the face image is perceived and activations in house sensitive areas (e.g., parahippocampal place area, PPA) when the house image is perceived (Tong, Nakayama, Vaughan, & Kanwisher, 1998). The transition principle leaves behavioral footprints as well, consistent with studies showing that perceptual integration of line features into emergent facial structures (state B) impairs visual search for line features (state A) (Suzuki & Cavanagh, 1995), that integrating multiple stimuli into parts of a single patch (state B) degrade speed discrimination of the individual part (state A) (Verghese & Stone, 1996), and that shape combinations embedded within a larger configuration (state B) are harder to learn and remember than when they stand alone (state A) (Fiser & Aslin, 2005). This transition principle is straightforward, yet explains previously seemingly conflicting findings, which are usually explained by evoking accounts such as predictive coding (Fang, Kersten, & Murray, 2008; D. He, Kersten, & Fang, 2012; S. O. Murray, Kersten, Olshausen, Schrater, & Woods, 2002; S. O. Murray, Schrater, & Kersten, 2004). For instance, in a bi-stable diamond illusion similar to Figure 1-1C, four lines in four apertures are sometimes integrated into an emergent, coherent moving diamond shape (state B) but sometimes are not integrated, appearing instead as disconnected, unrelated random lines (state A) (Lorenceau &

Shiffrar, 1992). When a coherent diamond is perceived, activity in the LOC increases but activity in V1 decreases, and vice versa (Fang et al., 2008); similarly, the integrated condition results in a stronger behavioral shape aftereffect but a weaker line aftereffect, and vice versa (D. He et al., 2012). To the extent that LOC is more sensitive to state B (coherent shape) than state A (fragmented lines) and V1 is more sensitive to state A than state B, as is well documented (Grill-Spector, Kourtzi, & Kanwisher, 2001), the state-dependent pattern is well explained and predicted, without having to invoke predictive coding, such as the proposal that higher areas attempt to explain away activities in lower areas through feedback projections (Fang et al., 2008; D. He et al., 2012; S. O. Murray et al., 2002; S. O. Murray et al., 2004). Indeed, "explaining away" proposals run into conflict with both human and monkey studies showing that when random short lines (state A) are grouped into collinear patterns such as long lines and shapes (state B), activity in lower areas such as V1 actually increases, as in higher areas such as LOC (Altmann, Bulthoff, & Kourtzi, 2003; Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003; W. Li, Piech, & Gilbert, 2008). However, this pattern in V1 makes sense according to the transition principle because V1 is more sensitive to state B (collinear lines and shapes) than state A (fragmented noisy short lines), with the intrinsic horizontal connections in V1 adept at collinear contour integration (W. Li & Gilbert, 2002; Z. Li, 1998). Thus, states rather than higher or lower areas per se matter in perceptual integration.

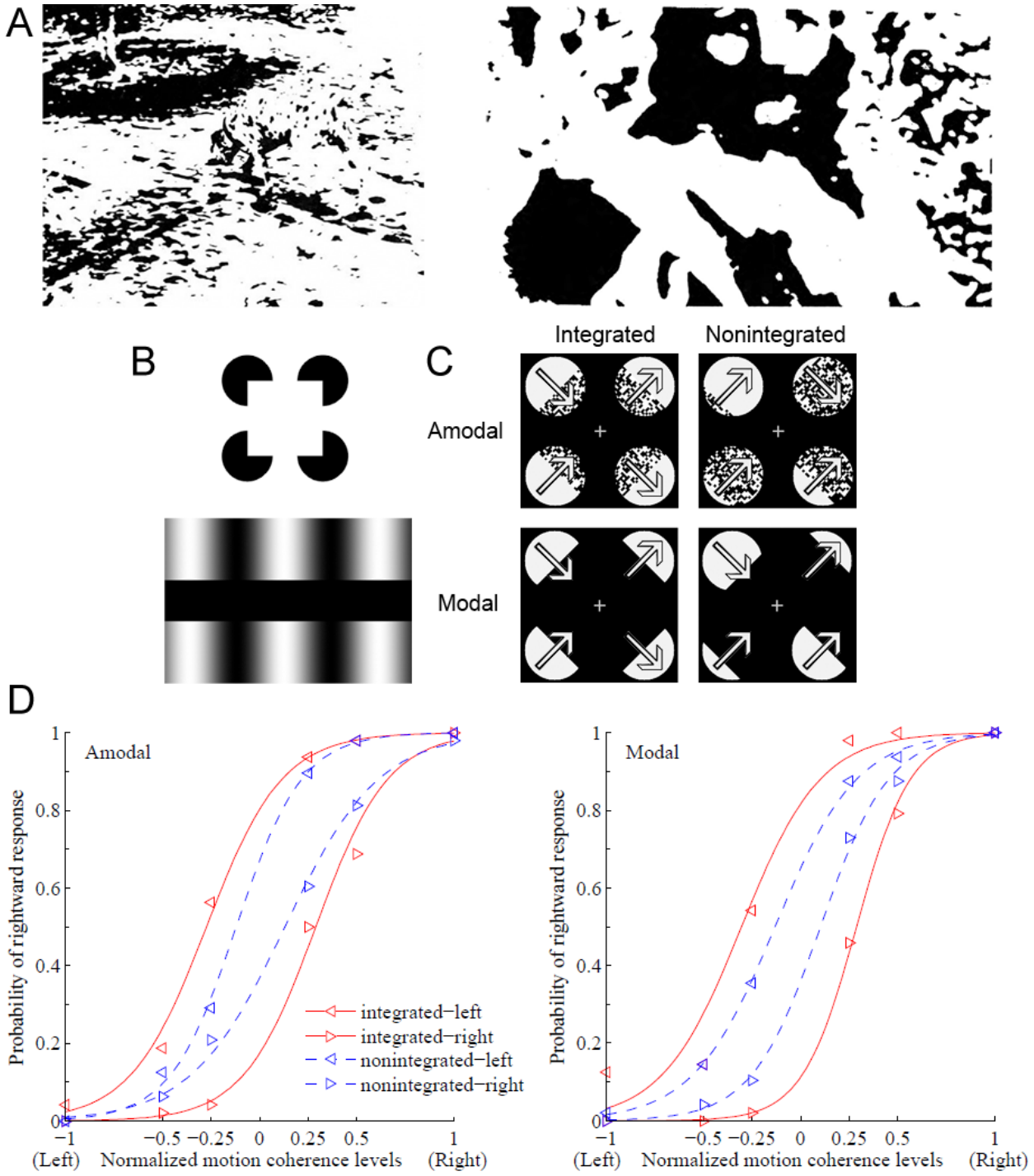


Figure 1-1. Emergence of visual description from perceptual grouping and during filling-in (perceptual completion). (A) Demonstrations of emergent properties from perceptual grouping: a sniffing Dalmatian dog (Left) and the face of a man (Right). (B) Emergence of visual description during filling-in of form and motion: in the Kanizsa square, the four notched pacmen induce a subjective square, with a sharp contour and

enhanced luminance; in the moving visual phantom, when the vertical black bars are moving, the central gaps are perceptually completed with the same pattern, color, speed, and direction of movement as the inducing grating, albeit dimmer. (C) Emergent motion filling-in: the arrows illustrate the local motion directions in each aperture (which change dynamically during each trial); in the integrated condition, the four apertures provides the boundary contour and motion direction information to be filled in, resulting in a vivid percept of a diamond moving leftward (not shown) or rightward (shown) distinct from local motion percept (uphill or downhill motion). (D) Adaptation aftereffects from emergent motion filling-in: In both the amodal and modal conditions, motion aftereffect (i.e., the motion response function is shifted leftward with more rightward responses following leftward adaptation compared with rightward adaptation) is stronger in the integrated than the nonintegrated condition, even though the local motion components are equal.

1.2 Unconscious binding: emergent vs. non-emergent perceptual integration

Notwithstanding being studied objectively as shown in Chapter 2, emergent properties imply a strong sense of human subjectivity. Of course, one can think of the whole chain of emergence in a pure information processing perspective without invoking any subjective experience; color vision, for example, can be described as light impinging on the retina being parsed into three major wavelengths by three classes of cone photoreceptors in the retina, which are then transformed into opponent signals through color-opponent cells in the retina and the lateral geniculate nucleus (LGN), which are further combined linearly and nonlinearly in V1 (Horwitz & Hass, 2012) and later areas. Computations devoid of subjective experiences (qualia) as such, though characteristics of machines, strike all too odd for humans, and this may be precisely the reason why computers and robots are yet to be able to discern images like those in Figure 1-1A. To the extent that objects embody subjective knowledge and expectation, such embodiment remains elusive for computations but is an integral part of the human visual system (Gregory, 1997). Indeed, even though one can conceive a computer or robot which has all

the computing power like us but devoid of any phenomenal experience, it probability exists only as a philosophical zombie (Kanai & Tsuchiya, 2012).

But what can the distinction between emergent and non-emergent properties tell us about visual awareness itself? The unconscious binding hypothesis proposes that perceptual integration is not limited to conscious vision and occurs in unconscious vision during binocular suppression as well (see **Box 1**), but it is unclear what distinguish the two. Is perceptual integration leading to *emergent* properties a signature of visual awareness, unique to conscious vision? This may be too extreme, but a likely scenario is that unconscious vision and conscious vision differ not so much in non-emergent perceptual integration, but are more readily dissociated along emergent perceptual integration. This proposal refines the unconscious binding hypothesis (Z. Lin & He, 2009) and the general notion that consciousness may be the result of information integration (Tononi, 2004). Consider lines and squares. The perception of a line is a non-emergent property, coded by neurons as early as in the retina (Leventhal & Schall, 1983; Levick & Thibos, 1982); the perception of a square is an emergent property as there is no phenomenal square in its constituent lines (only straight contours)(I. Rock, 1986), coded by neurons at multiple levels including V1, V2, and V4 (Hegde & Van Essen, 2007; Pasupathy & Connor, 2002). The same distinction applies to arcs (open-curvature) and ellipses (closed-curvature) as non-emergent and emergent, respectively. Supporting evidence for a critical role of awareness in emergent perceptual integration but not in non-emergent perceptual integration comes from adaptation studies. For instance, Figure 1-2A shows that coding of lines, as measured by tilt aftereffects, is little affected by awareness (Wade & Wenderoth, 1978), particularly for high contrast lines (Blake, Tadin, Sobel, Raissian, & Chong, 2006), as are coding or arcs, indexed by curvature aftereffect (Sweeny, Grabowecky, & Suzuki, 2011); yet coding of ellipses strongly depends on awareness (Sweeny et al., 2011).

Limiting unconscious processing to mainly non-emergent perceptual integration effectively draws a limit on the scope of eye-based integration, and seems to run into conflict with studies showing preserved processing of manipulable objects, that is, tools (Fang & He, 2005). For example, categorically congruent primes rendered invisible by CFS can facilitate categorization and naming of tools, but not nonmanipulable objects

such as animals and vehicles (Almeida, Mahon, & Caramazza, 2010; Almeida, Mahon, Nakayama, & Caramazza, 2008); the magnitude of priming effect for tools is insensitive to whether the prime and the test are the same pictures or different pictures, as long as they belong to the same tool category (Almeida et al., 2010). These findings suggest that unconscious eye-based integration is able to compute motor-relevant information (e.g., graspability). But all these studies use elongated tools (e.g. hammer, knife, axe, and wrench), making it possible that the effects are due to processing and representation of lines (elongated sticks) rather than a more abstract representation of tool category. This reinterpretation is consistent with tool priming being agnostic to the identity of the tools, as they all share the same line shape. Support of this reinterpretation comes from a study (Figure 1-2B) showing that nonelongated tool pictures generate no significant priming effects for tool categorization, but elongated vegetables or stick-like lines produce significant priming effects for tool categorization (Sakuraba, Sakai, Yamanaka, Yokosawa, & Hirayama, 2012). Of course, the lack of evidence for tool categorization without awareness in binocular suppression does not mean that unconscious processing is limited to primitive non-emergent perceptual integration. Indeed, non-emergent lines could well serve as graspable objects, affording actions without awareness (Milner & Goodale, 2008).

Box 1. The unconscious binding hypothesis.

As different visual features are processed by functionally distinct neural pathways and brain areas (Felleman & Van Essen, 1991; Livingstone & Hubel, 1988), a pressing challenge concerns how the visual system subsequently matches the correct features (e.g. a red bar moving rightward and a green bar moving leftward) and how it knows which feature belongs to which object (e.g. a red apple as red, and a yellow banana as yellow but not the reverse). Solutions to this perceptual binding problem are suggested to take place at two different stages of visual processing: an early and automatic stage based on spatiotemporal concurrence (Holcombe & Cavanagh, 2001) and a late, object-based stage mediated by attention to bind distributed features to correctly form coherent object representations (Treisman, 1999; Wolfe & Cave, 1999). Clearly, both conscious vision and unconscious vision face this binding problem (Treisman, 1996). The unconscious binding hypothesis proposes that binding during unconscious vision is possible, albeit fragile—the brain can associate, group, or bind certain features in an invisible scene to form a certain cortical representation, and such binding can be detected under optimal conditions (Z. Lin & He, 2009).

One potential mechanism for unconscious grouping is eye-based integration, in which unconscious information presented in one eye is integrated. This mechanism can also explain eye-specific

attentional modulation, in which attending to a monocular cue enhances the competition strength of a stimulus presented to the cued eye relative to the other, un-cued eye, even though eye of origin information is elusive for awareness (Zhang, Jiang, & He, 2012). Recently, the unconscious binding hypothesis has received some empirical support (Chou & Yeh, 2012; Mudrik, Breska, Lamy, & Deouell, 2011; Yang & Yeh, 2011) from studies using CFS (Fang & He, 2005; Tsuchiya & Koch, 2005). For example, a study (Mudrik et al., 2011) measuring the time to break CFS (Jiang, Costello, & He, 2007) found that complex scenes that included incongruent objects (e.g., a man "drinking" from a hairbrush) escaped perceptual suppression faster than normal scenes did (e.g., a man drinking from a glass). This congruency effect reveals binding of an object and its background scene without awareness.

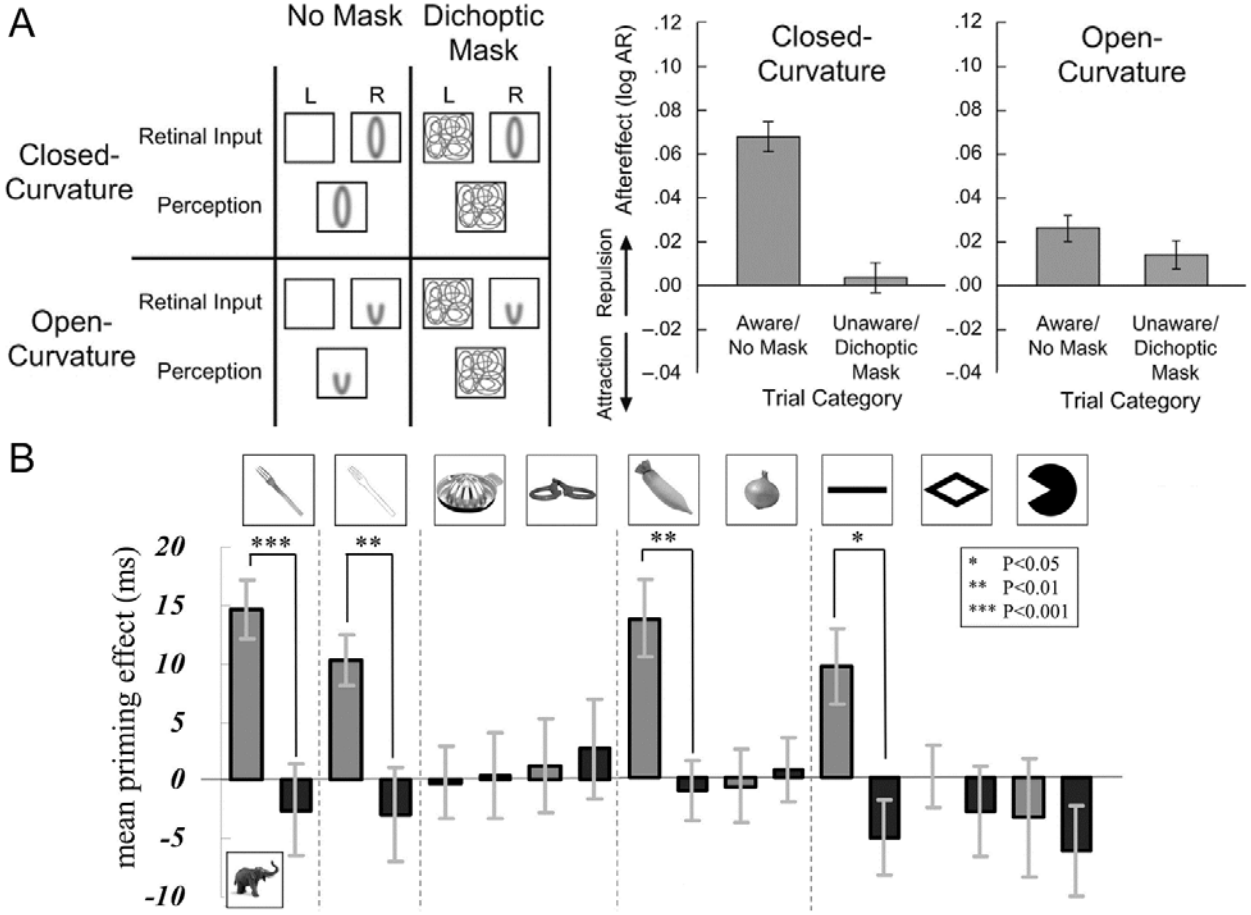


Figure 1-2. A critical role of awareness in emergent perceptual integration but not in non-emergent perceptual integration. (A) Curves vs. circles: Stimuli were presented either to the left eye (L; not shown), the right eye (R, shown), or both eyes. In the no-mask condition, the adaptor was presented to one eye with no mask on the other eye and thus was highly visible; in the dichoptic-mask condition, a dynamic mask was also presented to the other hand, rendering the adaptor invisible. Adaptation was measured as aftereffects in log aspect ratio (AR), with positive value implying a repulsive aftereffect

(e.g., a wide adaptor making a test appear thinner). Processing of open-curvature, a non-emergent property, does not depend on awareness as much as processing of closed-curvature, an emergent property. (B) Categorical priming for tools is due to the representation of lines (elongated sticks) rather than tools per se. Light gray bars represent mean priming effects for tool targets and dark gray bars for animal targets. The icons are examples of the prime stimuli used. Error bars indicate SEM.

1.3 Spatial configuration and the emergence of visual description

The emergence of visual description is inherently a product of interpretation because of the ambiguity in the image data; the success in interpretation then must lie in reaping the statistical regularities in the visual environment (Bar, 2004; Oliva & Torralba, 2007). Rich regularities are embedded in spatial configurations, and thus knowledge of spatial configurations is fundamental to the emergence of both emergent and non-emergent visual descriptions useful for object recognition and scene perception. For instance, the monitor in the office is above the keyboard, which is above the chair; the eyes on a face are above the mouth. Figure 1-3 illustrates how spatial configuration drives the emergence of visual description. A simple curve, when embedded within a circle and two dots, can give rise to an emergent percept of facial expression, such as sadness (down-bended curve) and happiness (up-bended curve, Figure 1-3A). Indeed, adapting to the emergent expression generates a much stronger emotion adaption effect than adapting to the curve alone, even though the circle and the dots provide no information for emotion; conversely, adapting to the emergent expression generates a much weaker curvature adaption effect than adapting to the curve alone, even though the circle and the dots provide no information for curvature (Xu, Dayan, Lipkin, & Qian, 2008). This finding is consistent with the state transition principle proposed early on: the curve as feature state and the curve embedded within the circle and dots as expression state, so that adaptation effect is maximized when the adaptor and the test recruit the same neural circuits (i.e., when they are in the same state).

Spatial configuration also drives the emergence of other descriptions such as visual saliency. For instance, visual search for a unique left tilted bar among three right

tilted bars is relatively easy, but adding homogenous, non-informative configurations to the search array can make the left tilted bar much more or less salient, known as object superior or inferior effects, respectively (Figure 1-3B) (Pomerantz & Pristach, 1989; Weisstein & Harris, 1974). Similarly, adding non-homogenous but non-informative configurations can also enhance or reduce the saliency of the unique bar; for instance, Figure 1-3C shows that adding alternating vertical and horizontal lines into an array of lines similar to Figure 1-3B (left) produces similar perceptual integration effects, resulting into percepts of emergent two-line shape structures(Zhaoping & Guyader, 2007). Such drastic changes in saliency and thus visual search performance is consistent with the state transition principle, as state changes (from lines to shapes) necessarily induce changes in property, saliency included, into those inherent to the second state.

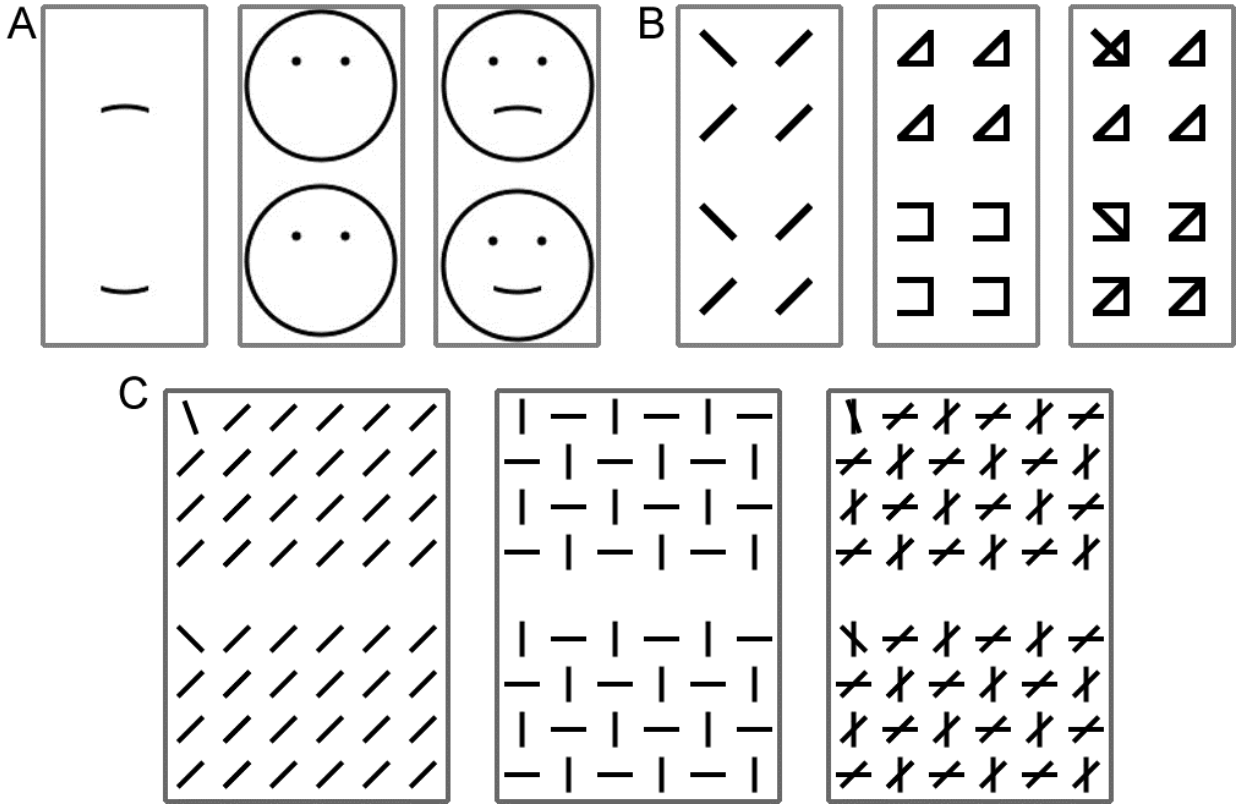


Figure 1-3. Spatial configuration drives the emergence of visual description. (A) The emergence of facial expressions based on simple curves embedded within spatial configurations: when in isolation, the two curves are merely curves bending in the opposite manner (left), but when placed properly within a simple spatial configuration (middle), they yield two distinct emergent descriptions, a sad facial expression and a

happy one, respectively (right). (B) The emergence of saliency through spatial configurations: when in isolation, the left tilted line stands out from the three right tilted lines equally well between the top and bottom configurations (left), but when embedded within non-informative, homogenous spatial configurations (middle), the left tilted line can become much more (top right) or less (bottom right) salient depending on the specific configuration. (C) The emergence of saliency based on lines embedded within spatial configurations: when in isolation, the left tilted line (25°) on the top stands out from its neighboring right tilted lines equally well (if not less) than its counterpart (45°) does in the bottom configurations (left), but when embedded within an identical, non-informative spatial configuration (middle), the left tilted line on the top now becomes much more salient than its counterpart does in the bottom configurations (right).

Spatial configuration effects imply that objects be coded orderly within their spatial contexts, but how robust and automatic is it? This issue of object and space integration is fundamental to contextual effects. To address this, in Chapter 3, using a moving frame technique, we directly manipulate the relative locations of objects within the contextual frames and assess how robust and automatic the coupling of objects to the contextual frames is (Figure 1-4). Coupling of objects to the contextual frames is indexed by non-retinotopic, frame-centered priming and masking effects. We show that the coupling of objects to the contextual frames is pervasive, present in tasks taping into working memory, implicit memory, and perception processes; it is relatively automatic, occurring without active memory involvement, object competition, or top-down anticipation; when objects are task irrelevant; and even when it is detrimental to the task. Pervasive and automatic coupling of objects to the contextual frames demonstrate a mechanism of frame-centered representation and integration, serving as the foundation of spatial configuration effects. Indeed, such a frame-centered representation and integration shapes the automatic deployments of visual attention. For instance, as reported in Chapter 4, we present two frames in succession, forming apparent translational motion, or in mirror reflection, with an uninformative, transient cue flashing at one of the item locations during the first frame. Despite that the cues are presented in a spatially separate frame, in both translation and mirror reflection, human performance is enhanced when the target in the second frame appears on the same relative location as the cue location

than on other locations. These observations lead us to propose the object cabinet theory, a modified object file account (Kahneman, Treisman, & Gibbs, 1992; Treisman, 1992): objects such as attentional cues (the files) within the reference frame (the cabinet) are orderly coded in their relations to the frame.

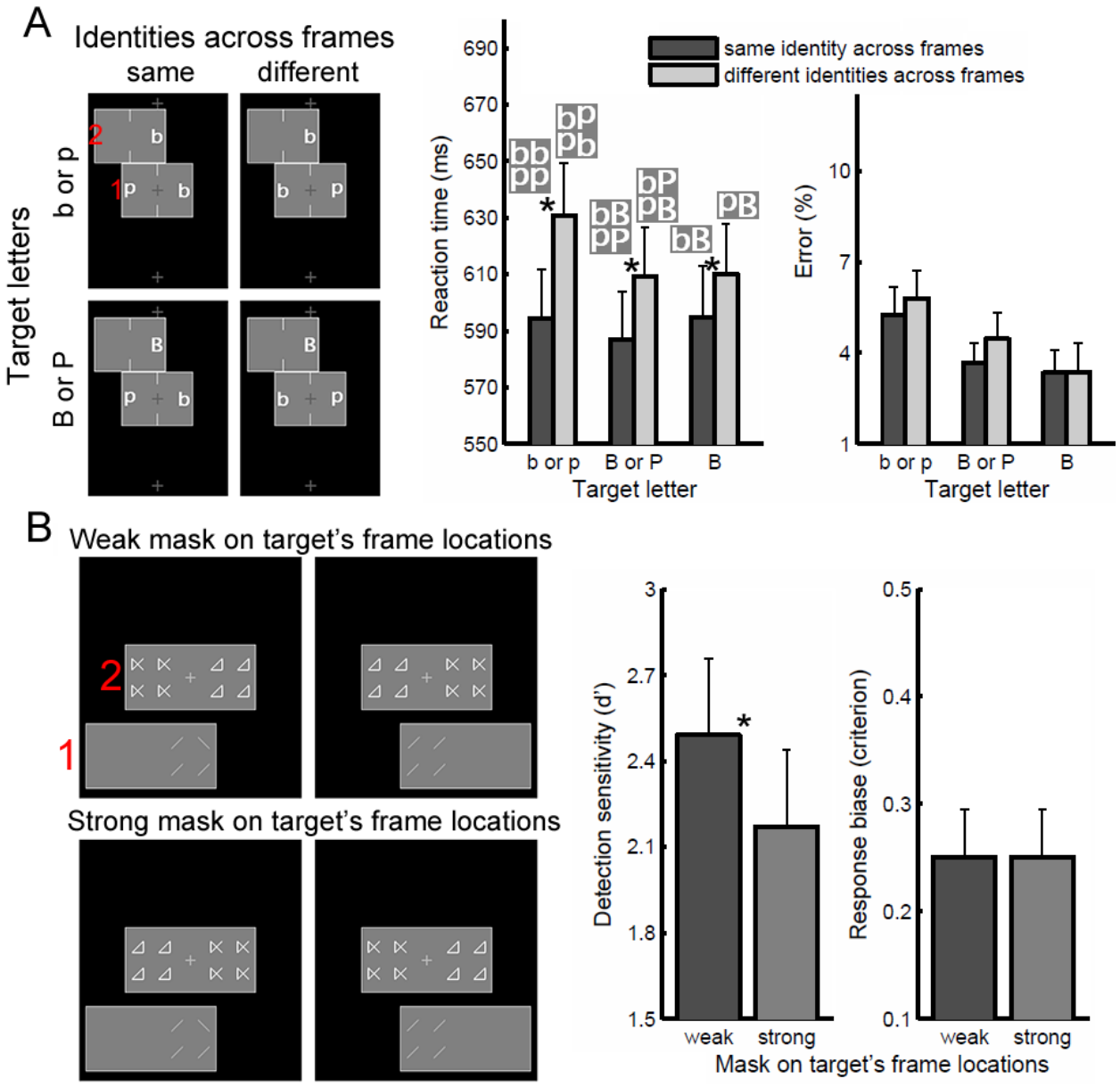


Figure 1-4. Robust and automatic coding of objects to the contextual frames as revealed by frame-centered priming and masking effects. (A) Frame-centered priming effect: The first frame with two lower case letters b and p (random order) are presented on the fixation, followed by the second frame on one of four corners (upper left shown) with a letter above or below the fixation; the task is to categorize whether the letter at the end is

b/B or p/P, case insensitive (e.g., the first row shows a lower case target b, the second row an upper case target B). Critically, the letter identities on the same frame locations are either the same (e.g. the first column) or different (e.g., the second column) across the two frames. Responses are faster when the target letter shares the same relative frame location as the same identity prime letter, regardless of target cases, even though the prime letters are uninformative and task-irrelevant. (B) Frame-centered masking effect: The target frame with a search display is now presented first on one of four corners, followed by the second frame with two mask displays; the task is to detect whether there is an odd, left-tilted target bar or not on the search display. Perceptual sensitivity but not response bias is higher when the target is followed by a weak mask on the same frame-centered location than a strong mask, even though perceptual integration is detrimental to the task. Error bars: standard error of the mean. *: statistically significant difference.

1.4 Associative knowledge and the emergence of visual description

Rich statistical regularities are embedded not just in spatial configurations, but also in the vast associative knowledge one learns in the life time, highlighting the importance of temporal structures. Indeed, objects and events in our daily life usually covary in predictable relations (Biederman, 1972), with semantic consistency information of objects and the background available within a brief glimpse (e.g., ~80 ms) (Davenport & Potter, 2004). These learned associations through past experiences constrain the emergence of visual description so that the presence of a particular object informs the probable presence and locations of other potential objects around it. For instance, associative knowledge can guide our expectation and hence the deployment of top-down attention: a target is thus searched faster when it appears within an old, predictive context than within a new context (Chun & Jiang, 1998; Chun & Jiang, 1999), possibly involving the interaction between brain areas for retrieval of memories for spatial context and the frontoparietal network for top-down attention deployment (Stokes, Atherton, Patai, & Nobre, 2012; Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006). Indeed, exploiting associative knowledge is an integral component in previous conceptualization of human cognition, such as schemata (Biederman, Mezzanotte, & Rabinowitz, 1982;

Hock, Romanski, Galie, & Williams, 1978; Mandler & Johnson, 1976), scripts (Schank, 1975), and frames (Minsky, 1975), wherein objects and their contextual information are bound in representation. For example, according to Minsky, encountering a new situation evokes the retrieval of a particular structure from memory—called a frame—so that the old frame can be adapted to accommodate new information.

The importance of associative learning in generating associative knowledge is thus apparent, making it critical to understand the principles that govern the learning of associations. Research in statistical learning indicates that passive viewing of complex visual scenes is sufficient for the extraction of shape cooccurrences (Fiser & Aslin, 2001), but how automatic is the association once it is established? A recent study shows that colors that were previously associated with monetary rewards can capture attention, and removing monetary rewards abolished the effect (Anderson, Laurent, & Yantis, 2011). In Chapter 5, we provide strong evidence that associated links can be established rapidly and without rewards: a physically inconspicuous stimulus, after being associated with targets through short-term associative learning in visual search, captures visual attention involuntarily, even when such capture is detrimental to task performance. Importantly, as revealed in Chapter 6, such an involuntary capture effect of attention is restricted to displays where the location of the target is uncertain and thus the associated colors are in competition with other neural colors. Indeed, in a flanker task where the target location is fixed at the fixation, incompatible flanking distractors with the associated colors actually cause less interference effect on target processing than incompatible flanking distractors with neutral colors, demonstrating an inhibition effect of the associated colors when the target location is known and fixed. These results thus point toward the automatic, pervasive nature of established associative links but also highlight the limit of associative learning.

1.5 Concluding remarks

Understanding how visual descriptions emerge from various inputs is a fundamental goal in human cognition. We assign an important role of perceptual integration in the emergence of visual description and distinguish emergent perceptual integration from non-emergent perceptual integration, which has important implications

for our understanding of unconscious processing (unconscious binding hypothesis). Spatial and temporal structures shape and even underlie the emergence of visual description. We argue that understanding state changes during the emergence of visual description is the key to understanding the various changes that accompany the emergence process (state transition principle).

Chapter 2

Emergent filling-in induced by motion integration reveals a high level mechanism in filling-in[2]

The visual system is intelligent, capable of recovering a coherent surface from an incomplete one, a feat known as perceptual completion or filling-in. Traditionally, surface features are interpolated in a way that resemble the fragmented parts. Using four circular apertures, here we show a distinct completed feature (horizontal motion) from local ones (oblique motions), which we term emergent filling-in. Adaptation to emergent filling-in motion generates a dynamic motion aftereffect (MAE) that is not due to spreading of local motion from the isolated apertures. The filling-in MAE occurs in both modal and amodal completions, and is modulated by selective attention. These findings highlight the importance of high-level interpolation processes in filling-in, and are consistent with the idea that, during emergent filling-in, the more cognitive/symbolic processes in later areas (e.g. the middle temporal visual area, MT, and the lateral occipital complex, LOC) provide important feedback signals to guide more isomorphic process in earlier areas (V1 and V2).

2.1 Introduction

The visual system is intelligent, capable of recovering a coherent surface from an incomplete one, a feat known as perceptual completion or filling-in. A celebrated example is the Kanizsa figure (Figure 2-1A), where a few notched pacmen induce a subjective shape, with a sharp contour and enhanced luminance (Kanizsa, 1979). Because the completed part shares the same "mode" (i.e. brightness) as the rest of the figure, it is termed modal completion, as opposed to amodal completion (Michotte, Thines, & Crabbe, 1964), where the completed part is occluded and thus lacks visible attributes (e.g. the notched pacmen are perceptually completed as occluded disks). Conceivably, to

[2] Parts of this chapter were published as Lin, Z, & He, S. (2012). Emergent filling-in induced by motion integration reveals a high level mechanism in filling-in. *Psychological Science*.

perceive Kanizsa-like subjective figures, the visual system has to interpolate critical missing links from the given fragmented image so that contour and surface can be completed, modally or amodally. Traditionally, surface features are interpolated in a way that resembles the fragmented parts. In the moving visual phantoms (Figure 2-1B), for instance, wherein a low-contrast moving grating is divided by an empty, orthogonal gap, the gap is perceptually completed with the same pattern, color, speed, and direction of movement as the inducing grating, albeit dimmer (Tynan & Sekular, 1975). Using a novel configuration, here we show that in both modal and amodal completions the interpolation processes can generate a solution distinct from the fragmented parts. Specifically, we presented a translationally moving diamond perceived through four circular apertures, as shown in Figure 2-1C. Integrating the four apertures provides the boundary contour and motion direction information to be filled in, resulting in a vivid percept of a diamond moving leftward or rightward. Because the filling-in motion percept from integration is distinct from local motion percept (uphill or downhill motion), we call this type of perceptual completion emergent filling-in.

Emergent filling-in affords us a unique opportunity to examine the long standing debate over isomorphic versus symbolic mechanisms of filling-in: the isomorphic theory argues that filling-in is achieved through pointwise neural representation of visual features in retinotopic visual areas (Gerrits & Vendrik, 1970), whereas the symbolic theory contends that contour and surface interpolations take place at higher areas, based on the contrast information at the surface border (Gregory, 1972). Emergent filling-in of motion implies the essential contributions from both early retinotopic areas and late non-retinotopic areas. Specifically, due to their small receptive fields, neurons in early visual cortex can only measure the component of motion perpendicular to a contour that extends beyond their receptive fields (e.g., uphill or downhill motion in our configuration) and thus cannot recover the true motion, known as the aperture problem in motion perception (Marr & Ullman, 1981). Downstream, the middle temporal visual area (MT or V5) solves this problem by integrating directional responses from V1 (Pack & Born, 2001). This highlights the synergetic interactions between local isomorphic processes (e.g. contour interpolation) and global symbolic processes (e.g. structure and global motion extraction) in enabling emergent filling-in.

If there is indeed neural emergent filling-in, one would expect to observe adaptation effects from the filling-in motion that is not due to spreading of local motions from the apertures. Moreover, if higher level symbolic processes (e.g. in MT/V5 and the lateral occipital complex, LOC) plays a critical role in guiding more isomorphic process (e.g. in V1 and V2), one would further expect that the filling-in specific adaptation effects, if observed, should be modulated by selective attention, as opposed to traditional filling-in that is relatively automatic, independent of attention (M. Meng, Remus, & Tong, 2005; Sasaki & Watanabe, 2004). To test these ideas, we adapted observers to the motion displays shown in Figure 2-1C (integrated) and measured the motion aftereffect (MAE) outside of the apertures using a dynamic test. Because MAE from local motion in the apertures could potentially transfer to the test region (Bex et al., 1999; Lalanne & Lorenceau, 2006; X. Meng et al., 2006; Price et al., 2004; Snowden & Milne, 1997; Weisstein et al., 1977), we also adapted observers to control displays, which had the same local apertures but could not be integrated to generate emergent filling-in, as shown in Figure 2-1C (nonintegrated). We found a much stronger adaptation effect in the integrated (with filling-in) condition than the nonintegrated (without filling-in) condition. The difference between the two conditions therefore reveals a MAE not confounded by transfers from nearby apertures, but due solely to adaptation to the emergent, perceived filling-in motion, an effect we call filling-in MAE. The filling-in MAE was observed in both our modal and amodal conditions, but, importantly, was reduced when the attention load of a fixation task was increased during adaptation, suggesting that the more symbolic processes in later areas might provide important feedback signals to guide more isomorphic process in earlier areas (Halgren et al., 2003), which critically depends on attention. Because the filling-in MAE was only present with dynamic tests and disappeared with static tests, the effect was likely a high-level one, involving MT/V5 (Culham et al., 1998; Nishida, Ashida, & Sato, 1994). Moreover, the filling-in MAE was reduced when the surface filling-in was removed with only contour filling-in remaining, suggesting the contribution of both surface filling-in and contour filling-in (see Figure 2-S1 in the Supplemental Material). In short, adaptation to emergent filling-in motion can generate an attention-dependent, high-level MAE.

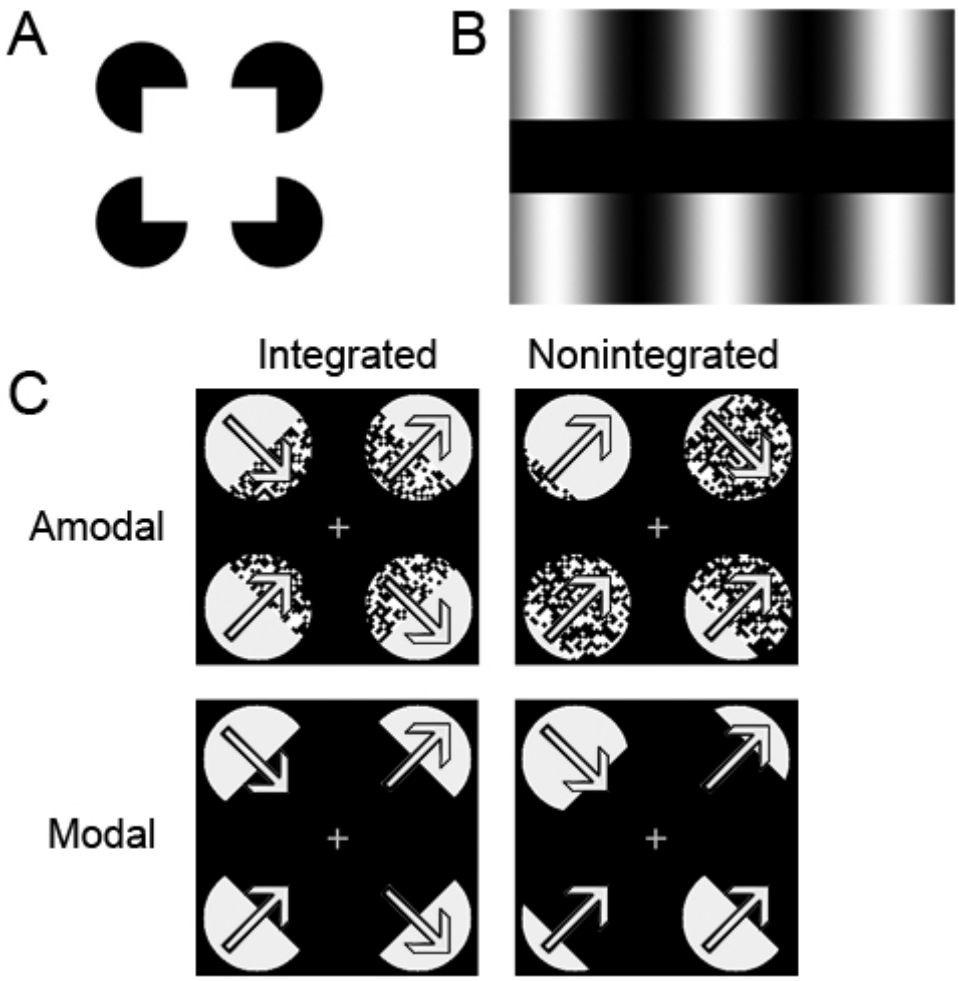


Figure 2-1. Filling-in (perceptual completion) illustrated. (A) Kanizsa square. (B) Moving visual phantom. (C) Snapshots of the adapting stimuli in the current study. Four types of adapting displays were tested, each of which could move either leftward or rightward. The arrows illustrate the local motion directions in each aperture (which change dynamically during each trial). In the passive task, observers fixated on the cross, the color of which changed every 500 ms. In the attention task, observers fixated on the cross and responded to it by pressing a key whenever it was a target, which could be any red item (easy task), or a red upper cross or a green lower crosses (difficult task). After 50 s (initial) or 5 s (top-up) of adaptation, a random dot kinematogram test immediately appeared at the center, without overlapping any aperture. Observers clicked the mouse to indicate the motion direction of the test.

2.2 Method

Observers. Six observers participated in the passive adaptation experiment (in both the amodal and modal conditions), 8 in the attention experiment (4 in the modal condition and 4 in the amodal condition). One observer was the author ZL; others, naïve to the study, were drawn from the University of Minnesota community and received money for their time. All had normal or corrected-to-normal vision, and signed an Institutional Review Board approved consent form.

Apparatus. The stimuli were presented on a gamma-corrected 22-inch CRT monitor (model: Hewlett-Packard p1230; refresh rate: 100 Hz; resolution: 1024 × 768 pixels) using MATLAB Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Subjects sat approximately 57 cm from the monitor with their heads positioned in a chin rest in a dimly lit room while an experimenter was present.

Stimuli. The adapting display consisted of four circular apertures (diameter: 4°; center-to-center distance: 6°; luminance: 23.4 cd/m^2 on a black background) centered on the fixation (center-to-center distance: 4.24°), as shown in Figure 2-1C. In the integrated condition, an occluded diamond (each size: 8.24°; speed: 11.68°/s) moved leftward or rightward behind the apertures. Thus, within each individual aperture, a sequence of uphill-downhill (or downhill-uphill motion) cycle was perceived, 150 frames each cycle (two frames in between direction reversal). In the nonintegrated condition, for each trial, the starting phases for the four apertures were randomly chosen among the 1st to 150th frame; the direction sequences were randomly distributed among the apertures with half moving uphill-downhill and half moving downhill-uphill. As such, a coherently moving diamond could not be perceived through integration. On average, the local motion signals in each aperture are the same across the integrated and non-integrated conditions. In both the integrated and nonintegrated conditions, the diamond could be a flickering surface consisting of half white patches (size: 0.16° × 0.16°; luminance: 26.7 cd/m^2) and half black patches, randomly distributed for each frame (amodal condition), or a uniform black surface as the background (modal condition). Flickering was removed in the attention task to reduce attention capture by fixing the distribution of white and black patches throughout each trial.

The test display was a random dot kinematogram (RDK, Britten, Shadlen, Newsome, & Movshon, 1992) that appeared within an imaginary circular aperture (diameter: 3.17°) centered on the fixation, with no spatial overlap with the adapting apertures (contour-to-contour distance: 0.66°). On each video frame, a fraction of dots (size: 0.12° × 0.12°; luminance: 19.1 cd/m^2 on a black background; density: 30 dots/deg^2/s) were randomly chosen and plotted at a displacement of 0.18° from the positions three frames (30 ms) earlier to generate apparent motion (6°/s); the other dots were plotted at random locations. The probability that a dot would be plotted in coherent motion is determined by the motion coherence.

The fixation consisted of crosses (length: 0.39°; width: 0.12°) in rapid serial visual presentation. In the passive task, the fixation's color cycled through red, green, and yellow (two items/s). In the attention task, the color for each cross could be red, green, or blue and the shape could be an intact cross, an upper cross (missing the lower part), or a lower cross (missing the upper part). In the easy task, red items, regardless of the shape, were the targets; in the difficult task, upper crosses in red and lower crosses in green, not the reverse, were the targets. Each target was presented for 25 frames, followed by a gray cross for 750 frames (i.e., two items/s).

Procedure. Motion sensitivity in each individual observer was calibrated prior to the adaptation phase. Sensitivity was measured using a RDK (duration: 800 ms) with eight coherence levels (2%, 5%, 10%, 15%, 20%, 30%, 40%, and 60%) and two directions (leftward and rightward), randomized across 160 trials in four blocks. Observers indicated the motion direction by clicking the mouse. For each individual observer, a logistic function was fitted to the responses to obtain the amount of coherence for asymptotic performance (the average of the absolute coherence levels for 97% ''rightward'' and 97% "leftward'' responses), which we defined as 1 unit of normalized coherence. For each direction of motion, stimuli with 100%, 50%, and 25% unit of normalized coherence served as test stimuli in the adaptation phase.

Following sensitivity calibration, observers adapted to four types of stimuli in a blocked-design, a full crossing of integrated vs. nonintegrated and modal vs. amodal, with the order randomized within observers. For each type of stimuli, observers adapted to two directions, left and right, in two blocks, with at least 2 minutes of rest in between.

The order of adaptation directions was counter-balanced across observers. Each block started with an initial adaptation trial (50 s), followed by 47 top-up adaptation trials (5 s). Each adaptation trial was followed immediately by a RDK test lasting 200 ms. The next adaptation trial began once the observer indicated the motion direction of the test through non-speeded mouse clicking.

The attention experiment was similar to the passive experiment, except for the addition of a detection task during adaptation, a reduction in the number of top-up trials (from 47 to 35), and unequal numbers of trials for each test level. During adaptation, observers were asked to detect the appearance of any fixation target (9 to 11 targets in the initial trial; 1 to 2 targets in the top-up trial) as accurately and quickly as possible through key pressing. Easy and difficult conditions were blocked. The order of adaptation directions and task difficulty was counter-balanced across observers. For the test, to make it difficult for the observers to guess the distribution of the numbers of trials for each coherence level, we randomly distributed the 36 trials in each block to the six coherence levels (ranging from 2 to 10).

2.3 Results

Dynamic MAE to the filling-in motion. Adaptation effects were observed in both the amodal and modal completion conditions. After looking at the fixation and adapting to the task-irrelevant surrounding motion in one direction, observers were more likely to perceive the RDK moving in the opposite direction. Figure 2-2 shows motion direction adaptation: the motion response function was shifted leftward (i.e. more rightward responses) following leftward adaptation compared with rightward adaptation. This MAE in the nonintegrated condition replicates a now widely reported “phantom” MAE (Weisstein et al., 1977), reflecting transfer of MAE from adapting locations to unadapted locations. Importantly, in both the amodal and modal conditions, the adaptation effect in the integrated condition was much larger than that in the nonintegrated condition, despite the same adapting stimuli in the individual apertures. Such a difference between the integrated and nonintegrated conditions thus reflected an adaptation effect to emergent filling-in, and likely involved both contour filling-in and surface filling-in (see also the supplementary experiment in Figure 2-S1).

To quantify the strength of the adaptation effect, the difference between the leftward and rightward adaptation in points of perceived null motion—the amount of motion coherence for which observers were equally likely to respond ‘‘left’’ or ‘‘right”—was used. Since the pattern between the amodal (the difference in adaptation effect between integrated and nonintegrated conditions: 0.14) and modal (0.16) conditions are similar (paired t-test, $t(5)=0.28$, $p=0.78$), the data are combined. The shift of null points following leftward relative to rightward adaptation was 0.33 ± 0.04 unit of normalized motion coherence for the integrated stimuli, significantly larger than 0.18 ± 0.06 for the nonintegrated stimuli, $t(5)=3.29$, $p=0.022$. In terms of actual, nonnormalized coherence, the shift was 0.19 ± 0.03 for the integrated stimuli, compared with 0.12 ± 0.04 for the nonintegrated stimuli, $t(5)=4.33$, $p=0.008$. However, because the points of perceived null motion are logistically fitted, it is not an assumption-free approach. To quantify the adaptation effect in a more objective manner, we calculated the percentage of rightward responses across all coherence levels tested for each condition, and then compared the difference in the percentage of rightward responses between the leftward and rightward adaptation, which therefore serves as an objective, quantitative index of the adaptation effect. As shown in Table 2-1, for the integrated stimuli, leftward adaptation (relative to rightward adaptation) resulted in 24.8% ± 4.3% more rightward responses to the test, much larger than 10.1% ± 4.9% for the nonintegrated stimuli, $t(5)=5.55$, $p=0.012$.

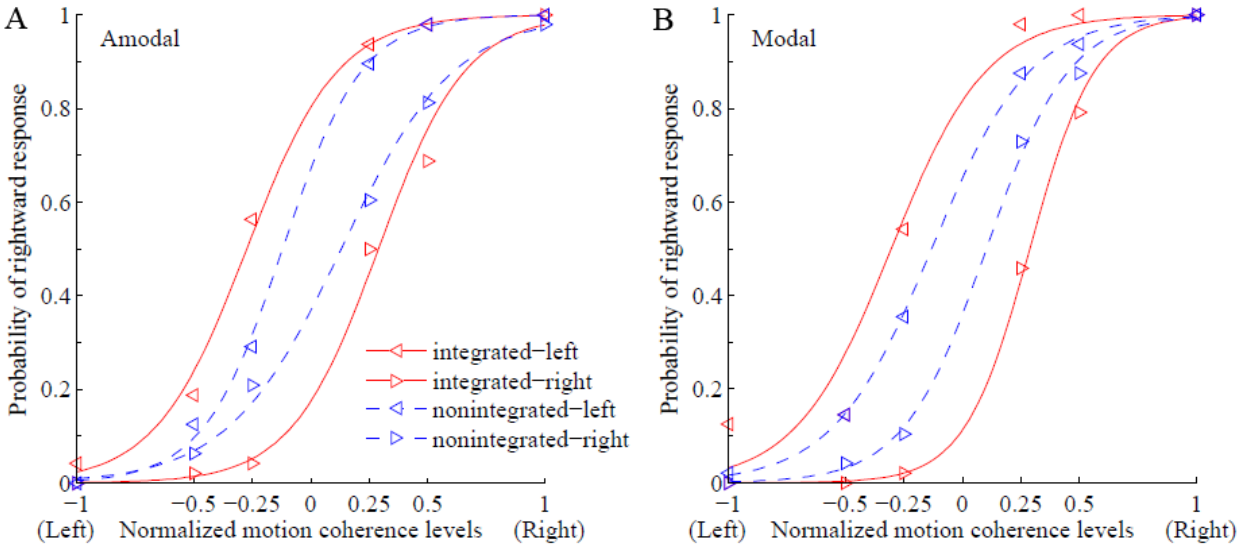
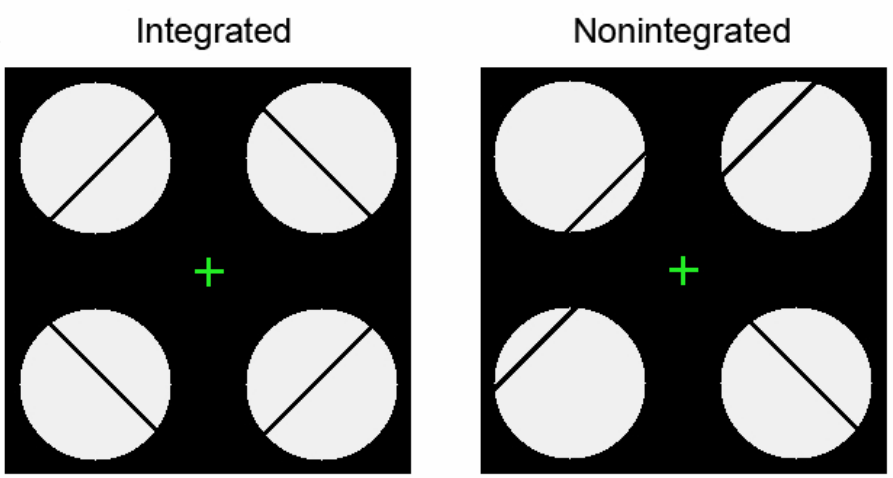


Figure 2-2. Results from the passive adaptation experiments. For both the amodal (A) and modal (B) completion conditions, the probability of rightward response following

adaptation to integrated-leftward, integrated-rightward, nonintegrated-leftward, and nonintegrated-rightward motion is plotted as a function of normalized motion coherence of the test stimuli. Each data point is the average of six observers; lines are logistically fitted to the data points in each condition.

A



B

	Difference in Magnitude of Motion Aftereffect Between the Integrated and Nonintegrated Conditions	
	Contour Filling-in	Average of Modal and Amodal Filling-in
Subject 1	0.076	0.337
Subject 2	0.053	0.184
Subject 3	0.056	0.318

Figure 2-S1. Contributions of both contour filling-in and surface filling-in in the emergent filling-in motion aftereffect. (A) Stimuli: Snapshots of the adapting stimuli with only contour filling-in and minimal surface filling-in. The same parameters and design was used as in the original integrated/nonintegrated conditions, except for the removal of surface from the apertures. (B) Results: Difference in the magnitude of motion aftereffect between the integrated and nonintegrated conditions across the two filling-in conditions

in three subjects (unit in normalized motion coherence). The filling-in MAE was reduced when the surface filling-in was removed with only contour filling-in remaining, suggesting the contribution of both surface filling-in and contour filling-in.

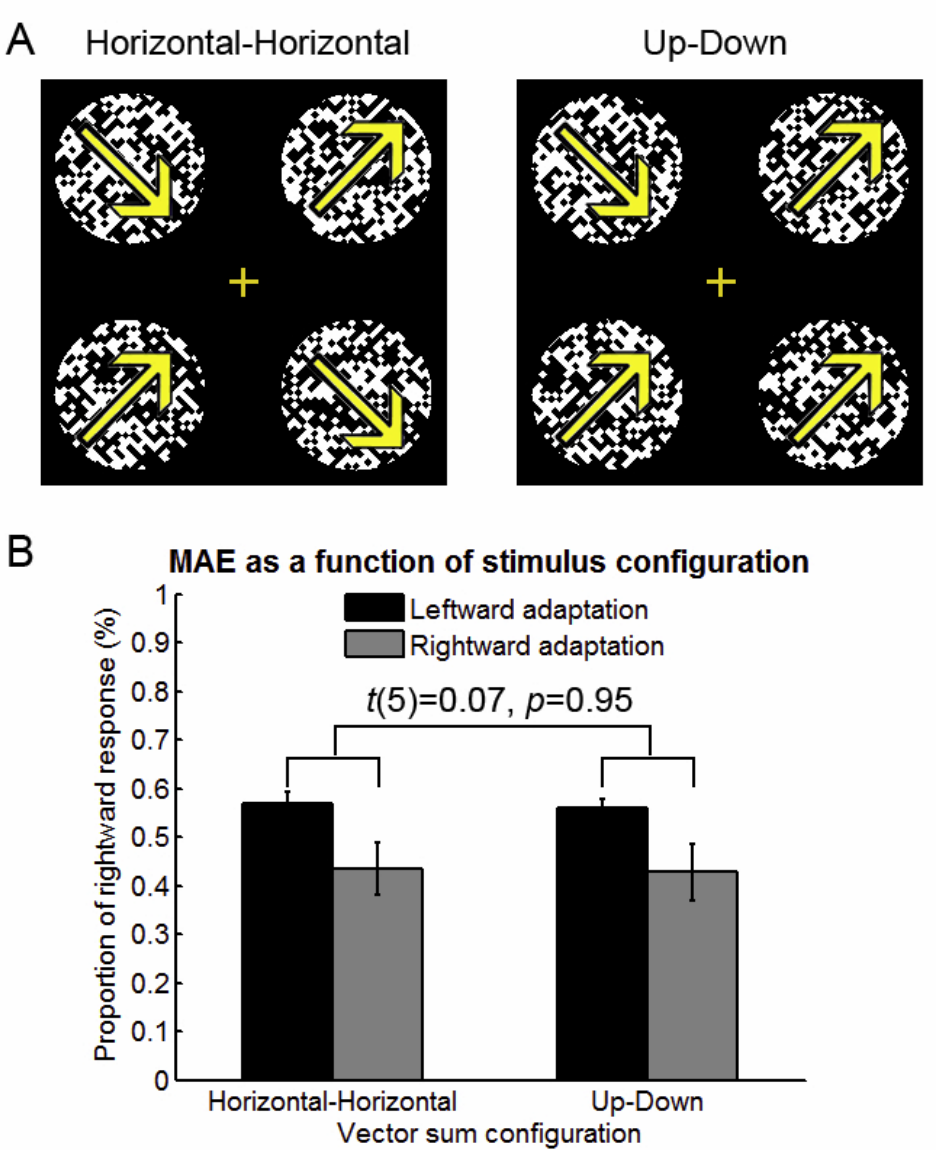


Figure 2-S2. Does a balanced configuration (i.e. two up and two down motions across the four apertures, as in the integrated condition) produce a similar adaptation effect as a sequentially balanced configuration (i.e. one up and three down motions sequentially alternate with three up and one down motions across the four apertures, as in the nonintegrated condition)? (A) Stimuli: Snapshots of the adapting stimuli with a balanced configuration (Horizontal-Horizontal) and a sequentially balanced configuration (Up-Down).The same parameters and design was used as in the original

integrated/nonintegrated conditions, except for the differences in the stimuli. (B) Results: Data from six observers show that MAE in the Horizontal-Horizontal condition (13.5%) is similar to that in the Up-Down condition (13.2%), $t(5)=0.07$, $p=0.95$, suggesting that a balanced configuration produce a similar adaptation effect as a sequentially balanced configuration in a configuration similar to the original emergent filling-in configuration.

Filling-in MAE depends on attention. Does the motion adaptation effect from completion depend on the attention resource to the adaptation stimuli (isolated and completed)? To address this, the adaptation effect due to completion (i.e. the difference between the integrated and the nonintegrated conditions) is compared between the easy and the hard fixation tasks (Lak, 2008). In both the amodal and modal completion conditions, we consistently found that the adaptation effect due to completion was much bigger in the easy fixation task than that in the hard one, as shown in Table 2-1. To quantify this, we again calculated the percentage of rightward responses across all coherence levels tested for each condition and used as an index the difference in the percentages of rightward responses between the leftward and rightward adaptation conditions. In the easy task, for the integrated stimuli, leftward adaptation (relative to rightward adaptation) resulted in 24.3% ± 7.9% more rightward responses to the test, much larger than 0.03% ± 6.5% for the nonintegrated stimuli, $t(7)=2.57$, $p=0.037$. However, in the hard task, leftward adaptation to the integrated stimuli resulted in 13.9% ± 8.6% more rightward responses, similar to adaptation to the nonintegrated stimuli 15.3% ± 4.6%, $t(7)=-0.25$, $p=0.812$. The difference in adaptation between integrated vs. nonintegrated stimuli is thus robust in the easy task, 21.5% ± 7.8%, but essentially absent in the hard task, -1.4% ± 5.3%.

Table 2-1. Magnitude of motion aftereffect as measured by the difference in the percentage of rightward responses between the leftward and rightward adaptation.

	Magnitude of Motion Aftereffect (%)	
	Integrated	Nonintegrated
Passive Task (*N*=6)	24.8 ± 4.3	10.1 ± 4.9
Low Attention Load (*N*=4)	24.3 ± 7.8	2.8 ± 6.5
High Attention Load (*N*=4)	13.9 ± 8.6	15.2 ± 4.6

2.4 Discussion

Interpolation processes during perceptual completion usually follow the features from the available parts to fill in the gaps. In this study, using motion apertures we showed that filling-in of motion direction could be distinct from local motion directions, hence the term emergent filling-in. Moreover, adaptation to this emergent filling-in motion, either modally or amodally, generated a MAE in a region outside of the adapting apertures when a dynamic test was used. This MAE was much reduced when the same local adapting motion signals were presented in the apertures but could not be integrated to fill in the test region. This contrast demonstrates a MAE due to emergent filling-in. The filling-in MAE was reduced when the surface filling-in was removed with only contour filling-in remaining, suggesting the contribution of both surface filling-in and contour filling-in (see Figure 2-S1 in the Supplemental Material). By manipulating the attention load of the central fixation task during adaptation, we further showed that the filling-in MAE was dramatically reduced when the central fixation task demanded more attention than when it demanded less.

The attention modulation observed is in contrast with previous studies showing an early, isomorphic mechanism for completion. For example, in psychophysical studies, it has been shown that both modal and amodal completion appear to be obligatory and occur preattentively (Davis & Driver, 1994; Z. J. He & Nakayama, 1992; Rauschenberger & Yantis, 2001; Rensink & Enns, 1998), before object-based attention selection (Moore, Yantis, & Vaughan, 1998) and surface-based attention selection (Davis & Driver, 1997), and at least for amodal completion before perceptual grouping by shape similarity (Palmer, Neff, & Beck, 1996). A recent study with fMRI also suggests that filling-in of

visual motion from the phantom illusion in V1 and V2 could occur independently of attention (M. Meng et al., 2005), similar to color filling-in (Sasaki & Watanabe, 2004). The distinction of attention modulation between our study and these previous studies suggests the important role of high level symbolic processes (e.g. in MT/V5 and/or LOC) in emergent filling-in, possibly by providing feedback signals to guide isomorphic processes in early visual cortex. In principle, attention modulation of MAE can be traced to attention effect on the interpolation processes, or on the representation of emergent motion, or both. Because motion integration in MT is sensitive to attention (Cook & Maunsell, 2004), attention effect on the interpolation processes is likely to play an important role.

The idea that during emergent filling-in high level symbolic processes provides feedback signals to guide isomorphic processes in early visual cortex is intriguing. Over the last two decades, many studies have provided evidence in support for the isomorphic theory, leading to the belief that perceptual completion is hardwired in early visual cortex to provide an anticamouflage device for discovering hidden objects (Komatsu, 2006; Pessoa, Thompson, & Noe, 1998; M. L. Seghier & Vulleumier, 2006). For example, evidence from psychophysics (Davis & Driver, 1994, 1998; Z. J. He & Nakayama, 1992; Pillow & Rubin, 2002; Rensink & Enns, 1998; Smith & Over, 1975; Weisstein et al., 1977), modeling (Grossberg, 1994; Z. Li, 1998), neuropsychology (Mattingley, Davis, & Driver, 1997), neurophysiology (De Weerd, Gattass, Desimone, & Ungerleider, 1995; Fiorani Junior, Rosa, Gattass, & Rocha-Miranda, 1992; Lee & Nguyen, 2001; Roe, Lu, & Hung, 2005), and fMRI (M. Meng et al., 2005; Sasaki & Watanabe, 2004) generally reveals a low-level mechanism in V1 and/or V2 responsible for completion, but not a cognitive, symbolic mechanism (Dennett, 1992; Gregory, 1972). While modally completed feature surfaces sometimes have been shown with fMRI to activate high level regions—such as the lateral occipital complex (LOC) to modally completed Kanizsa figures (Mendola, Dale, Fischl, Liu, & Tootell, 1999), independent of the contour (Stanley & Rubin, 2003), and associative areas in the dorsal pathway (the caudal region of the intrapariatal sulcus and LOC) to modally completed surface from the Craik-O'Brien-Cornsweet illusion (Perna, Tosetti, Montanaro, & Morrone, 2005), dependent on the edges—the activation might be due to feedforward activity from V1 and V2 (M.

Seghier et al., 2000), and even without invoking filling-in at all (Cornelissen, Wade, Vladusich, Dougherty, & Wandell, 2006). In general previous research showing early mechanism of perceptual completion invokes spreading of contour and features as a mechanism of integration, which could be achieved through horizontal connections in early visual cortex (Gilbert, 1992). For instance, the long-range integration seen in Kanizsa figure can be computed in early visual cortex via a cascade of activity percolating through a chain of lateral connections (Gilbert, Das, Ito, Kapadia, & Westheimer, 1996; S. Ullman, 1976). Due to the aperture problem in motion perception (Marr & Ullman, 1981), the emergent filling-in phenomenon relies on higher-order integration processes that extract complex global structures from local information.

It is debated whether modal completion and amodal completion involve the same contour and surface completion mechanisms (Kellman & Shipley, 1991; M. M. Murray, Foxe, Javitt, & Foxe, 2004). Given the large projecting angle between adjacent inducing edges in the current study (roughly 90°), the completed shape in the amodal condition might be more angular (i.e. closer to a corner) than the modal condition (Singh, 2004). In addition, the filling-in motion surface in the test region in the modal condition, compared with the amodal condition, might invoke neural representations of motion in early visual cortex (M. Meng et al., 2005). Despite these potential differences in contour and surface, similar filling-in MAE and similar attention modulatory effect were observed across amodal and modal conditions, suggesting that the mechanism responsible for the filling-in MAE is insensitive to the differences between modal and amodal completions, but sensitive to the emergent filling-in.

In our configuration, the temporal vector sum motion direction remained constant in the integrated condition but changed in the nonintegrated condition due to phase randomization. Could the larger MAE in the integrated (i.e. with filling-in) condition compared with the nonintegrated (i.e. without filling-in) condition be due to this difference? A subsequent control experiment shows that adapting motion with directions balanced over time is just as effective in producing MAE (see Figure 2-S2 in the Supplemental Material). Would completed motion, in particular in the modal completion condition, attract attention more than noncompleted motion, or could it be that observers simply imagine motion in the integrated condition but not in the nonintegrated condition?

These attention and imagination confounds, although valid concerns in previous studies on MAE (Snowden & Milne, 1997; Weisstein et al., 1977), could not explain our data. Subjectively, in the passive experiment where observers simply fixated on the central color changing cross, the motion in the nonintegrated condition actually appeared more attention capturing than that in the integrated condition. As for the imagination confound, a global net leftward or rightward motion was apparent in the nonintegrated condition and thus the suspected engagement of motion imagination would similarly apply. Objectively, in the attention experiment we deliberatively controlled the attention load of the central fixation task and made the motion stimuli completely task irrelevant. With a central fixation task continuously demanding for close inspection in both the integrated and nonintegrated conditions, attention was controlled; yet, for both the modal and amodal condition, a significant MAE difference between the integrated and nonintegrated conditions was still observed in the low attention load condition. Similarly for the imagination confound, with the central fixation task continuously loading observers' attention, imagination factor was controlled.

These results suggest that not all interpolation processes are the same; some rely more on low level mechanisms whereas others rely more on high level mechanisms, which potentially differentiates filling-in mechanisms. For example, filling-in at the blind spot or retinal scotomas is instantaneous, with the contour and feature to be filled in readily available from the surrounding retinotopic areas in V1/V2. Similarly, filling-in during troxler fading and stabilized retinal images could be achieved through such retinotopic mechanisms. Perceiving Kanizsa-like subjective figures usually requires integrating spatially disjoined parts to provide the contour and surface features to be filled in. If the integration process could be computed through an early mechanism, such as in Kanizsa figures, then the mechanism of perceptual completion may be mainly an early one (Shimojo, Kamitani, & Nishida, 2001). If, on the other hand, the integration process must be computed through a late mechanism, such as motion integration across apertures, then the mechanism of perceptual completion would rely heavily on later mechanisms. This framework might help to unite the vast filling-in phenomena in the literature. Further studies of the emergent filling-in phenomenon are likely to yield more insight regarding the mechanisms of filling-in.

Chapter 3

Frame-centered object representation and integration revealed by non-retinotopic priming and masking

Object identities (“what”) and their spatial locations (“where”) are processed in distinct pathways in the visual system, raising the question of how the “what” and “where” information is integrated. Because of object motions and eye movements, the retinotopically based representations are unstable, necessitating non-retinotopic representation and integration. A potential mechanism is to code and update objects according to their reference frames (i.e., frame-centered representation and integration). To test this mechanism, based on apparent motion, we have developed a psychophysical paradigm where two frames are shown in sequence and objects’ relative frame locations across frames are directly manipulated. We show that priming effects (the facilitation effect on processing of objects on the second frame from objects on the first frame) systematically depend on the relative frame locations of the prime and the target; similarly, backward masking effects (the detrimental effect on processing of objects on the first frame from objects on the second frame) also depend on the relative frame locations of the target and the mask. These findings thus reveal non-retinotopic, frame-centered priming and backward masking effects, and demonstrate that object representation in visual perception and memory is robustly coupled to its reference frame and continuously being updated through a frame-centered, location specific mechanism. By tagging objects to the reference frame, this frame-centered mechanism may underlie the crucial perceptual continuity and individuation of objects when multiple objects are moving, which is vital for proper and accurate actions toward moving objects.

3.1 Introduction

As objects in a typical scene are distributed on different locations, daily acts such as perception, navigation, and action require knowing both what the objects are and where they are located. Object identities and object locations are thought to be subserved

by distinct ventral and dorsal pathways, respectively (Ungerleider & Mishkin, 1982). Despite this distributed nature of identity and location representations, our phenomenal experience of the world is seamless and stable and at the same time retains the spatial relationships among objects. How does the visual system achieve this stable topographic perception, without confusing and mixing identity and location information? Topographic representations based on the retinal coordinate, though ubiquitous in the early visual system (Wandell, Dumoulin, & Brewer, 2007), are inherently susceptible to eye movements and object motions and thus are unstable (Melcher & Colby, 2008; Wurtz, 2008). Disruption of retinotopic (eye-centered) representation by eye movements can be compensated by perceptual (Melcher & Morrone, 2003) and neural mechanisms (d'Avossa et al., 2007; Duhamel, Colby, & Goldberg, 1992) tuned in spatiotopic (world-centered) coordinates, anchored in stable external world coordinates (Burr & Morrone, 2011; but see Cavanagh, Hunt, Afraz, & Rolfs, 2010; Gardner, Merriam, Movshon, & Heeger, 2008). Yet, disruption of retinotopic representation by object motions presents unique challenges that cannot be solved by spatiotopic mechanisms: the object representation is no longer anchored in world coordinates. Therefore, maintaining visual stability in face of object movements requires a more flexible, non-retinotopic coding and integration mechanism that can operate independently of eye movement contingent processes. Here we report that mobile objects in visual perception and memory are robustly represented and updated in a frame-centered manner.

Probing non-retinotopic representation by eye movements, though a frequently used approach, suffers from low temporal resolution because of the latency, duration, and variability of saccadic eye movements (Boi, Ogmen, Krummenacher, Otto, & Herzog, 2009), and hence is ill-suited for short-lived processes such as the updating of continuously moving reference frames. To circumvent this limitation, we have developed a psychophysical paradigm wherein two frames were presented in sequence and, because of apparent motion, perceived as one frame moving from one location to another (Figure 3-1A). The interstimulus interval (ISI) between the two frames could be as short as 0 ms, making it ideal for testing short-lived non-retinotopic processing between frames. This approach is reminiscent of recent research demonstrating object-based temporal-spatial information integration, such as color integration along a motion path (Nishida,

Watanabe, Kuriki, & Tokimoto, 2007; Watanabe & Nishida, 2007), form and motion integration (Boi et al., 2009) (but not tilt or motion aftereffects (Boi, Ogmen, & Herzog, 2011)) across apparent motion Ternus-Pikler displays, letter identity transfer through apparent motion (Kahneman et al., 1992), and motion and color (but not letters or digits) integration through attention tracking during continuous apparent motion (Cavanagh, Holcombe, & Chou, 2008). Our experiments build on these findings of object-specific updating process by investigating a more general frame-centered updating mechanism in perception and memory, wherein object representation is robustly coupled to its reference frame and updated through a frame-centered, location specific mechanism.

To directly probe frame-centered object representation and integration, we explicitly manipulated objects' relative frame locations across the two frames. We presented the target randomly above or below the fixation along the vertical meridian, and the two preceding (Experiments 1-4) or following (Experiments 5-7) objects on the left and right sides of the fixation along the horizontal meridian, thus ensuring the same retinal distance between the two preceding or following objects and the target. Critically, we also manipulated the target's relative location within its frame. If object representation is robustly coupled to its reference frame and continuously being updated in a frame-centered fashion, target performance should be strongly modulated by which one of the two preceding or following objects shared the same relative frame location as the target, even though they were equidistant from the target. That is, target performance should be facilitated if the target and the prime shared the same relative frame location than not, and should be impaired if the target and the mask shared the same relative frame location than not. Across the experiments, we observed non-retinotopic, frame-centered priming and backward masking effects, demonstrating that object representation is robustly coupled to its reference frame and continuously being updated through a frame-centered, location specific mechanism. By tagging objects to the reference frame, this novel frame-centered mechanism may underlie the crucial perceptual continuity and individuation of objects when multiple objects are moving, which is vital for proper and accurate actions toward moving objects. It may also be ideal for coping with disruption of retinotopic representation by eye movements, in which case objects are updated in reference with the world.

3.2 Methods

Observers and apparatus. Seventeen human observers (10 females) with normal or corrected-to-normal vision from the University of Minnesota community participated in Experiment 1, 19 (9 females) in Experiment 2, 17 (8 females) in Experiment 3, 13 (6 females) in Experiment 4A, 14 (7 females) in Experiment 4B, 12 (6 females) in Experiment 5, 13 (6 females) in Experiment 6, 12 (7 females) in Experiment 7 in return for money or course credit. The experiments were conducted in accordance with the IRB approved by the Committee on Human Research of University of Minnesota and the Declaration of Helsinki.

The stimuli were presented on a black-framed, gamma-corrected 22-inch CRT monitor (model: Hewlett-Packard p1230; refresh rate: 100 Hz; resolution: 1024 × 768 pixels) using MATLAB Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Observers sat approximately 57 cm from the monitor with their heads positioned in a chin rest in an almost dark room while the experimenter (ZL) was present.

Stimuli and procedure. To train observers to maintain stable fixation, before the main experiment, observers took part in a fixation training session, in which they viewed a square patch of black and white noise that flickered in counterphase (i.e., each pixel alternated between black and white across frames). Each eye movement during the viewing would lead to perception of a flash, and observers were asked to maintain fixation using the flash perception as feedback (i.e. to minimize the perception of flashes). Each experiment included 32 practice trials (in 1 block) and 320 experimental trials (in 5 blocks; for experiment 1 to 4) or 192 experimental trials (in 3 blocks; for experiment 5 to 7).

The basic structure of each trial in all the experiments involved a sequence of two frames against a black background: in the priming experiments, the first frame was presented on the center and the second frame was presented on one of four peripheral locations; in the masking experiments, the order was reversed. Because of apparent motion, the two frames were perceived as moving from one location to another. In the priming experiments (Experiment 1 to 4), the target was presented on the second frame and non-retinotopic priming effect from the stimuli on the first frame was gauged; in the masking experiments (Experiment 5 to 7), the target was presented on the first frame and

non-retinotopic backward masking effect from the stimuli on the second frame was measured.

In Experiment 1, after 800 ms presentation of a fixation cross (length: 0.23°; width: 0.08°; luminance: 80.2 cd/m^2), the first frame (size: 4° horizontal × 3° vertical; luminance: 40.1 cd/m^2 outlined by a white rectangle), centered on the fixation cross, was presented for 150 ms with two different prime letters (randomly selected from the following list: K, M, P, S, T, Y, H, and E; font: Lucida Console; size: 24; luminance: 80.2 cd/m^2) presented on two sides of the frame (center-to-center distance between letters: 3°) during the last 100 ms. After an interstimulus interval (ISI) of 30 ms, the second frame (shift from the first frame: ± 1.5° × ± 3°) was presented for 150 ms with a target letter (randomly selected from the same prime letter list) presented on one side of the frame but always right above or below the fixation (i.e. equidistant from the two prime letters) during the first 100 ms. Observers were asked to indicate whether the target letter, whose relative frame location corresponded to either the left or the right prime letter, was one of the two prime letters (i.e. either match or mismatch, a variant of match-to-sample task) as quickly and accurately as possible. When the target was one of the two prime letters, the target and the same prime letter could occupy either the same relative frame location or different locations. Experiment 2 was similar to Experiment 1 except that only one prime letter was presented during the first frame. In Experiment 3, the letters on the first frame (lowercase b and p with the order randomized) were task irrelevant and observers were asked to indicate whether the target letter (randomly selected from the following list: b, B, p, and P) presented on the second frame was a letter b/B or p/P, regardless of its case. Two extra fixation crosses (luminance: 30.1 cd/m^2) were added along the vertical meridian to aid target localization. The frame duration was 200 ms and the letter duration was 150 ms; other aspects were the same as Experiment 1A. Experiment 4A was the same as Experiment 1 except that the first frame was presented all through the offset of the second frame. Experiment 4B was the same as Experiment 4A except that two additional letters O were added during the presentation of the target letter.

In Experiment 5, after 1500 ms presentation of a fixation cross (length: 0.23°; width: 0.08°; luminance: 40.1 cd/m^2), the first frame (size: 4° × 3°; luminance: 40.1

cd/m^2 outlined by a white rectangle; shift from the central fixation: ± 1.3° × ± 2.6°) was presented for 200 ms with a target letter (either the letter N or its mirror-reversal: 0.6° × 0.6° ; line width: 0.04°; luminance: 70.2 cd/m^2) presented on one side of the frame but always right above or below the fixation during the last 10 ms. Immediately after the offset of the first frame (i.e. ISI of 0 ms), the second frame, centered on the fixation cross, was presented for 200 ms with two masks (weak mask: a circle of 0.16° × 0.16°; strong mask: a cross of 0.78° × 0.78° superimposed with three Xs of 0.55° × 0.78°, 0.78° × 0.78°, and 0.78° × 1.01° each; line width: 0.04°; luminance: 80.2 cd/m^2) presented on two sides of the frame (center-to-center distance between masks: 2.6°) during the first 30 ms. Observers were asked to indicate whether the target letter, whose relative frame location corresponded to either the left or the right mask, was the normal letter N or its mirror reversal as accurately as possible without emphasis on reaction time. In Experiment 6, a visual search task was used. Each frame was now enlarged (size: 8° × 4°; shift from the central fixation for the first frame: ± 2.4° × ± 4.8°). The search display consisted of 4 lines (center-to-center distance between lines: 1.6°; line length: 1.0°; line width: 0.08°; luminance: 60.2 cd/m^2); on half of the trials, all lines were rightward tilted 45° relative to vertical and on the other half of the trials, 3 were rightward tilted 45° and 1 was leftward tilted 45°. The search display was presented during the last 50 ms of the first frame. Two search display-like masks were presented on the second frame (weak mask: a filled square of 0.08° × 0.08°; strong mask: a vertical line of 0.98° superimposed centrally with an X of 0.98° × 0.98°; line width: 0.08°; luminance: 80.2 cd/m^2; center-to-center distance between masks: 2.4°). The mask display was presented during the first 50 ms of the second frame. Observers were asked to indicate whether there was a leftward titled bar on the search display as accurately as possible. In Experiment 7, the same search display and task were used as in Experiment 6. The weak mask now consisted of four right triangles (leg length: 0.7°; line width: 0.08°) with the hypotenuses rightward tilted 45°; the strong mask now consisted of an X (0.98° × 0.98°) and a vertical bar (length: 0.98°) touching right on the left side of the X.

3.3 Results

Experiment 1: A non-retinotopic, frame-centered priming effect. Two prime letters were presented first, one on each side of the fixation, followed by a target letter equidistant from the two prime letters and randomly above or below the fixation (Figure 3-1A). The task was to indicate whether the target letter was one of the two prime letters, i.e., either match or mismatch (a variant of match-to-sample task). Critically, although equidistant from the two prime letters, the target's relative frame location corresponded to either the left or the right prime letter, so that in the match trials, the two matched letters occupied either the same relative locations or different frame locations (Figure 3-1B). A repeated-measures ANOVA revealed that response times (RTs) differed significantly among these three conditions ($F(2, 32) = 19.48$, $P < 0.001$, $\eta_p^2 = 0.549$, Figure 3-1C). When the target matched one of the two prime letters, observers performed significantly faster when the matched letters occupied the same relative frame location than when they occupied different locations ($t(16) = -5.67$, $p < 0.001$, Cohen's $d = -1.38$), revealing a non-retinotopic, frame-centered priming effect. The error rate data were consistent with the RT data, with a lower error rate in the same frame location condition than the different frame locations condition for the matched letters ($t(16) = -2.04$, $p = 0.058$, $d = -0.50$). These results demonstrate that, even though object's location is task irrelevant, object's identity (e.g., shape) is bound to its reference frame and continuously being updated in a frame-centered manner in visual short-term memory (i.e. working memory).

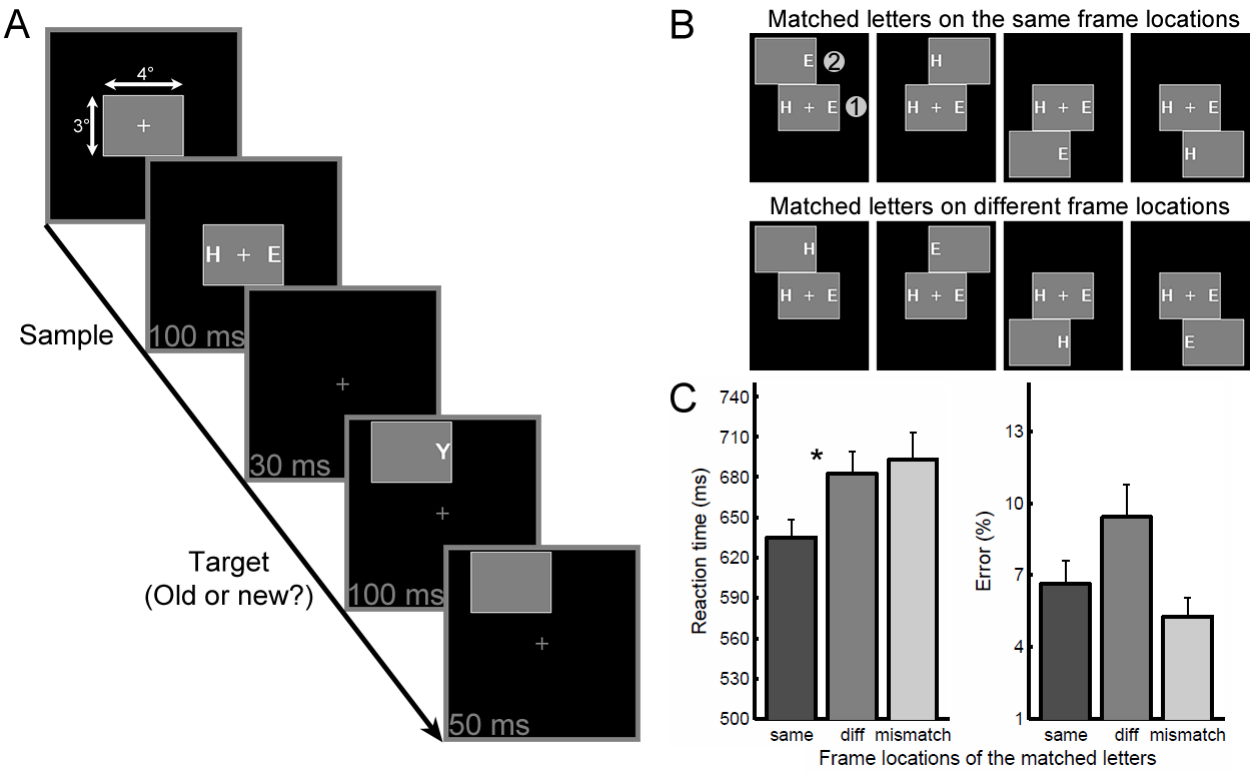


Figure 3-1. Procedure, stimuli, and results from Experiment 1. (A) Procedure: Observers fixated on a central cross throughout the experiment. In an apparent motion configuration, the first frame was presented at the center for 150 ms, with two prime letters presented during the last 100 ms along the horizontal meridian. After an interstimulus interval of 30 ms, the second frame was presented on one of four corners for 150 ms, with a target letter presented during the initial 100 ms along the vertical meridian. The task was to indicate whether the target was one of the two primes. (B) Sequential displays where the target matched one of the two prime letters: The first row illustrates displays where the target and the matched prime letter occupied the same relative frame location; the second row illustrates matched letters on different frame locations. (C) Results (n = 17): Responses were faster when the target and the matched prime occupied the same relative frame location than different frame locations. Error bars: standard error of the mean. *: statistically significant difference.

Experiment 2: Frame-centered priming without object competition and not due to response bias. An alternative account for this new priming effect is response bias based on the motion congruency between letters and frames: the matched letters on the same frame location moved in the same direction as the frame motion, whereas the

matched letters on different frame locations did not. This explanation is not likely because apparent motion directions tend to be unstable when two objects were followed by one object (Shimon Ullman, 1979). Nevertheless, to directly test this motion congruency account, we generated clear letter apparent motion direction by presenting only one letter during the first frame, followed by another letter on the second frame. This design also has an added benefit of allowing us to test whether encoding more than one object during the first frame is necessary for the frame-centered priming effect (presenting more than one object, as in Experiment 1, entails an intrinsic relationship and order among the objects, which might encourage the visual system to bind objects to their corresponding locations in working memory, even though location information is task irrelevant).

The letters on the two frames could either matched or mismatched, which the subjects were asked to indicate. Importantly, the relative frame locations of the two letters were also manipulated orthogonally, so that the letters could be on the same or different frame locations (Figure 3-2A). Both the frame-centered representation account and the motion congruency account predict better performance when the matched letters ("same" identities) were on the *same* relative frame location than on different locations. However, the motion congruency account also predicts better performance when the mismatched letters ("different" identities) were on *different* relative frame locations than on the same frame locations. A repeated-measures ANOVA on RTs revealed a significant interaction between these two factors ($F(1, 18) = 5.83$, $P = 0.027$, $\eta_p^2 = 0.245$, Figure 3-2B). Consistent with both accounts, observers performed significantly faster when the matched letters occupied the same relative frame location than when they occupied different locations ($t(18) = -3.46$, $p = 0.003$, $d = -0.79$). However, for the mismatched letters, there was no difference between the same relative frame location condition and the different locations condition ($t(18) = -0.07$, $p = 0.944$, $d = -0.02$). This result is thus inconsistent with the motion congruency account and provides further support for our frame-centered representation account by extending the frame-centered priming effect to displays imposing no object competition. The error rate data were consistent with the RT data, with a lower error rate in the same frame location condition than the different frame

locations condition for the matched letters ($t(16) = -2.20$, $p = 0.041$, $d = -0.50$) but not for the mismatched letters ($t(16) = 0.86$, $p = 0.399$, $d = 0.20$).

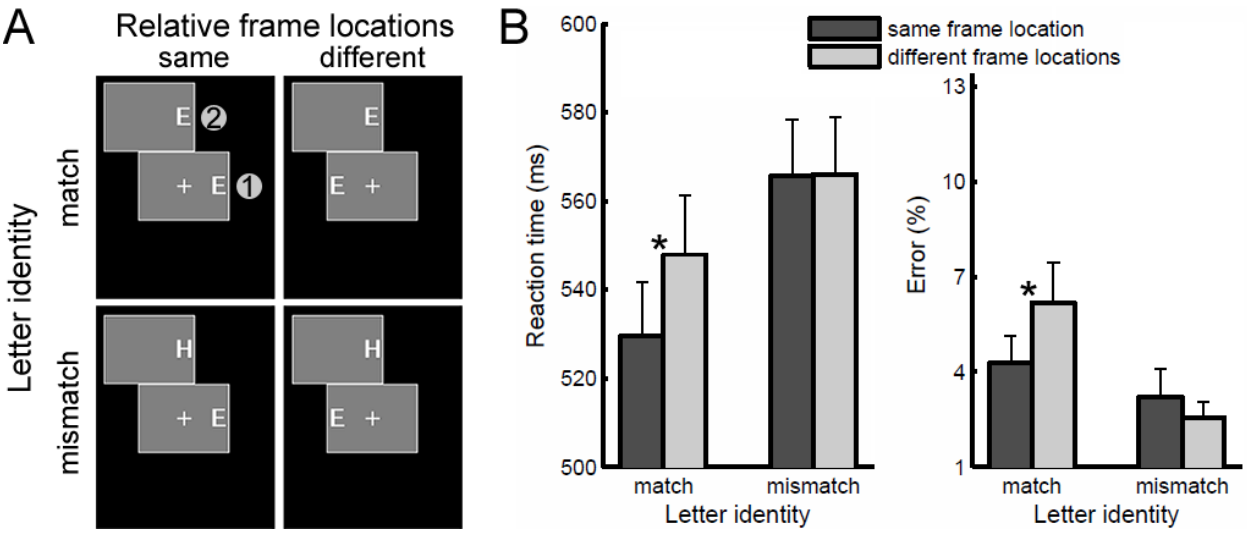


Figure 3-2. Stimuli and results from Experiment 2. (A) Stimuli: As Experiment 1, the fixation frame was presented first, and the task was to indicate whether the first letter and the second letter were the same or not. The first row illustrates sequential displays where the two letters matched; the second row illustrates mismatched letters. Letters on the first column occupied the same relative frame location; letters on the second column were on different frame locations. (B) Results ($n = 19$): Responses were faster for the same relative frame location condition than different relative frame locations condition only when the two letters matched. Error bars: standard error of the mean. *: statistically significant difference.

Experiment 3: Frame-centered priming without explicit memory demand and involving both low-level visual featural representation and high-level identity representation The purpose of Experiment 3 was two-fold. First, the match-to-sample task of Experiment 1 and 2 involved encoding of objects on the first frame into short-term memory and retrieval of stored objects in memory during the second frame, which may induce strategies such as allocating attention to a specific frame-centered location during a given trial (Kristjansson, Mackeben, & Nakayama, 2001; Olson & Gettner, 1995). To determine whether memory demand is necessary for the frame-centered

priming effect, in Experiment 3, we designed a perceptual task by asking new observers to directly categorize whether the target letter presented on the second frame was b/B or p/P, regardless of the case, while ignoring the two prime letters—lower case b and p with the order randomized—on the first frame (Figure 3-3A). Because both b and p were presented on the first frame and targets b/B and p/P were presented equally often on the second frame, the prime letters provided no information for the task. Critically, although the target was always of the same identity as one of the prime letters, the target and the matched prime letter could occupy either the same relative frame location or different locations. Thus, there were two factors in our design (Figure 3-3A): target case (lower vs. upper case) and prime letter identity on the same relative frame location as the target's (same vs. different). Second, the introduction of upper case targets afforded us to test alternative accounts based on low-level visual features and thus probe the level of representation in the frame-centered priming effect. In particular, for the letter pairs b-B and p-B in our study, B is physically equidistant from b and p (i.e., the overlap in low-level visual features between the letters is identical between the two pairs), but conceptually B is more similar to b than top (Lupyan, Thompson-Schill, & Swingley, 2010). Thus, accounts based solely on low-level visual features would predict similar performances when the target B shared the same relative frame location as b than as p.

A repeated-measures ANOVA on RTs revealed a significant interaction between target case and identity across frames ($F(1, 16) = 8.71$, $P = 0.009$, $\eta_p^2 = 0.352$), indicating that case consistency (i.e., visual featural overlap) plays a substantial role in frame-centered priming (Figure 3-3B). Observers performed significantly faster when the target and the matched prime occupied the same relative frame location than different relative frame locations, for both lower case targets ($t(16) = -7.76$, $p < 0.001$, $d = -1.88$) and upper case targets ($t(16) = -3.81$, $p = 0.002$, $d = -0.92$), indicating that frame-centered representation also retains a higher level, more abstract component that can survive case transformation. Because lower case p and upper case P were visually quite similar, we analyzed trials with target B separately to further assess whether low-level visual features alone can account for the frame-centered priming effect we observed. Although letter pairs b-B and p-B had the same visual feature similarities, RTs to targets B were much faster when targets shared the same relative location as letters b than letters p ($t(16) = -$

3.13, p = 0.006, d = -0.76), further demonstrating the frame-centered priming effect cannot be attributed to accounts based solely on low-level visual features or motion analysis and must also involve higher-level representations. No differences were found on the error rates.

Clearly, these results indicate that the frame-centered priming effect can occur without memory demand; object identity is thus bound to its reference frame and continuously being updated in a frame-centered manner not just in explicit visual short-term memory but also in implicit memory.

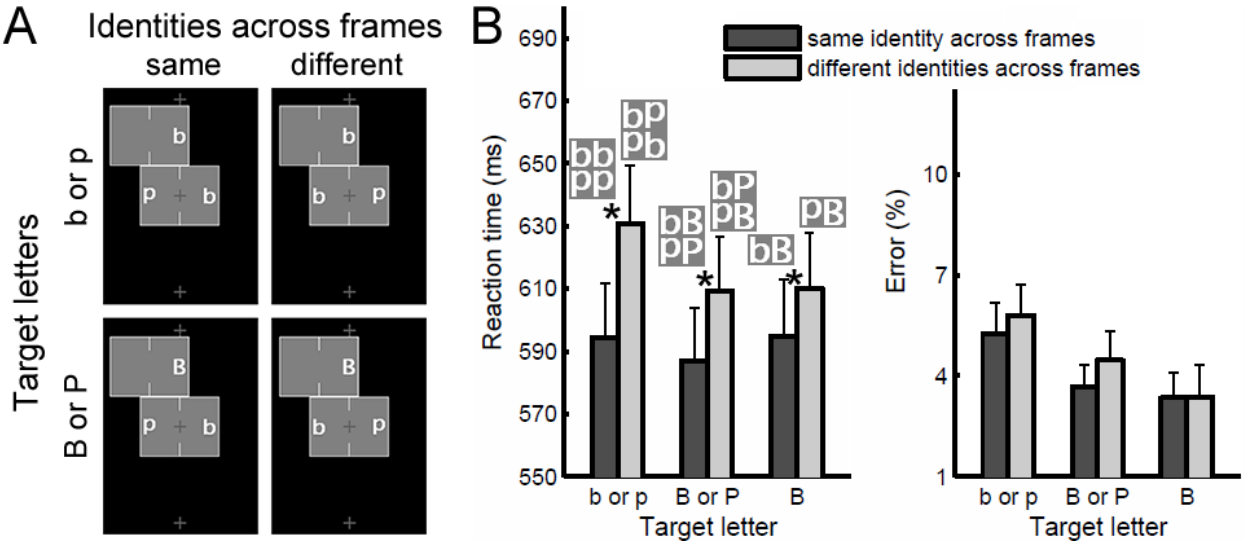


Figure 3-3. Stimuli and results from Experiment 3. (A) Stimuli: The fixation frame with two lower case letters, b and p (random order), was presented first, and the task was to categorize whether the letter at the end was b (B) or p (P), regardless of the case. The first row illustrates sequential displays where the target letters were lower case (b and p); the second row illustrates upper case targets (B and P). The first column shows sequential displays with the same letter identities on the same frame locations; the second column shows distinct letters on the same frame locations. (B) Results (n = 17): Responses were faster when the target letter shared the same relative frame location as the same identity prime letter, regardless of target cases. Error bars: standard error of the mean. *: statistically significant difference.

Experiment 4A and 4B: Frame-centered priming without apparent motion. Throughout Experiment 1 to 3, apparent motion was used a powerful tool to explore frame-centered object representation and updating. A natural question is the extent to which frame-centered representation and updating depends on motion mechanisms. To address this, in Experiment 4A, the apparent motion of frames was disrupted by continuously presenting the first frame until the offset of the second frame (Figure 3-4A); in Experiment 4B, the apparent motion of letters was also disrupted by presenting two letter Os during the target presentation (Figure 3-4C). Without the apparent motion of frames, the same pattern of frame-centered priming effect persisted (Figure 3-4B, same vs. different frame locations of the matched letters in RTs: $t(12) = -3.42$, $p = 0.005$, $d = -0.95$; error rate: $t(12) = -0.94$, $p = 0.368$, $d = -0.26$), as it did without the apparent motion of frames and letters (Figure 3-4D, same vs. different frame locations of the matched letters in RTs: $t(13) = -2.37$, $p = 0.034$, $d = -0.63$; error rate: $t(13) = -1.22$, $p = 0.242$, $d = -0.33$). These results thus demonstrate that apparent motion, exploited as a powerful tool in our study, is not necessary for frame-centered priming.

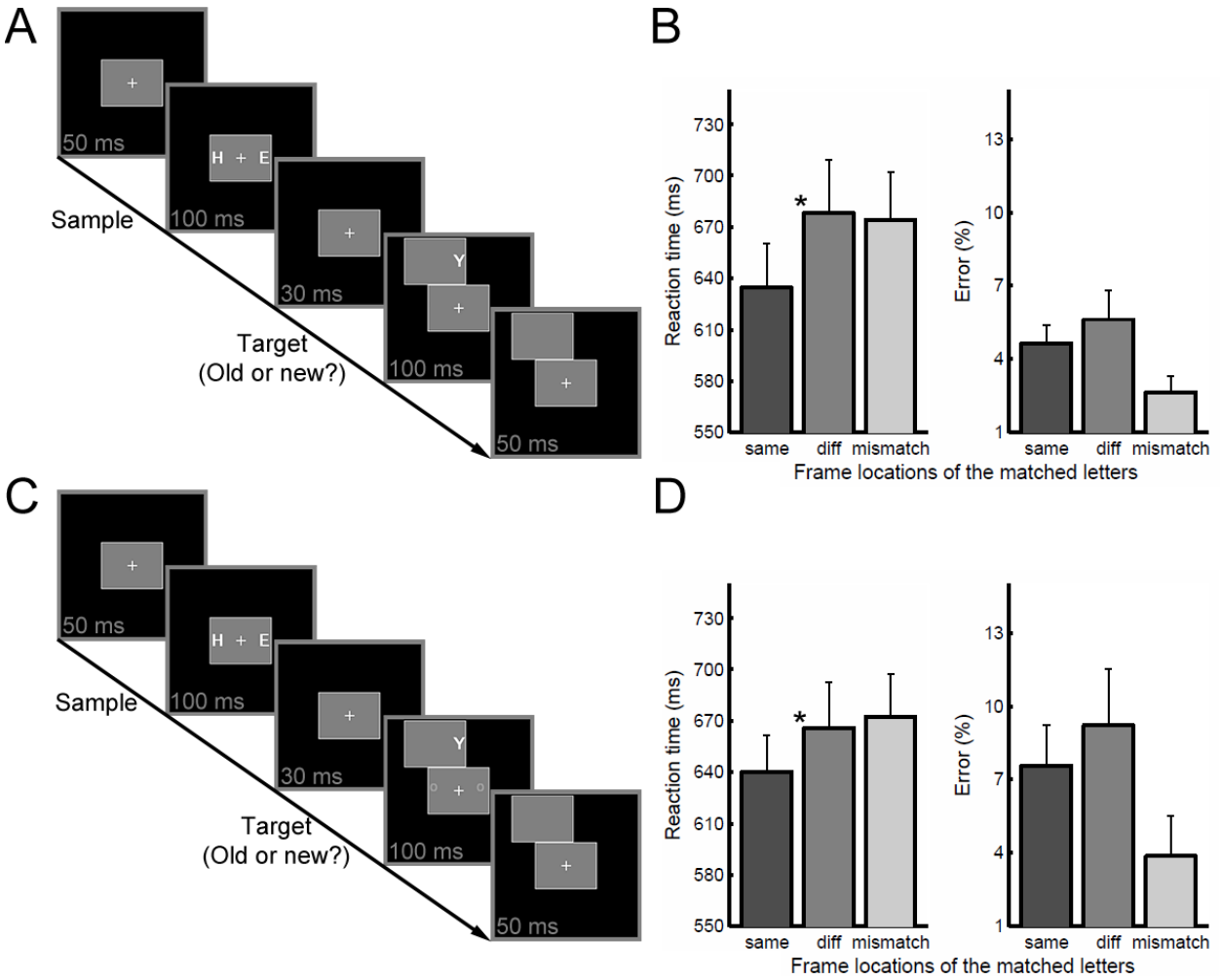


Figure 3-4. Procedure and results from Experiment 4A and 4B. (A, C) Procedure for 4A and 4B: In A, the same design was used as in Experiment 1, except that the first frame stayed until the offset of the second frame, thus disrupting the apparent motion of the frames; in C, the same design was used as in A with the addition of two letter Os during the presentation of the target, thus disrupting the apparent motion of both the frames and the letters. (B, D) Results for 4A (n = 13) and 4B (n = 14): In both cases, responses were faster when the target and the matched prime occupied the same relative frame location than different frame locations. Error bars: standard error of the mean. *: statistically significant difference.

Experiment 5: A non-retinotopic, frame-centered backward masking effect. With the retinotopic distances between the primes and the target controlled across the conditions compared, Experiment 1-4 demonstrate a novel frame-centered, non-retinotopic priming effect: target processing is affected by preceding prime objects in a

way that is determined by the relative frame-centered locations of the primes and the target. Such a frame-centered priming effect indicates that object representation is robustly coupled to its reference frame and continuously being updated through a frame-centered mechanism. Because prime objects were presented before the target object, in anticipation of the target, processes responsible for target categorization might have been activated before target presentation and thus act upon the prime objects. It remains unclear then whether such high-level anticipation processes are necessary for the frame-centered updating mechanism to operate. One way to address this issue is to present the target first, followed by two objects equidistant from the target, and assess whether backward masking of target processing from the two following objects is determined by the relative frame-centered locations of the target and the two masks. Because of the detrimental nature of backward masking, observers have no incentive to integrate objects across frames and will be actually better off not integrating across frames; this design thus also allows us to test the relative automaticity of frame-centered representation and updating.

In Experiment 5, a target letter, either N or mirror-N, was presented randomly above or below the fixation, immediately followed by two masks, a strong one and a weak one, one on each side of the fixation (Figure 3-5A). The task was to indicate whether the target letter was N or mirror-N, with emphasis on accuracy. Critically, although equidistant from the two masks, the target's relative frame location corresponded to that of either the left or the right mask, depending on the location of the target frame (first frame). Thus, the target could occupy the same relative frame location as the strong mask or the weak mask (Figure 3-5B). Figure 3-5C shows that observers performed both significantly faster and more accurately when the target was followed by a weak mask than by a strong mask on the same relative frame location (RTs: $t(11) = -2.54$, $p = 0.028$, $d = -0.73$; error rates: $t(11) = -2.80$, $p = 0.017$, $d = -0.81$), revealing a non-retinotopic, frame-centered backward masking effect.

These results demonstrate that, even when it is detrimental to do so, object representation is strongly and obligatorily coupled to its reference frame, with masking effects determined by the relative frame-centered locations of the target and the masks. Moreover, because targets were presented before the masks, the frame-centered masking

effect demonstrates that frame-centered updating can operate without high-level anticipation processes and can be based on low level, relatively automatic processes. In sum, frame-centered backward masking extends frame-centered presentation and updating from the memory realm to the perception realm.

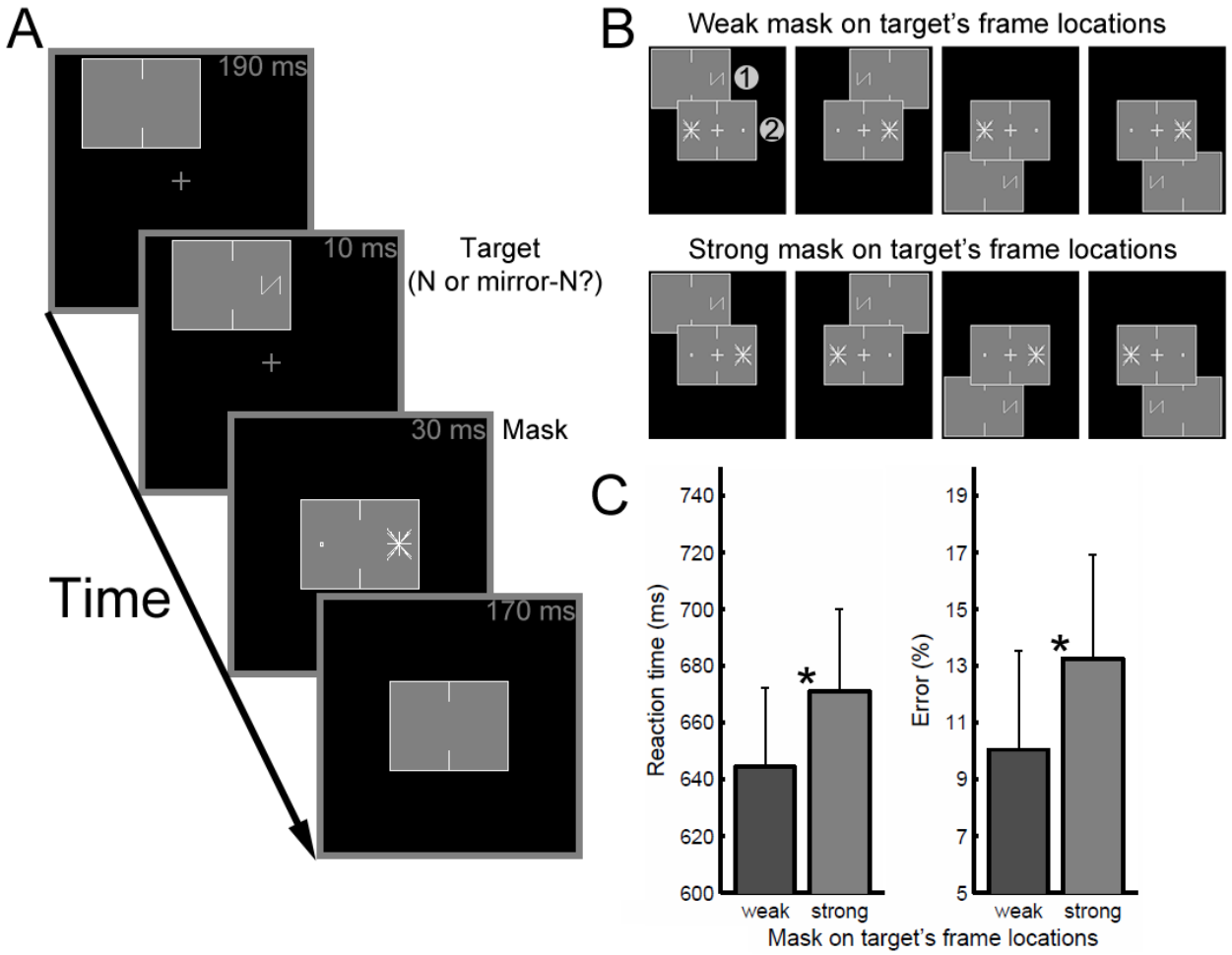


Figure 3-5. Procedure, stimuli, and results from Experiment 5. (A) Procedure: The first frame was presented either on the upper left, upper right, lower left or lower right corner for 200 ms, during the last 10 ms of which a target letter was presented either within the left or the right side of the display but always right above or below the fixation. Immediately after the first frame, the second frame was presented at the center for 200 ms, during the initial 30 ms of which two masks, one strong and one weak, were presented. The task was to indicate whether the target was N or mirror-N. (B) Sequential displays of the target and the masks: The mask on the same relative frame location as the target could be a weak one (first row) or a strong one (second row). (C) Results (n = 12):

Responses were faster and more accurately when the target was followed by a weak mask than by a strong mask on the same relative frame location. Error bars: standard error of the mean. *: statistically significant difference.

Experiment 6: Frame-centered backward masking due to perceptual sensitivity, not bias. How does frame-centered, non-retinotopic backward masking affect perceptual sensitivity and response bias? To address this, we employed a detection task on visual search, with signal (a left tilted bar) present on half of the trials and absent on the other half (Figure 3-6A). Observers were asked to detect, as accurately as possible, whether the search display on the first frame (randomly on one of four corners) contained an odd, left-tilted bar. The search display was followed by two mask displays, one strong and one weak, on the second frame centered on the fixation. The mask on the same relative frame location as the search display could be the strong one or the weak one, with the mask strength determined by physical energy. Using signal detection theory (Green & Swets, 1966), perceptual sensitivity was measured by d', based on hit rate and false alarm rate (i.e., Z(HR) - Z(FAR)), with higher d' meaning better perceptual sensitivity; response bias was measured by c, based on hit rate and false alarm rate (i.e., -(Z(HR) + Z(FAR))/2), with lower absolute c meaning lower bias towards one type of response regardless of the stimuli. Using d' and c, we found that observers' perceptual sensitivity was higher when the target was followed by a weak mask than by a strong mask ($t(12) = 3.41$, $p = 0.005$, $d = 0.95$), but response bias was not affected by the type of mask ($t(12) = -0.51$, $p = 0.617$, $d = -0.14$, Figure 3-6B). These results thus indicate that frame-centered backward masking acts by modifying perceptual sensitivity rather than response bias.

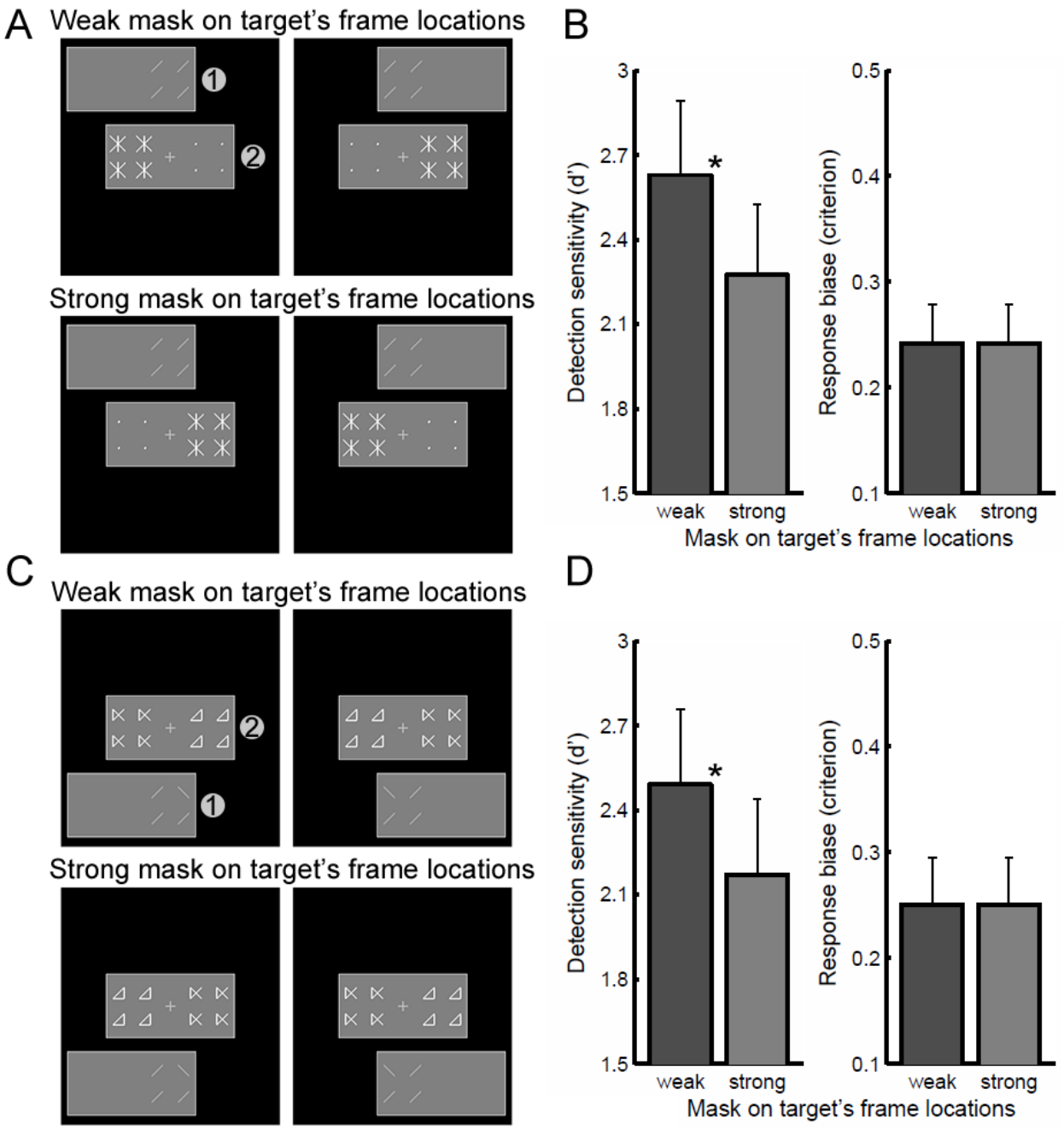


Figure 3-6. Procedure and results from Experiment 6 and 7. (A, C) Sample stimuli for Experiment 6 and 7: A visual search detection task was used. The task was to detect whether the search display on the first frame (on one of four corners) had an odd, left-tilted target bar or not. The mask on the same relative frame location as the search display on the second frame (centered on the fixation) could be a weak one or a strong one. (B, D) Results for Experiment 6 (n = 13) and 7 (n = 12): In both cases, perceptual sensitivity

but not response bias was higher when the target was followed by a weak mask on the same frame-centered location than a strong mask. Error bars: standard error of the mean. *: statistically significant difference.

Experiment 7: Form integration during frame-centered backward masking resulting in a non-retinotopic object superiority effect. The results of Experiment 6 prompted us to examine a more intriguing case of cross frame integration: can form information be integrated across frames in a frame-centered fashion (Hayhoe, Lachter, & Feldman, 1991)? To test this, in Experiment 7, we extended the object superiority effect (Weisstein & Harris, 1974), in which adding a non-informative part to each search element can significantly enhance visual search performance, into a form of non-retinotopic object superiority effect. In our configurations (Figure 3-6C), the search displays were again tilted bars, which shared the same relative frame location with either four triangles, which, if integrated with the targets, would result in a traditional retinotopic object superiority effect, or four knots matched in the number of lines with the triangles, which, if integrated with the targets, would produce no object superiority effect. Thus, form integration, if it occurs, would render the four triangles mask as a relatively weak mask and the four knots as a strong mask. Strikingly, the pattern of results paralleled that of Experiment 6 (Figure 3-6D): observers' perceptual sensitivity was higher when the target was followed by four triangles (weak mask) than by four knots (strong mask), $t(11) = 3.85$, $p = 0.003$, $d = 1.11$, with response bias unaffected by the type of mask, $t(12) = -0.08$, $p = 0.937$, $d = -0.02$. These results indicate that certain form information can be integrated across frames in a frame-centered fashion.

3.4 Discussion

Using a moving frame technique, we show that objects are robustly coupled to their reference frames in visual perception and memory, resulting in non-retinotopic, frame-centered priming and masking. Specifically, the current series of experiments show that frame-centered representation and integration is 1) pervasive (in working memory, implicit memory, and perception), 2) relatively automatic (without active memory involvement, object competition, and top-down anticipation; when objects are task irrelevant; and even when it is detrimental to the task), and 3) can be retrospective

(affecting processing of stimuli in the preceding frame). Frame-centered representation and integration is highly adaptive for perception, navigation, and action, enabling object information to be continuously updated relative to a reference frame as the frame moves.

Frame-centered representation and integration sheds new light on our understanding of object and space integration. In both non-human primates and humans, information specifying object identity, such as shape and color, is processed primarily by the ventral pathway, whereas information specifying object location and spatial relations among objects is processed primarily by the dorsal pathway (Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). This segregation of identity and space processing can extend to the prefrontal cortex (Wilson, Scalaidhe, & Goldman-Rakic, 1993). Integrating information about object identity with information about object location is thus vital for perception, memory, navigation, and action. Although identity information is relatively invariant and free from reference frames, space information is intrinsically relative and must be specified in a given coordinate. This frame-sensitive nature of space coding is surprisingly overlooked in previous research studying identity and space integration (Cichy, Chen, & Haynes, 2011; Rao, Rainer, & Miller, 1997; Schwarzlose, Swisher, Dang, & Kanwisher, 2008; Voytek, Soltani, Pickard, Kishiyama, & Knight, 2012), with space defined only in the retinotopic coordinate. Retinotopic coordinate, though widespread in the early visual system, is insufficient because of the ubiqutous eye movements and object movements in natural vision. In our study, objects were presented in sequence, with the retinal distance between the two preceding or following objects and the target kept identical, allowing us to explicitly examine the effect of reference frame on identity and location integration. Our results show that object identity is robustly coupled to its reference frame, resulting in frame-centered priming and masking. These findings thus reveal the flexible nature of identity and space integration. An essential next step is to understand how this frame-centered property is coded in the brain, which is vital for fully understanding how identity and space is integrated to support our many daily acts.

Frame-centered representation implies that objects are explicitly coded in reference to the frame, complementing previous research showing that features and parts are coded in reference to the object (Marr & Nishihara, 1978; Olshausen, Anderson, &

Van Essen, 1993). Unit recordings in monkeys, for example, show that some neurons in V4 are tuned for contour elements at specific locations relative to a larger shape (Pasupathy & Connor, 2001), and, in a delayed sample-to-target saccade task, some neurons in the Supplementary Eye Field are selective for a particular side of an object independent of the retinal location of the object (Olson & Gettner, 1995). These findings in monkeys are consistent with observations from hemineglect patients (who are impaired in orienting spatial attention to information on the contralateral side of the brain lesion): these patients neglect the contralesional side of an object regardless of its position in egocentric coordinates (Driver, Baylis, Goodrich, & Rafal, 1994; Tipper & Behrmann, 1996). Conversely, normal humans are able to direct attention toward a specific location within an object when the object changes its retinal location (Kristjansson et al., 2001; Umilta, Castiello, Fontana, & Vestri, 1995). These findings thus suggest that, when the task explicitly requires it, encoding of objects to the reference frame is possible (Olson, 2003). Our results extend these findings by showing that object representation is robustly coupled to its reference frame in a way that is independent of eye movements and top-down cueing, and more importantly, when the frame moves, such frame-centered representation can be continuously updated through a frame-centered, location specific mechanism.

Frame-centered representation and integration leads to a new theoretical understanding of perceptual continuity and object representation. Previous research suggests that non-retinotopic information integration during motion can be object-specific. For instance, perceiving an object is suggested to induce a temporary, episodic representation in an object file. Re-perceiving that object automatically retrieves its object file, so that perception is enhanced when all of the attributes match those in the original object file and impaired if some attributes are changed (Treisman, 1992). Similar mechanism has been suggested for lower level feature processing, such as color integration (Nishida et al., 2007; Watanabe & Nishida, 2007). What can be integrated in such an object-based manner remains unclear. During continuous apparent motion, it is shown that low level features such as motion and color but not high level representations such as letters or digits can be integrated through attention tracking (Cavanagh et al., 2008). During apparent motion Ternus-Pikler displays, however, it is shown that the

opposite is true, that form and motion can be integrated across frames (Boi et al., 2009) but tilt and motion aftereffects cannot transfer across frames (Boi, Ogmen, et al., 2011). The differences in the specific paradigms used notwithstanding, our results show that both low level (e.g. orientation) and high level (e.g. form) information can actually be integrated across frames. Most importantly, previous research on mobile updating has been limited to a fixed relative position and thus unable to reveal the effect of reference frame; by directly manipulating the relative frame locations, our results reveal that such mobile integration relies on a frame-centered, location specific mechanism. These results naturally lead to the concept of object "cabinet", in the sense that the objects (the files) within the reference frame (the cabinet) are orderly coded in their relative spatial relations to the frame. This object cabinet coding may be the foundation of scene perception, which involves coding the spatial and contextual relationships among objects and their frames. A hierarchical file and cabinet system might thus underlie our ability to represent multiple objects into a scene and help to keep track of multiple moving objects.

Frame-centered priming and masking effects also fuel our understanding of the spatiotemporal characteristics of object representation. In general, information integration during motion is inherently non-retinotopic, involving dynamic spatiotemporal updating. Our results show that the effect of such integration is bi-directional: object information on the first frame affects processing of information on the second frame (i.e. priming) and vice versa (i.e. masking). This observation links two prominent theoretical proposals on mobile information integration: object files (Kahneman et al., 1992) and object updating (Enns, Lleras, & Moore, 2010), which are usually discussed without much reference to each other. Object files theory is primarily concerned with priming effect (e.g. object-specific priming) and suggests that mobile objects are tagged in object files that collect, integrate, and update information of individual perceived objects as they move. Object updating theory on the other hand is primarily concerned with masking effect (e.g. object-substitution masking) and suggests that mobile objects are continuously been updated as long as the objects are perceived as the same ones moving rather than new ones emerging (Enns & DiLollo, 1997; Lleras & Moore, 2003). Based on these two notions, we suggest that when changes to the mobile object are none or minor (e.g. in Experiment 1, from a letter E to another E), object information is integrated into an object

file, which is updated continuously so that the perceptual sameness of the object is maintained and perceptual stability achieved (e.g. priming). Yet, when changes to the mobile object are significant (e.g. in Experiment 5, from a letter N to a noise patch), the object file is destructed, resulting in a change of perceptual sameness and thus a breakdown in perceptual stability (e.g. masking). This construction, updating, and destruction scheme captures the priming and masking effects reported.

More generally, given that frame-centered priming persists in spite of disrupted apparent motion, our object cabinet framework is not necessarily limited to integration across motion, and can be seen as an extension of earlier concepts in memory research, including schemata (Biederman et al., 1982; Hock et al., 1978; Mandler & Johnson, 1976), scripts (Schank, 1975), and frames (Minsky, 1975), wherein objects and their contextual information are bound in representation. For example, according to Minsky, encountering a new situation evokes the retrieval of a particular structure from memory—called a frame—so that the old frame can be adapted to accommodate new information. However, unlike frames, which are conceptualized as representations of stereotyped situations (e.g. a living room, a birthday party), our object cabinet concept is more general, including not only familiar situations but also novel objects and novel situations, and constrains not only higher level memory processes but also lower-level visual perception. Thus, the object cabinet framework has important implications for both memory and perception. For instance, non-retinotopic integration across successive, moving frames (transframe integration) is functionally similar to integration across locations on the retina during eye movements (transsaccadic integration) (Irwin, 1996; Melcher & Colby, 2008). It is possible that a similar frame-centered representation could account for transsaccadic integration such that information is integrated across eye movements in a frame-center, location specific manner. Thus, the frame-centered mechanism can cope with changes in retinotopic representation induced by object movements by updating the object's representation in reference with the frame, and may similarly cope with changes in retinotopic representation induced by eye movements by updating objects in reference with the world. Such a non-retinotopic frame-centered updating mechanism would be advantageous for coping with ubiquitous object

movements and eye movements and holds promise as a powerful and parsimonious account for maintaining visual stability in our everyday vision and action.

Chapter 4

Frame-centered object representation supports flexible exogenous visual attention across translation and reflection

Unexpected changes or sudden movements in the visual environment usually attract visual attention in an automatic fashion, known as exogenous attention. Decades of research has shown that exogenous attention is rigid, based on the retinotopic locations of the salient events. To the extent that salient events in the natural environment usually form specific spatial relationships with surrounding contexts and are dynamic, adaptive exogenous attention mechanisms should be responsive to statistical and structural regularities. A potential mechanism is one based on a moving reference frame, with exogenous attention being dynamically attracted to a relative, frame-centered location. To test this non-retinotopic mechanism, we presented two frames in succession, forming apparent translational motion, or in mirror reflection, with an uninformative, transient cue presented at one of the item locations in the first frame. Despite that the cues are presented in a spatially separate frame, in both translation and mirror reflection, human performance is enhanced when the target in the second frame appears on the same relative location as the cue location than on other locations. In other words, induced by frame-centered representations of the cues, exogenous attention cueing generalizes across translation and mirror reflection, revealing a structural, frame-centered mechanism in exogenous attention. As an extension of object file theory, these findings are explained in an object cabinet framework.

4.1 Introduction

Our senses are constantly flooded with streams of information, necessitating mechanisms of attentional gating and selection. Attention allocation is mainly achieved through two distinct types (Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Nakayama

& Mackeben, 1989; Theeuwes, 2010): a flexible, endogenous (i.e. top-down, goal-directed) mechanism whereby one voluntarily selects information of interests for further processing, and a rigid, exogenous (i.e. bottom-up, stimulus-driven) mechanism whereby one's attention is involuntarily drawn to the location of salient information. Exogenous attention drawn to salient events is adaptive, allowing quick orienting of attention toward urgent events and important objects, e.g. a snake moving close. Decades of research suggests that exogenous attention, being attracted to the location of the salient event, is rigid (Carrasco, 2011). In support of this rigidity characteristic of exogenous attention, the paradigmatic task in the last few decades has been the Posner cueing paradigm (Posner, 1980), where a transient, uninformative cue is randomly flashed on one of the item locations, and attention is found to be involuntarily drawn to the cue location.

To the extent that urgent and important information in our daily visual environment is usually dynamic (e.g. a dashing arrow or a jumping lion), an exogenous attention mechanism rigidly locked to the original cue location may be maladaptive. Moreover, because urgent and important visual cues in the natural world usually form specific spatial relationships with their contexts, an exogenous attention mechanism blind to such statistical and structural regularities also seems inefficient. Therefore, coping with salient and dynamic visual cues would benefit from a more flexible exogenous attention mechanism incorporating environmental regularities, such as one based on a reference frame that can be continuously updated when it moves or when the eyes move, with exogenous attention being attracted to a relative, frame-centered location. Such a possibility is hinted by studies showing attention orienting to a remote, frame-centered location when the cues were *predictive* of the target location in the test frame and thus with a strong component of endogenous attention (Boi, Vergeer, Ogmen, & Herzog, 2011; Umilta et al., 1995). To directly test the flexibility of exogenous attention without being confounded by informative cues, here we adopted the logic of the Posner cueing paradigm with *uninformative* cues and tested exogenous attention across translational apparent motion. Moreover, translational flexibility may reflect a frame-based spatial mechanism, such as an attentional spotlight mounted on a moving frame, or a frame-based structural mechanism where statistical and structural regularities automatically guide attention, that is, a frame-centered mechanism. To distinguish between these two

possibilities, we capitalize on mirror image confusion—a phenomenon in object recognition that has fascinated artists, philosophers, cosmologists, and scientists for centuries (Corballis & Beale, 1976), in which many species, monkeys and humans included, confuse images with their bilateral, left-right mirror-reflection counterparts (i.e. enantiomorphs)—and test exogenous attention across mirror image reflection.

4.2 Methods

Observers and apparatus. Thirteen observers (5 females) participated in Experiment 1, 10 (8 females) in Experiment 2, 8 (6 females) in Experiment 3A, and 8 (6 females) in Experiment 3B. One observer was the author ZL; others, naïve to the study, were drawn from the University of Minnesota community and received course credits or money for their time. All had normal or corrected-to-normal vision, and signed an Institutional Review Board approved consent form.

The stimuli were presented on a black-framed, gamma-corrected 22-inch CRT monitor (model: Hewlett-Packard p1230; refresh rate: 100 Hz; resolution: 1024 × 768 pixels) using MATLAB Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Observers sat approximately 57 cm from the monitor with their heads positioned in a chin rest in an almost dark room while the experimenter (ZL) was present.

Stimuli and procedure. To train observers to maintain stable fixation, before the main experiment, observers took part in a fixation training session, in which they viewed a square patch of black and white noise that flickered in counterphase (i.e., each pixel alternated between black and white across frames). Each eye movement during the viewing would lead to perception of a flash, and observers were asked to maintain fixation using the flash perception as feedback (i.e. to minimize the perception of flashes). After the fixation training session, observers were exposed to the two frame apparent motion configurations and indicated being able to perceive apparent motion before proceeding to the experimental tasks. There were 96 practice trials (in 2 blocks) and 576 experimental trials (in 12 blocks) in Experiment 1, 64 practice trials (in 1 block) and 256 experimental trials (in 4 blocks) in Experiment 2, 128 practice trials (in 2 blocks) and 384 experimental trials (in 6 blocks) in Experiment 3A, and 216 practice trials (in 6 blocks) and 216 experimental trials (in 6 blocks) in Experiment 3B.

The basic structure of each trial in all the experiments involved a sequence of two frames against a black background: the first frame with an exogenous attention cue was presented randomly on the left or the right of the fixation, followed by the second frame with a visual search display also randomly on the left or the right. When the two frames were on different sides of the fixation, because of apparent motion, they were perceived as a single frame moving from one location to the other (as subjectively confirmed by each observer).

In Experiment 1, after 800 ms presentation of a fixation dot (diameter: 0.39°; luminance: 30.1 cd/m^2), the first frame consisting of a central square (size: 3.1°; luminance: 40.1 cd/m^2 outlined by a white rectangle; square center to fixation distance: 1.85°) and two flanking squares (size: 2°; luminance: 40.1 cd/m^2; center-to-center distance between the central square and each flanker square: 3.16°) was presented either on the left or the right of the fixation for 200 ms, with a dot cue (size: 0.3°; color: randomly selected from blue, yellow, and cyan for each trial) inserted randomly on one of six locations within the central square (center-to-center distance between cues and the square center: 1.1°) from 80 ms to 120 ms. After an interstimulus interval (ISI) of 60 ms, the second frame was presented either on the same side of the fixation or on the other side of the fixation for 200 ms with six orientation bars (length: 0.5°; width: 0.1°) presented during the first 80 ms. Three bars were green, with two tilted 30° either clockwise or counterclockwise relative to vertical and one titled 30° to the opposite side; the remaining three bars were red, with two vertical and one tilted 30° either clockwise or counterclockwise. Observers were asked to indicate whether the non-vertical red bar was tilted clockwise (right) or counterclockwise (left) as quickly and accurately as possible.

In Experiment 2, after 1250 ms presentation of a fixation dot (diameter: 0.39°; luminance: 30.1 cd/m^2), a white disk (diameter: 1.2°; luminance: 80.2 cd/m^2) was presented randomly on one of four locations making up four corners of an isosceles trapezoid (center-to-center distance between the two left cues: 4°; between the two right cues: 2°; trapezoid height: 2°; trapezoid centroid to fixation distance: 2.5°) either on the left or the right of the fixation for 30 ms. The cue was immediately followed by four disks of the same size and luminance on the same side of the fixation for 10 ms, which again were immediately followed by four disks of the same size and luminance either on

the same side of the fixation or on the other side of the fixation for 40 ms, during which four black orientation bars (length: 0.8°; width: 0.125°) were superimposed on the disks. Three bars were horizontal and vertical (two horizontal and one vertical or two vertical and one horizontal), with one tilted 15° either clockwise or counterclockwise relative to vertical. Observers were asked to indicate whether the tilted bar was tilted clockwise (right) or counterclockwise (left) as accurately as possible (without emphasis on reaction time).

In Experiment 3A, after 750 ms presentation of a fixation dot (diameter: 0.39°; luminance: 30.1 cd/m^2), the first frame of a shell-like shape outlined by a light gray line (width: 0.08°; luminance: 70.2 cd/m^2) and consisting of an isosceles trapezoid (length of the longer base: 4.5°; length of the shorter base: 2.5°; trapezoid height: 2.2°; trapezoid centroid to fixation distance: 2.4°; luminance: 40.1 cd/m^2) and an arc (contained within an imaginary rectangle of 1.25° × 4.5°; luminance: 40.1 cd/m^2) was presented for 200 ms, with a dot cue (size: 0.3°; color: randomly selected from blue, yellow, and cyan for each trial) inserted randomly on one of four locations within the shell from 140 ms to 180 ms. The four locations made up four corners of an imaginary isosceles trapezoid (length of the longer base: 3.24°; length of the shorter base: 1.54°; trapezoid height: 1.72°) with the longer base partially overlapping that of the isosceles trapezoid of the shell frame. After an ISI of 60 ms, the second frame was presented for 200 ms with four orientation bars (length: 0.5°; width: 0.1°) presented during the first 80 ms. Two bars were green, with one tilted 30° clockwise and one one tilted 30° counterclockwise relative to vertical; the remaining two bars were red, with one vertical and one tilted 30° either clockwise or counterclockwise. Observers were asked to indicate whether the non-vertical red bar was tilted clockwise (right) or counterclockwise (left) as quickly and accurately as possible

In Experiment 3B, after 1000 ms presentation of a fixation dot (diameter: 0.39°; luminance: 30.1 cd/m^2), the first frame of a shell-like shape outlined by a red line (width: 0.08°; luminance: 26.7 cd/m^2) and consisting of an isosceles trapezoid (length of the longer base: 4.8°; length of the shorter base: 2.2°; trapezoid height: 2.2°; trapezoid centroid to fixation distance: 2.4°; luminance: 40.1 cd/m^2) and an arc (contained within an imaginary rectangle of 1.25° × 4.8°; luminance: 40.1 cd/m^2) was presented for 100 ms either on the left or the right of the fixation, with a horizontal, white bar (size: 0.16° ×

1.0°; luminance: 80.2 cd/m^2) inserted randomly on one of three locations within the shell during the last 40 ms. The three locations made up three corners of an imaginary isosceles triangle (length of the base: 3.0°; triangle height: 1.55°) with the base partially overlapping that of the isosceles trapezoid of the shell frame. After an ISI of 30 ms, the second frame was presented for 100 ms either on the same location as the first frame or on the opposite location, with three horizontal, white bars presented during the first 40 ms, which were immediately followed by the same three bars plus two tilted bars (one tilted 30° clockwise and one one tilted 30° counterclockwise relative to vertical; one just above its abutting horizontal bar and one below its abutting horizontal bar) and one vertical bar (randomly above and below its abutting horizontal bar) for 60 ms. The size of the vertical or tilted bar close to the fixation was 0.08° × 0.5°; the sizes of the two farther away ones were 0.08° × 0.67°. Observers were asked to indicate whether the vertical bar was above or below its abutting horizontal bar as quickly and accurately as possible.

4.3 Results

Flexible exogenous attention across translation. Figure 4-1A illustrates the basic structure of each trial in Experiment 1: two frames were presented in sequence, first a cue frame containing a transient, uninformative cue, followed by a target frame containing a target and five distractors. The two frames appeared equally often on the same side of the fixation (retinotopic) or on different sides (translation), making the cue (and the cue frame) uninformative of the target *frame* position. When the two frames were on different sides, they were perceived as one frame moving from one location to another. The cue was randomly flashed at one of six locations within the cue frame, so that the cue was uninformative of the *target* location. The target, a non-vertical red bar tilted clockwise or counterclockwise, randomly occupied one of six locations within the target frame, with the remaining five locations occupied by five distractors. Observers were instructed to maintain fixation on a central dot (all participated in a fixation training session before the main experiments; see Methods) and asked to discriminate and report the target orientation as quickly and accurately as possible. Note that the cue provided no information for the future frame position or the target location, fulfilling a critical criterion of exogenous attention cueing. Therefore, if target performance was facilitated

for the same relative frame locations compared with different relative frame locations, this would constitute evidence for flexible exogenous attention across translation.

A repeated-measures ANOVA revealed that, regardless of retinotopic or translation display conditions (insignificant interaction of display conditions and cue conditions, $F(1, 12) = 1.06$, $P = 0.32$), response times (RTs) were significantly faster when the cue and the target coincided on the same or same relative location (valid) than on different or different relative locations (invalid; $F(1, 12) = 16.92$, $P = 0.001$, Figure 4-1B). While the superior performance in the valid trials than the invalid trails for the retinotopic display condition ($t(12) = -3.55$, $P = 0.004$) is consistent with classic findings from the Posner cueing paradigm, the superior performance in the valid trials than the invalid trails for the translation display condition ($t(12) = -2.54$, $P = 0.026$) reveals a novel flexible exogenous attention mechanism, in which exogenous attention transfers across translation. No significant differences were found in the accuracy data.

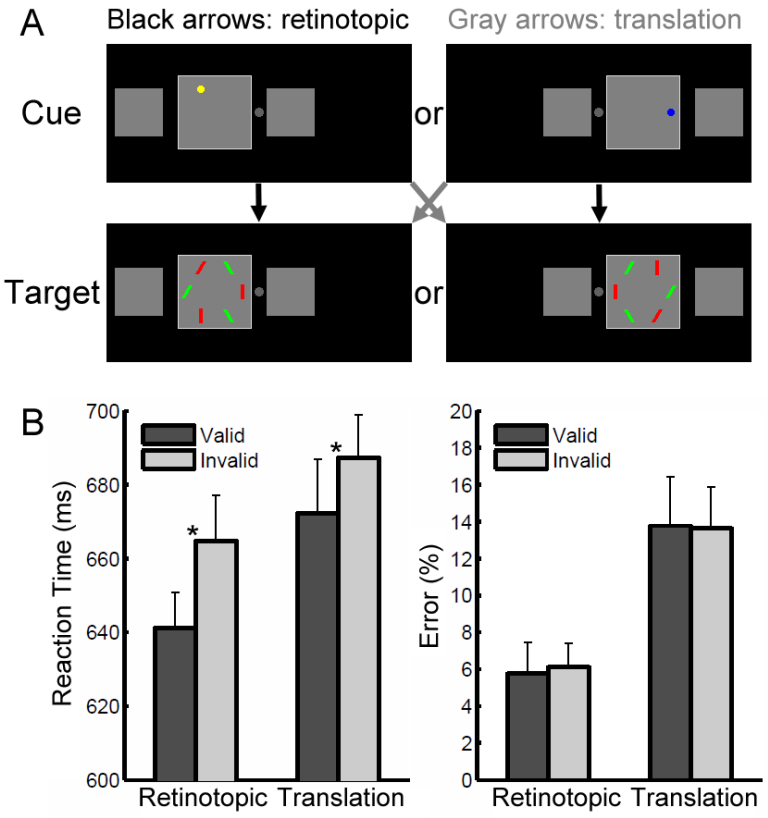


Figure 4-1. Flexible exogenous attention across translation. (A) Stimuli: Observers fixated on a central dot throughout the experiment. The cue frame was presented first, followed by the target frame; they appeared either on the same side of the fixation

(retinotopic display condition; both left or right) or on different sides (translation display condition; one left and one right, forming translation). The cue was uninformative, neither predicting the location of the target frame nor predicting the target location within the search display. Observers were asked to indicate whether the non-vertical red bar was tilted clockwise or counterclockwise. (B) Results (n = 13): In both the retinotopic and translation display conditions, responses were faster for valid (cue and target at the same relative frame position) trials than invalid (cue and target at different relative frame positions) trials. Error bars: standard error of the mean. *: statistically significant difference.

Implicit frames and unique target apparent motion in flexible exogenous attention across translation. Because only one item (the cue) was presented during the cue frame, the translation effect in Experiment 1 might be due to the unique apparent motion between the cue and the target in the valid trials, limiting the scope of flexible exogenous attention. To address this issue, in Experiment 2 we developed a cue-masking procedure for the cue frame (Figure 4-2A), where a transient cue was presented ahead of other items and disappeared at the same time with other items (Mulckhuyse, Talsma, & Theeuwes, 2007). As such, when the test items were subsequently presented on the other side of the fixation as the target frame, all masking items on the cue frame were engaged in apparent motion, allowing us to rule out the confound of unique apparent motion for the valid condition. Since both the cue frame and the target frame were implicit frames made up of the masking and search items, this design also has the added benefit of allowing us to test whether explicit frames were necessary for flexible exogenous attention. A final goal was simply to test the robustness of flexible exogenous attention by using different stimuli (black bars) and a more difficult task with emphasis on accuracy only (discriminating the tilted bar as tilted clockwise or counterclockwise).

A repeated-measures ANOVA revealed that, regardless of retinotopic or translation display conditions (insignificant interaction of display conditions and cue conditions, $F(1, 9) = 0.28$, $P = 0.61$), observers performed significantly more accurately when the cue and the target coincided on the same or same relative location (valid) than on different or different relative locations (invalid; $F(1, 9) = 16.85$, $P = 0.003$, Figure 4-2B). Again, consistent with classic retinotopic exogenous cueing, for the retinotopic

display condition, performance in the valid trials was better than the invalid trails ($t(9) = 3.10$, $P = 0.01$). Importantly, as in Experiment 1, a novel exogenous attention mechanism across translation was observed in the translation display condition, with superior performance in the valid trials than the invalid trails ($t(9) = 3.34$, $P = 0.009$). These findings thus extend flexible exogenous attention across translation into implicit frames, and reveal the robustness of exogenous attention across translation, without being confounded by unique apparent motion between the cue and the target.

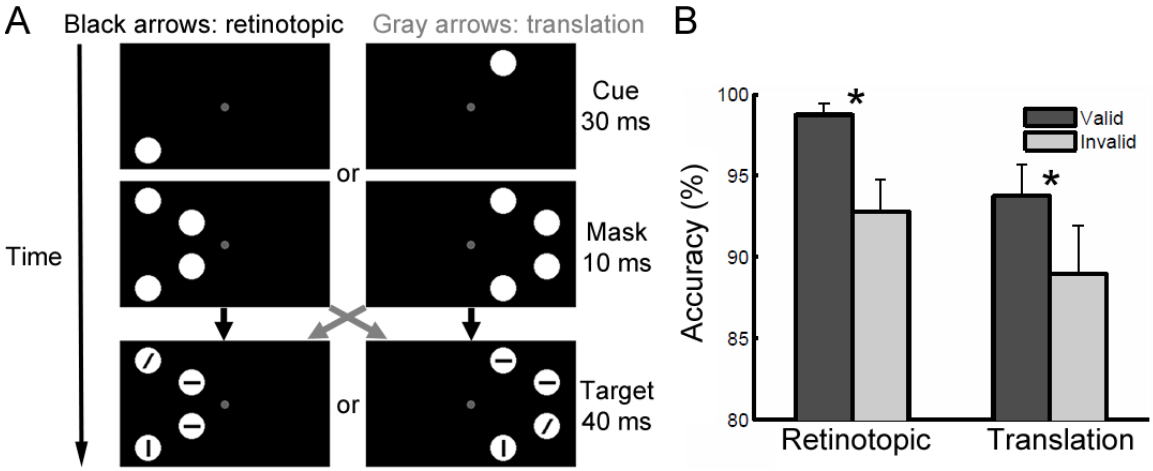


Figure 4-2. Flexible exogenous attention across translation through an implicit frame. (A) Stimuli: The cue frame (first an uninformative disk cue alone, then immediately masked by four disks on the same side of the fixation) was presented first, followed by the target frame either on the same side as the cue frame or on the opposite side. Observers were asked to indicate whether the tilted bar was tilted clockwise or counterclockwise. (B) Results ($n = 10$): In both the retinotopic and translation display conditions, responses were more accurate for valid trials than invalid trials. Error bars: standard error of the mean. *: statistically significant difference.

Flexible exogenous attention across mirror reflection: toward a frame-centered mechanism. Flexibility across translation in exogenous attention as observed in Experiment 1 and 2 represents a non-retinotopic mechanism. Does it reflect a frame-based *spatial* mechanism, such as an attentional spotlight mounted on a moving frame, or a frame-based *structural* mechanism where structural regularities automatically guide attention (i.e., a frame-centered mechanism)? To address this, here we examined an intriguing case of flexibility: generalization of exogenous attention across lateral, left-

right mirror images, as shown in Figure 4-3A. Confusion of left-right mirror images is ubiqutous across many species (Corballis & Beale, 1976): in young children, for instance, letters and their mirror-reflection counterparts sometimes get confused (e.g. letter R and its mirror reflection, as in the Toys "R" Us logo) (Rudel & Teuber, 1963); in human adults, perceiving an object automatically primes its mirror-reflected counterpart (Lavie, Lin, Zokaei, & Thoma, 2009; Stankiewicz, Hummel, & Cooper, 1998). Thus, a similar mechanism might also exist for exogenous attention to generalize across mirror images, allowing the organism to quickly orient to dynamic salient visual information. If target performance was facilitated when targets appeared on the mirror-reflected locations of the cues, compared with other locations, this would constitute evidence for flexible exogenous attention across mirror reflection, and point to a structural, frame-centered mechanism rather a space-based mechanism in play.

In Experiment 3A, exogenous attention across mirror images was examined by using a cue frame and a target frame in lateral, left-right mirror reflection (Figure 4-3A). Similar stimuli and design were used as Experiment 1. In the mirror reflection condition, when the cue and the target were on the same relative frame location (e.g. upper left of the left frame vs. upper right of the right frame), forming mirror reflection, the trial was considered a valid cue trial; otherwise, it was an invalid cue trial. Critically, in the mirror reflection display condition, RTs were significantly faster for valid trials than invalid trial ($t(7) = 4.73$, $P = 0.002$, Figure 4-3B), revealing a novel flexible exogenous attention mechanism across mirror reflection. A significant effect of cue validity was also observed in the retinotopic display condition ($t(7) = 3.16$, $P = 0.016$), but the effect was much smaller than the mirror reflection display condition (significant interaction of display conditions and cue conditions, $F(1, 7) = 8.93$, $P = 0.02$). There are two possible reasons for the smaller cueing effect in the retinotopic condition: 1) stronger masking effects from the cues in the retinotopic display condition than the mirror reflection display condition, and/or 2) the tendency of mirror reflection, which would, by definition, impair retinotopic cueing. No significant differences were found in the accuracy data. Thus, this flexible exogenous attention mechanism across mirror reflection points to a structural, frame-centered mechanism rather a space-based mechanism in play.

Although all observers subjectively reported perceiving mirror reflection when the cue frame and the target frame were on different sides of the fixation and none reported perceiving translational motion, it is unclear whether the translational generalization effect observed in Experiment 1 and 2 played a role in the present mirror reflection generalization effect. To test this, we directly compared translational effect and mirror reflection effect in the mirror reflection condition by classifying trials into three categories: mirror trials (when the cue and the target forming mirror reflection, e.g. upper left of the left frame vs. upper right of the right frame), translation trials (when the cue and the target forming translation, e.g. upper left of the left frame vs. upper left of the right frame), and invalid trials (all remaining trials). We found that RTs were significantly faster in the mirror trials (648 ms) than the translation trials (668 ms), $t(7) = -2.75$, $P = 0.03$; RTs in the translation trials (668 ms), though slightly faster, were statistically comparable to the invalid trials (676 ms), $t(7) = -0.92$, $P = 0.39$. No significant differences were found in the accuracy data. These results therefore demonstrate that translational generalization played an insignificant role in our mirror reflection configuration.

To replicate the effect of flexible exogenous attention across mirror reflection observed in Experiment 3A and to reduce the masking effect from the cues, in Experiment 3B a different configuration and a different task were used (Figure 4-3C). Again, exogenous attention across mirror images was examined by using a cue frame and a target frame in lateral, left-right mirror reflection. Now, a horizontal bar was used as a cue on the cue frame; the target frame consisted of three identical horizontal bars, immediately followed by three t-shape stimuli, one vertical and two tilted. Observers were asked to find the vertical T stimulus and discriminate whether the vertical bar was above or below its abutting horizontal bar. A similar pattern was observed as in Experiment 3A (Figure 4-3D). Again, in the mirror reflection display condition, RTs were significantly faster for valid trials than invalid trial ($t(7) = 4.02$, $P = 0.005$), revealing a flexible exogenous attention mechanism across mirror reflection. A significant effect of cue validity was also observed in the retinotopic display condition ($t(7) = 3.58$, $P = 0.009$), although the effect was smaller than the mirror reflection display condition (significant interaction of display conditions and cue conditions, $F(1, 7) = 6.31$,

P = 0.04), suggesting that the exogenous attention system is highly sensitive to mirror reflection. No significant differences were found in the accuracy data.

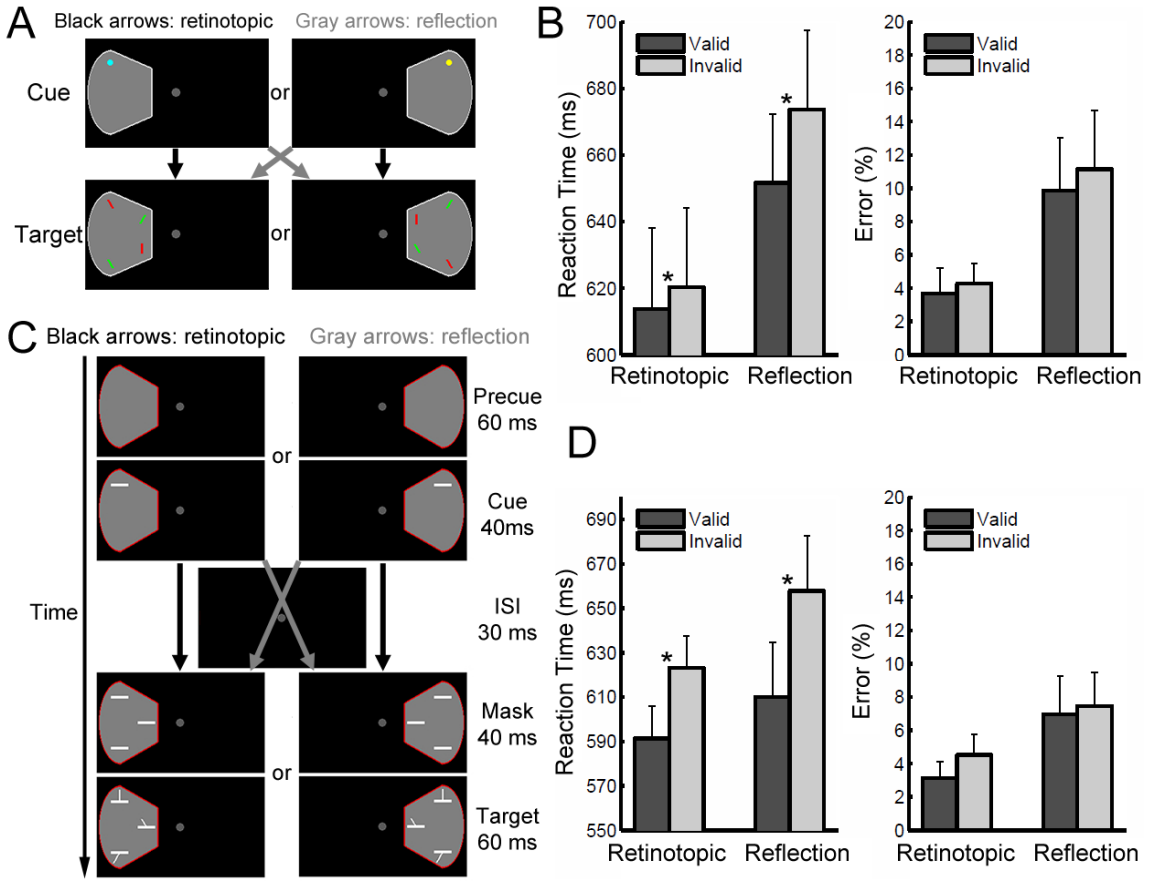


Figure 4-3. Flexible exogenous attention across mirror reflection. (A) Stimuli in Experiment 3A: The cue was uninformative, and the cue frame and the target frame either appeared on the same side of the fixation (retinotopic display condition; both left or right, indicated by the black arrows) or on different sides (mirror reflection display condition; one left and one right, forming mirror reflection, indicated by the gray arrows). Observers were asked to indicate whether the non-vertical red bar was tilted clockwise or counterclockwise. (B) Results from Experiment 3A (n = 8): In both the retinotopic and mirror reflection display conditions, responses were faster for valid trials than invalid trials. (C) Stimuli in Experiment 3B: The cue frame started with a precue, followed by an uninformative cue (precue and cue were on the same side of the fixation). After a brief interval, the mask and then the target frame appeared (mask and target were on the same side), either on the same side as the cue frame or on the opposite side. Observers were asked to indicate whether the vertical bar (but not the tilted bars) was above or below its

abutting horizontal bar. (B) Results from Experiment 3B (n = 8): In both the retinotopic and mirror reflection display conditions, responses were faster for valid trials than invalid trials. Error bars: standard error of the mean. *: statistically significant difference.

4.4 Discussion

Decades of research inspired by the Posner cueing paradigm (Posner, 1980) has suggested that exogenous visual attention operates in a rigid manner, allowing rapid allocation of attention resource to locations of salient events (Carrasco, 2011; Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Theeuwes, 2010). By adopting the logic of the Posner cueing paradigm but using a sensitive moving frame technique, we demonstrate unambiguously that exogenous cueing generalizes across translation and mirror reflection in a frame-centered, relative-location specific manner, revealing a structural, frame-centered mechanism in exogenous attention.

This flexible mechanism in exogenous attention represents an important addition to the classic rigid mechanism in exogenous attention and suggests that exogenous attention is highly adaptive in a dynamic visual environment. For instance, exogenous attention can show rapid visuospatioal learning: performance is improved when targets are constantly presented on the same relative location (Kristjansson et al., 2001), color, or shape (Kristjansson & Nakayama, 2003) within the cue frame (e.g. two parallel lines). Moreover, cueing attention to a moving object results in subsequent inhibition at the locus the object later occupied (Gibson & Egeth, 1994; Jordan & Tipper, 1999; Tipper, Driver, & Weaver, 1991); cueing attention before a saccade results in subsequent inhibition at the same spatiotopic (world-centered) position as the initial cue (Pertzov, Zohary, & Avidan, 2010). Together with these findings, our results thus suggest that exogenous attention is highly plastic and adaptive, able to accommodate different visual stimulations and tasks. As such, flexible exogenous attention prepares the organism to rapidly and efficiently respond to unexpected changes or sudden movements that could signal danger or opportunity.

That exogenous cueing generalizes across translation and mirror reflection suggests that cue representation during the cue frame is not lost when the cue and the cue frame disappear. In addition, the cueing effect is apparent when the cue frame is

perceived to move to the target frame because of apparent motion, revealing the important role of perceptual continuity in establishing flexible exogenous attention. Crucially, unlike object-based processing, the cueing effect was contingent on the relative frame location in a frame-centered, relative-location specific mechanism, consistent with studies showing attention cueing to a remote, frame-centered location when the cue is predictive (Boi, Vergeer, et al., 2011; Umilta et al., 1995). These characteristics are explained by a modified object file account (Kahneman et al., 1992; Treisman, 1992), which we call object cabinet theory. In this object cabinet theory, perceiving the cue frame induces a temporary, episodic representation in an object cabinet. Because of apparent motion, the cue frame is perceived to be moving to the target frame as if they were the same frame, and hence processing items on the target frame automatically retrieves the object cabinet from the cue frame. Critically, the cue representation during the cue frame is frame-centered, location specific, resulting in frame-centered, relative-location specific generalization of exogenous attention across translation and mirror reflection. Such a modified object file account is so called object cabinet theory because the cues (the files) within the reference frame (the cabinet) are orderly coded in their relations to the frame. Object cabinet theory thus readily explains the flexibility of exogenous attention induced by moving frames. In other words, frame-centered representations of the cues support flexible exogenous attention across translation and mirror reflection.

In particular, flexible exogenous attention across mirror reflection may provide an adaptive functional account of left-right mirror images confusion and may be a functional consequence of such perceptual generalization of lateral mirror images. Developmental and comparative psychologists have long been puzzled by the extreme difficulty children and other animals such as octopuses, rats, pigeons, cats, and monkeys have in differentiating left-right mirror images (e.g. / and \) but not mirror images about oblique axes (e.g. — and |) (Corballis & Beale, 1976; Rudel & Teuber, 1963; Sutherland, 1957). It has been debated whether confusion of lateral mirror images represents a limitation of the visual system, or it actually represents a mode of adaptive visual processing (Carrasco, 2011). For instance, although confusion of letters b and d or symbols < and > can be costly through the course of human civilization (e.g., reading and mathematics), in

the natural world, mirror images are usually perceptually equivalent to each other (e.g. symmetry in the real world abounds where mirror images are simply two halves of the same object—the two sides of a face and a vase, or in the modern world a car and a TV). Moreover, given the abundant bilateral symmetries in the natural world and the fact that even the bodies and nervous systems have evolved to be more and more bilateral symmetrical (Weyl, 1952), mirror-image confusion or generalization may be an adaptive strategy of the visual system (Gross & bornstein, 1978). Indeed, although through learning, literate adults easily differentiate b vs. d, TOM vs. MOT, and > vs. <, which have acquired the status of distinct objects, objects in the natural world rarely change their identities through mirror reversal—looking at the mirror, your toothpaste and toothbrush in the bathroom appear just as familiar as they are. Therefore, in the majority of cases, a perceptual system that treats mirror images as equivalent would be adaptive, facilitating rapid object recognition. Our results of flexible exogenous attention across mirror reflection suggest that mirror images are treated as equivalent not just in the perceptual system but also in the attention system, such that cueing within one image generalizes to its mirror image in a relative location specific manner. The perceptual and attentional equivalence of mirror images might thus provide a flexible mechanism for adaptive processing of the ubiquitous mirror images in the natural world.

In conclusion, without being confounded by informative cues, our studies reveal that exogenous visual attention is flexible, generalizing across translation and mirror reflection through a structural, frame-centered mechanism. By exploiting spatial and temporal statistical and structural regularities of the dynamic visual environment, such a flexible exogenous attention mechanism is highly adaptive, facilitating rapid and efficient perception and action.

Chapter 5

Association-driven attentional capture

Attention and learning are two core processes in human cognition. Attention is critical for learning and memory; learning and memory constrain and shape the deployment of attention. How associative learning affects attention deployment is elusive. Learned associations are widely assumed to guide attention in a top-down manner, affording implicit or explicit predictions about item locations. Here we show that a physically inconspicuous stimulus, after being associated with targets through short-term associative learning, captures visual attention involuntarily, even when such capture is detrimental to task performance. In other words, learned associations are not just passive knowledge for exploitation—they capture our attention. Complementing previous stimulus-driven and goal-driven mechanisms, this associative learning based attentional capture represents a novel mechanism in attentional control, and explains previously puzzling findings on reflective attention orienting by eye gaze, arrows, words, numbers, and reward. Evidently, our visual experiences can determine what we see in the first place.

5.1 Introduction

Attention and learning are two fundamental aspects of human cognition: attention determines what aspects of the sensory input to gain access to awareness and memory; learning shapes the way we perceive and react to the environment. Although they are often studied in isolation, increasing consensus suggests that understanding each may require understanding the other (Chun & Johnson, 2011; Desimone, 1996). Attention is crucial for associative learning to take place (Jiang & Chun, 2001; Mitchell & Le Pelley, 2010), but the effect of associative learning on attention deployment remains poorly understood. Closing this gap of knowledge is critical if we are to understand how attention deployment and associative learning interact. Because objects and events in our daily life usually covary in predictable relations (Biederman, 1972), these learned

associations through past experiences can guide our expectation and hence the deployment of top-down attention: a target is thus searched faster when it appears within an old, predictive context than within a new context (Chun & Jiang, 1998; Chun & Jiang, 1999), possibly involving the interaction between brain areas for retrieval of memories for spatial context and the frontoparietal network for top-down attention deployment (Stokes et al., 2012; Summerfield et al., 2006). This effect of associative learning on top-down attention is revealing, but a critical link is missing regarding how associative learning controls bottom-up attention, an essential attention mechanism in which attention is involuntarily captured to the location of physically salient information (Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Nakayama & Mackeben, 1989; Theeuwes, 2010). In other words, can a stimulus, after associative learning, capture one's attention even when it is not physically salient or task relevant?

This notion of attentional capture through associative learning clashes with both stimulus-driven and goal-driven accounts of attentional control, wherein attention is deployed solely by physically salient and task-relevant stimuli. Indeed, previous research on attentional capture seems pessimistic about this possibility. Bottom-up attention is usually thought to be stimulus-driven (Corbetta & Shulman, 2002; Egeth & Yantis, 1997; Nakayama & Mackeben, 1989; Theeuwes, 2010), in which attention is captured by stimulus saliency, such that a red dot among green ones or an abrupt onset in an otherwise static display captures visual attention automatically (Itti & Koch, 2001). However, although such stimulus-driven effect seems to provide compelling evidence for a saliency-based account of attentional capture, to the extent that a salient object or event is usually associated with an important opportunity or an impending danger, the stimulus-driven effect might well reflect long-term association experience through development and evolution. In other words, we suggest that our attention is captured to, say, moving and looming stimuli (Abrams & Christ, 2003; Franconeri & Simons, 2003; J. Y. Lin, Franconeri, & Enns, 2008) not because of moving and looming *per se* but because of the association that moving and looming stimuli usually have with danger and threat (J. Y. Lin, Murray, & Boynton, 2009). This reinterpretation of attentional capture thus raises a counterintuitive prediction: a physically inconspicuous stimulus, after being associated

with targets through short-term associative learning, may involuntarily capture one's attention during extinction, even when such capture is detrimental to task performance.

To test this hypothesis, here we used a protocol of short-term associative learning (a few hundred trials) followed by test (extinction): during learning, an otherwise arbitrary color was consistently paired with the target (hereafter termed "associated color"); during the test, such pairing was removed. This design of associative learning plus test taped into relatively short-term but stable associative knowledge (Anderson et al., 2011), making the results uncontaminated by the influence of working memory contents (Downing, 2000) or semantic memory (Moores, Laiti, & Chelazzi, 2003). The results show that a color previously associated with targets involuntarily captures one's attention during extinction, even when such capture is detrimental to task performance, revealing an involuntary mechanism of attentional selection by associative learning.

5.2 Methods

Subjects and apparatus. Eight subjects (6 females; mean age 22.0) with normal or corrected-to-normal vision from the University of Minnesota community participated in Experiment 1, 16 (12 females; mean age 20.3) in Experiment 2, 12 (10 females; mean age 21.5) in Experiment 3, and 5 (2 females; mean age 19.0) in Experiment 4 (control experiment) in return for money or course credit. The experiments were conducted in accordance with the IRB approved by the Committee on Human Research of University of Minnesota and the Declaration of Helsinki.

The stimuli were presented on a black-framed, gamma-corrected 22-inch CRT monitor (model: Hewlett-Packard p1230; refresh rate: 100 Hz; resolution: 1024 × 768 pixels) using MATLAB Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Subjects sat approximately 57 cm from the monitor with their heads positioned in a chin rest in an almost dark room while an experimenter (ZL) was present.

Stimuli and procedure. To train subjects to maintain stable fixation, before the main experiment, subjects took part in a fixation training session, in which they viewed a square patch of black and white noise that flickered in counterphase (i.e., each pixel alternated between black and white across frames). Each eye movement during the

viewing would lead to perception of a flash, and subjects were asked to maintain fixation using the flash perception as feedback (i.e. to minimize the perception of flashes).

The general procedure in all of the experiments comprised an associative learning phase and a test phase. During the learning phase, a particular color—either red or green, counterbalanced across subjects—was constantly associated with the target (the color was termed the "associated color"; the association relationship was not explicitly told to the subjects in Experiment 1 and 2); during the test phase, no such association existed, and the subjects were explicitly informed about this and told to ignore the colors. Experiment 1 included 30 practice trials (in 1 block), 600 learning trials (in 10 blocks), and 256 test trials (in 4 blocks); Experiment 2 included 30 practice trials (in 1 block), 600 learning trials (in 10 blocks), 24 practice trials, and 288 test trials (in 4 blocks); Experiment 3 included 30 practice trials (in 1 block), 360 learning trials (in 6 blocks), 12 practice trials, and 288 test trials (in 4 blocks); Experiment 4 included 30 practice trials (in 1 block), 360 learning trials (in 6 blocks), 12 practice trials, and 144 test trials (in 2 blocks).

In Experiment 1, each trial in the learning phase included the following sequence: after 400, 500, or 600 ms (randomly selected for each trial) presentation of a white fixation dot (diameter: 0.31°; luminance: 80.2 cd/m^2) on a black background, four colored circles (diameter: 1.6°; width: 0.08°; colors: red, green, blue, and yellow) equidistant from the fixation cross (circle center to fixation distance: 1.5°) were presented on four corners of an imaginary diamond for 200 ms, followed by the same four circles plus four white letters (size: 0.8° × 0.8°; luminance: 80.2 cd/m^2) within the circles for 800 ms or until response. Critically, though not explicitly told to the subjects, the target letter was always presented within the red circle or the green circle, counterbalanced across subjects. The target letter was a T rotated 90° or 270º; the distractors were three Ls rotated 0°, 90°, 180°, or 270° (three rotation degrees randomly selected for each trial). Subjects were asked to discriminate whether the letter T was rotated leftward or rightward and "respond as quickly as possible while minimizing errors." If subjects did not respond within the maximal duration of the letters (800 ms), the trial was considered incorrect; feedback was provided at the fixation for 1000 ms: a minus sign for each incorrect response, a plus sign for each correct one (in the practice session, feedback for each incorrect response also included a re-display of the original letters without the

circles for 2 seconds). The feedback sign was followed by a blank black screen for 1000 ms. Each trial in the test phase was the same as the learning phase except that the target duration was 1200 ms instead of 800 ms, and, more importantly, the target letter appeared equally often within each color circle. Thus, subjects were explicitly instructed that "colors/circles are irrelevant to the task and should be ignored."

In Experiment 2, the trial structure in the learning phase was the same as in Experiment 1, except that the 200 ms frame where the cue circles were presented alone was now omitted (i.e. the circles and the letters appeared at the same time), and that the color set now included red, green, blue, yellow, and cyan (the color associated with the targets was red or green, counterbalanced across subjects; the other three colors were randomly selected from the remaining four colors for each trial). Each trial in the test phase was the same as the learning phase except that the target duration was 1200 ms instead of 800 ms, and, more importantly, the associated color only appeared on half of the trials, and within these trials the target was never presented within this color. (The 24 practice trials were those without the associated color.)

In Experiment 3, the target display in the learning phase now included six colored circles (diameter: 1.8°; width: 0.12°; colors: red, green, blue, yellow, cyan, pink, and gray) equidistant from the fixation cross (circle center to fixation distance: 3°), presented on six corners of an imaginary hexagon for 1000 ms. Subjects were explicitly told that the target letter T in each trial was always within the red circle (or green circle); the five distractors were randomly colored five colors other than red and green . Other aspects were the same as Experiment 1. Each trial in the test phase was the same as the learning phase except that the target duration was 2000 ms and more importantly the target was always within a unique square shape (side: 1.6°; width: 0.12°) among circles. The associated color only appeared on half of the trials, and the square took on each color equally often. (The 24 practice trials were those without the associated color.) Subjects were explicitly instructed that the square had a random color and so colors were irrelevant, and that they should focus on the square.

In Experiment 4, the learning phase was the same as that of Experiment 3. Each trial in the test phase included a horizontal (2.12° × 1.65°) or vertical (1.65° × 2.12°) oval among five circles without letters. The associated color only appeared on half of the

trials, and for these trials the oval took on each color equally often. Subjects were asked to discriminate whether the oval was horizontal or vertical. Other aspects were the same as Experiment 3.

5.3 Results

Experiment 1: Attentional capture by uninformative cues after associative learning. Although subjects were fully aware that colors during the test were uninformative regarding the location of the target and thus were irrelevant to the task and should be ignored, their performance was enhanced when the target happened to appear within the associated color than other colors (Figure 5-1B). Reaction times (RTs) were significantly faster when the target appeared within the associated color (valid trials) than when a distractor appeared within the associated color (invalid trials; $t(7) = -3.37$, $P = 0.012$, Cohen's $d = -1.19$). Error rates of the valid and invalid conditions were comparable ($t(7) = -0.43$, $P = 0.682$, $d = -0.15$). To confirm that the effect of attentional capture we found was consistent and not due to the beginning portion of the test, we split the four test sessions into first half and last half (Figure 5-1B). A repeated-measures ANOVA on RTs revealed only a significant main effect of validity ($F(1, 7) = 11.38$, $P = 0.012$, $\eta_p^2 = 0.619$) without a significant interaction of validity and test sessions ($F(1, 7) < 0.01$, $P = 0.970$, $\eta_p^2 < 0.001$), supporting a robust, consistent effect of attentional capture during the test. Indeed, RTs were significantly faster in the valid trials than in the invalid trials for both the first half ($t(7) = -2.84$, $P = 0.025$, $d = -1.01$) and the last half ($t(7) = -3.05$, $P = 0.019$, $d = -1.08$). No effects were found on the error rates.

These results from uninformative cues thus demonstrate that, after being associated with targets through short-term associative learning, a physically inconspicuous, task-irrelevant stimulus can involuntarily capture one's attention during extinction, even when it does not share any identifying features with the target (Folk, Remington, & Johnston, 1992). Moreover, because the color was associated the target without monetary reward during learning, our finding reveals that merely being associated with the targets is sufficient for attentional capture, indicating that monetary reward is not necessary and might not be special for attentional capture (Anderson et al.,

2011). Therefore, this finding contradicts both stimulus-driven and goal-driven accounts of attentional control and provides support for the associative learning account of attentional capture.

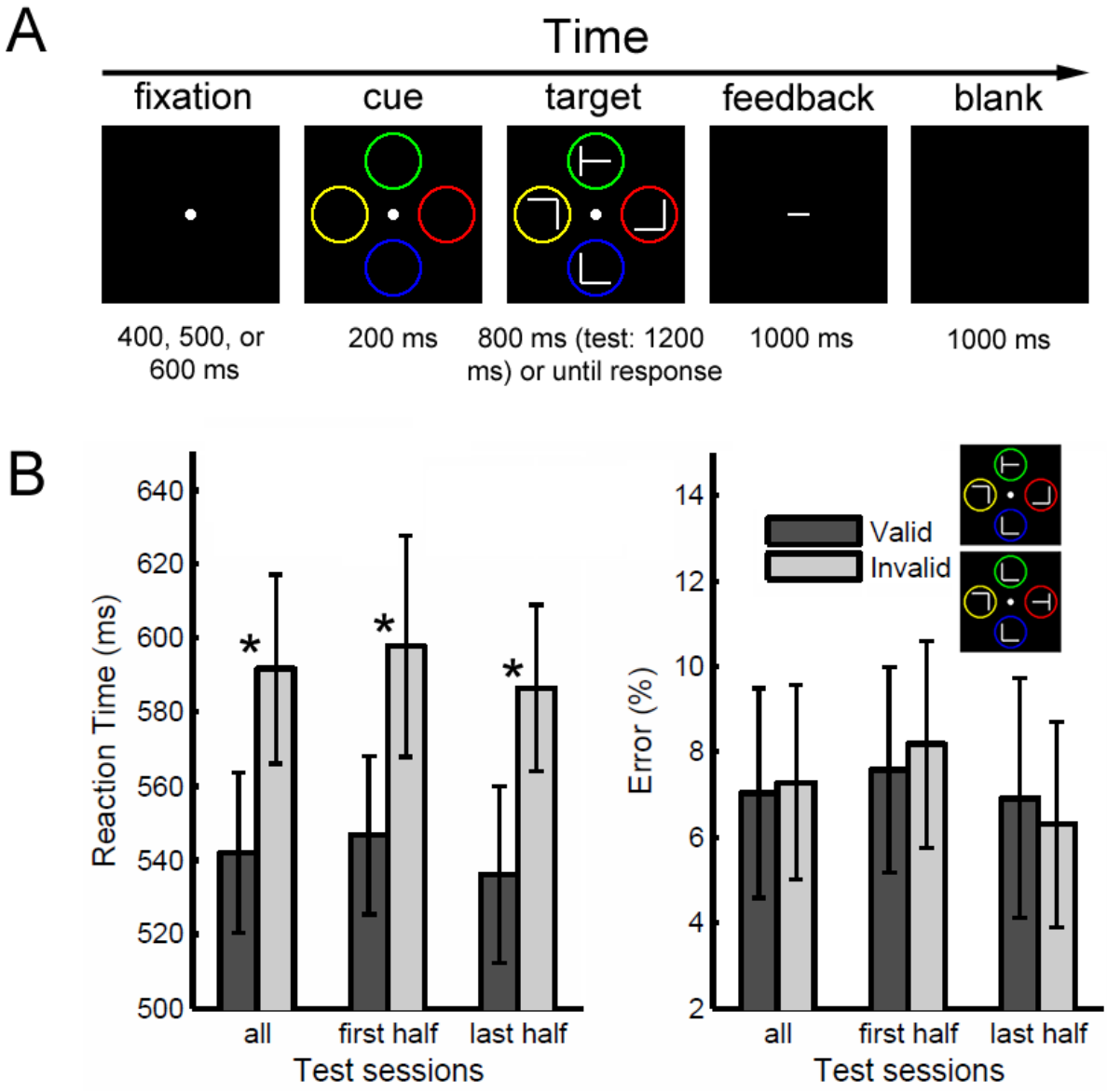


Figure 5-1. Attentional capture by uninformative cues after associative learning. (A) Stimuli in the associative learning phase: Subjects fixated on a central dot throughout the experiment and were asked to find the rotated letter T and reported whether it was rotated to the left or to the right. The target was constantly associated with the green circle (or red circle, counterbalanced across subjects). (B) Results from the test phase (n = 8): In the test phase, the target letter appeared equally often within each color circle (i.e., color

cues were uninformative). In valid trials, the target happened to appear within the associated color; in invalid trial, a distractor appeared within the associated color. Responses were faster for valid trials than invalid trials. Error bars: standard error of the mean. *: statistically significant difference.

Experiment 2: Attentional capture by invalid cues. Using uninformative cues during the test, in Experiment 1 we showed that responses were faster for valid trials than invalid trials. Although this cueing paradigm is widely used in the attentional capture literature (Folk et al., 1992), the effect of attentional capture could be due to a familiarity effect of discriminating letters within the associated color in the valid trials but not in the invalid trials. Another possibility is that subjects might use a strategy of searching for the associated color circle during the test, which, though not optimal, imposed only a low cost for our cueing task since after all the target did appear within the associated color for a fair share of trials (by chance, 25%).

To rule out the familiarity account and also to pit the top-down, search-for-color strategy account with our bottom-up, involuntary capture account, in Experiment 2 we introduced a new design in the test phase: the associated color would only appear on half of the trials, and within these trials the target would never appear within this color. Thus, the test trials included the following two types: those without the associated color (absent trials) and those with the associated color (present trials). Critically, the two accounts—the search-for-color strategy account and the involuntary capture account—made the opposite predictions. The strategy account predicted that performance would be better in the present trials because the color search process by itself would take about twice as much time in the absent trials than in the present trials (Treisman & Gelade, 1980), in addition to the informational factor that attention needed to visit four items to search for the target letter for the absent trials but only three items for the present trials (since the target never appeared within the associated color). The involuntary capture account, however, predicted the opposite, that performance would be better in the absent trials simply because attention would be involuntarily captured to the associated color in the present trials and thus delayed the deployment of attention to the target letter. Moreover, since the targets never appeared within the associated color, the performance advantage in the absent trials could not be explained by familiarity. The results support the

involuntary capture account (Figure 5-2B): RTs were significantly faster in the absent trials than in present trials ($t(15) = -2.49$, $P = 0.025$, $d = -0.62$), with comparable error rates in the two conditions ($t(15) = 0.83$, $P = 0.418$, $d = 0.21$).

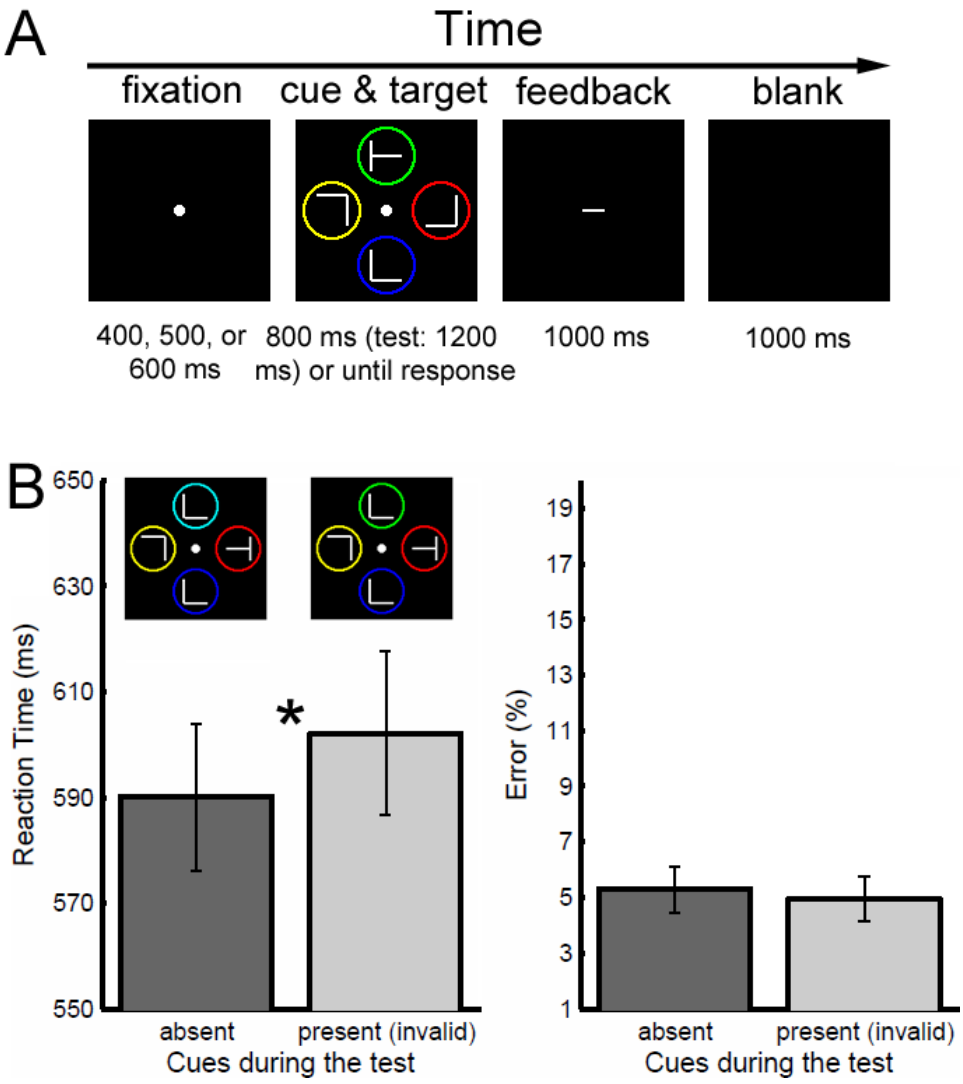


Figure 5-2. Attentional capture by invalid cues. (A) Stimuli in the associative learning phase: The task was to report whether the target T, constantly associated with the green circle (or red circle, counterbalanced across subjects), was rotated to the left or to the right. (B) Results from the test phase (n = 16): The associated color only appeared on half of the trials, within which a distractor (but never the target) appeared within the associated color. Responses were faster for trials without the associated color (absent)

than with the associated color (present-invalid). Error bars: standard error of the mean. *: statistically significant difference.

Experiment 3: Attentional capture by uninformative color cues even when attention is guided by a clear top-down goal (search-for-a-square). To examine the automaticity of the attentional capture by associative learning, in this experiment we directly pitted the bottom-up capture effect with a clear top-down goal of the subjects: subjects in the test were asked to search for a unique square among circles and respond to the target letter T inside the square (a compound search task). The associated color appeared on half of the trials, and for these trials the square took on each color, including the associated color, equally often; thus, colors provided no information for the task. The associative learning account made two predictions: first, for those trials with the associated color, performance would be impaired when the associated color shape coincided with a distractor (invalid) than with a target (valid); second, performance would also be impaired when the associated color shape coincided with a distractor (invalid) than trials without the associated color (absent).

Both predictions were confirmed (Figure 5-3B): for trials with the associated color, RTs were significantly faster in the valid trials than the invalid trials ($t(11) = -3.23$, $P = 0.008$, $d = -0.93$), with comparable error rates in the two conditions ($t(11) = -0.10$, $P = 0.922$, $d = -0.03$). Moreover, RTs were significantly faster in the absent trials than the invalid trials ($t(11) = -6.48$, $P < 0.001$, $d = -1.87$), with comparable error rates in the two conditions ($t(11) = -0.93$, $P = 0.374$, $d = -0.27$). (RTs were faster in the valid trials than the absent trials with marginal significance, $t(11) = -2.03$, $P = 0.067$, $d = -0.59$, with comparable error rates, $t(11) = 0.50$, $P = 0.624$, $d = 0.15$.) These results from nonsalient, task irrelevant stimuli is strikingly similar to the classic attentional capture effect from physically salient stimuli (Theeuwes, 1992), supporting the associative learning account of attentional capture. In a control experiment (Experiment 4), we also showed that the attentional capture effect generalizes to a different task (Figure 5-3C).

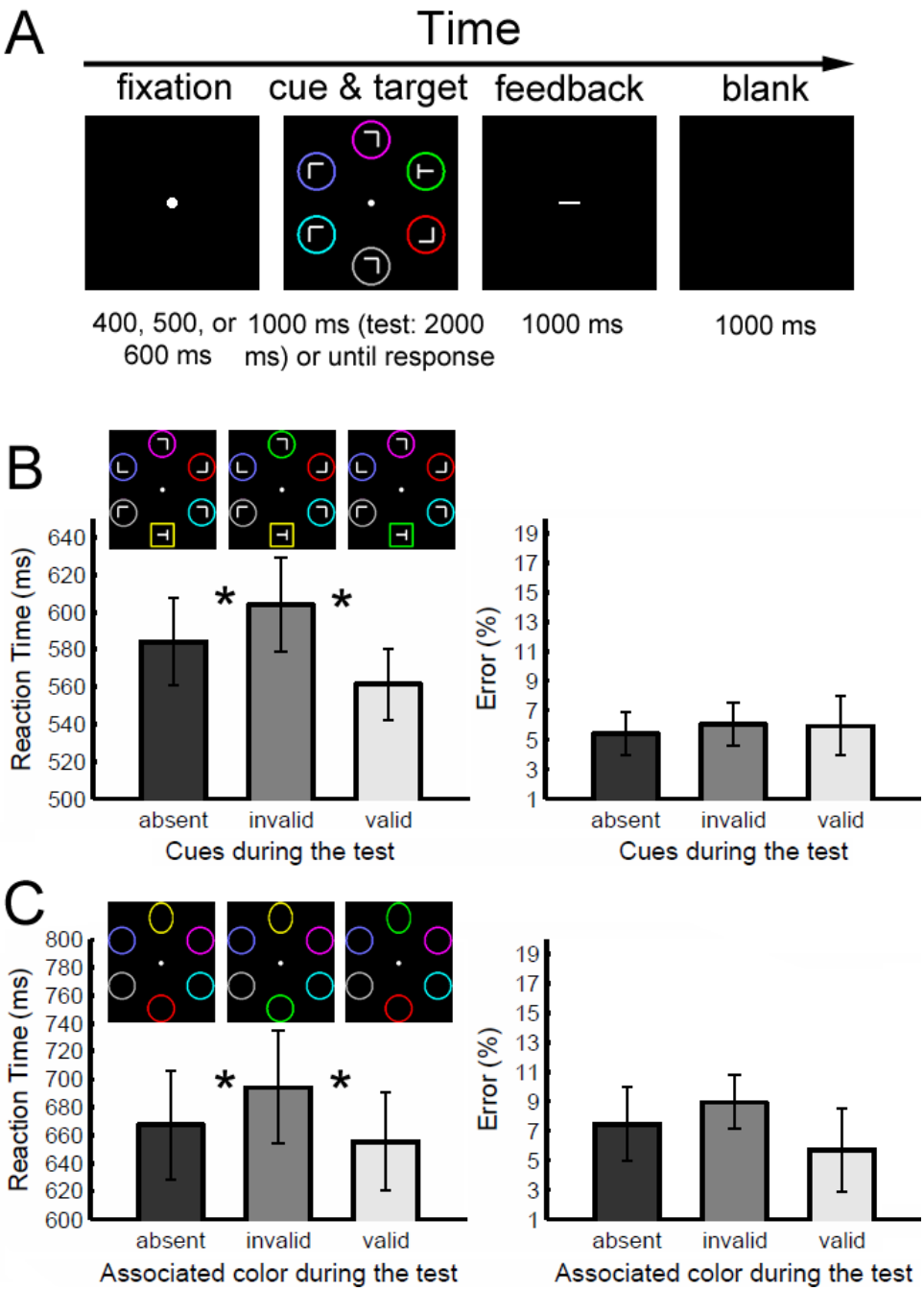


Figure 5-3. Attentional capture by uninformative color cues even when attention is guided by a clear top-down goal (search-for-a-square) and its generalization to a different task. (A) Stimuli in the associative learning phase: The task was to report whether the target T, constantly associated with the green circle (or red circle, counterbalanced across subjects), was rotated to the left or to the right. (B) Results from the test phase (n = 12) in Experiment 3 (attentional capture with a top-down goal): The target always appeared within a square and the distractors within circles. The associated color appeared on half

of the trials, and for these trials the square took on each color (including the associated color) equally often. Responses were faster for trials without the associated color (absent) than when a distractor appeared within the associated color (invalid); when the associated color appeared, responses were also faster if the target appeared within the associated color (valid) than when a distractor appeared within the associated color (invalid). (C) Results from the test phase (n = 5) in Experiment 4 (generalization to a different task): Subjects searched for an oval shape among circles and report whether it was horizontal or vertical. The associated color appeared on half of the trials, and for these trials the oval took on each color (including the associated color) equally often. The same pattern of results was found as in (B). Error bars: standard error of the mean. *: statistically significant difference.

5.4 Discussion

Serving as a gatekeeper of visual inputs, attention is known to be critical for cognitive processes such as learning and memory, but interactions among cognitive processes go both ways, so learning and memory ought to constrain and shape the deployment of attention. Previous research suggests that learned associations provide top-down cues for attention (Chun, 2000), so that we can use the stored knowledge to make informed predictions regarding the locations of items and thus improve our performance. The experiments reported here show that a physically inconspicuous stimulus, after being associated with targets through short-term associative learning, captures one's attention involuntarily, even when such capture is detrimental to task performance. In other words, not just as indispensable knowledge awaiting the visual system to exploit, learned associations demand attention.

An associative learning mechanism of attentional control. Associative learning based attentional capture represents a novel mechanism in attentional control, complementing two well-established attentional control mechanisms, namely, stimulus-driven attention and goal-driven attention, wherein attention is driven solely by physically salient and task-relevant stimuli. Using nonsalient and task-irrelevant stimuli, our results demonstrate that, after being associated with targets through short-term associative learning, these seemly neutral stimuli capture attention. Naturally, the

associative learning account holds that the strength of association determines the strength of attentional capture. This proposal is consistent with the reward magnitude-dependent attentional capture, wherein colors that are associated with high rewards captures attention more strongly than those with low rewards (Anderson et al., 2011), presumably because of higher levels of motivation to attend and react to stimuli with higher rewards. It is thus conceivable that increasing the number of items that are associated with targets would weaken the strength of association for each item, resulting in weaker effects of attentional capture by these items, i.e., a dilution effect by multiple items. This dilution effect may underlie the lack of attentional capture by colors when two colors are associated with targets and without rewards (Anderson et al., 2011).

More generally, this new associative learning mechanism may be a general principle for attentional deployment, parsimoniously explaining several previously puzzling findings, including "reflective" orienting of visual attention to the direction of eye gaze (Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 1999), the meaning of symbolic stimuli such as arrows (Eimer, 1997; Shepherd, Findlay, & Hockey, 1986) and words (Hommel, Pratt, Colzato, & Godijn, 2001), and the magnitude of numbers (Fischer, Castel, Dodd, & Pratt, 2003). Although reflective orienting effects do not necessarily reflect attentional capture, they can be understood as manifestations of attentional effects as a consequence of these cues being associated with spatial location and direction in our daily experiences—the direction of eye gaze is associated with specific direction through our social interactions, the direction of arrows through civilization (e.g. traffic signs), the meanings of words through education, and the location of valuable stimuli through evolution and development. This interpretation is consistent with a recent finding that, just like eye gaze, arrows, words, and numbers, neutral central cues, after being associated with spatial locations, can influence the allocation of spatial attention (Dodd & Wilson, 2009). It is also consistent with the fragile effects of attentional capture by the magnitude of numbers (Galfano, Rusconi, & Umilta, 2006; Ristic, Wright, & Kingstone, 2006) because of the weak number-to-space association. In sum, associative learning is a general mechanism in vision, providing a powerful tool for the brain to make sense of the environment and direct visual attention, voluntarily and involuntarily.

Learning, memory, and automaticity: implications for attention research. Practice makes perfect, and automatic (Shiffrin & Schneider, 1977). Repeated training makes behaviors more and more automatic so that many acts in daily life seem effortless (e.g., typing and driving). This effect of practice and learning on behavior seems obvious, yet it is often overlooked in research on attention, probably because of the belief that it would require intensive training (e.g. days and weeks) to make a mental act automatic. For instance, using apparent motion, a recent study had subjects participate in a 100% valid cueing task and later on used a subset of these subjects in an uninformative cueing task; results from the uninformative cueing task were used to make claims regarding stimuli-driven, exogenous attention (Boi, Vergeer, et al., 2011). As the experiments reported here show, associative learning is extremely powerful and rapid (as few as 360 trials in Experiment 3), so that a cue that was associated with the target in the 100% valid cueing task would very likely capture one's visual attention even when it became nonpredictive in the neutral cueing task. The effect of learning on attention deployment is thus of practical significance for designing experiments and interpreting results. Although different mental acts are likely to have different sensitivities to learning and memory effects, it behooves us to be careful and perhaps conservative in our assumptions.

Contextual attention: visual experiences shape attentional priorities. Our findings of attentional capture by associated learning raise a more general question regarding the plasticity of attention: to what extent does attention depend on contexts? In visual search, attending to a visual feature (e.g. color or spatial frequency) leaves a short-term implicit memory trace of the attended feature, automatically priming the deployment of attention to the same feature in the following trials (e.g. up to 30 secs), known as priming of pop-out (Maljkovic & Nakayama, 1994). Priming of attention deployment is also evident for spatial locations (Maljkovic & Nakayama, 1996). A similar episodic memory effect on attention selection occurs in flanker tasks, showing that repetition priming across trials can account for the conflict-adaptation effect (Mayr, Awh, & Laurey, 2003): performance on incongruent trials are faster after incongruent than after congruent trials; performance on congruent trials are faster after congruent than after incongruent trials (Gratton, Coles, & Donchin, 1992). Our findings from associative learning extend these episodic memory effects to relatively long term, stable contexts,

with the effect of association on attentional capture lasting for hundreds of trials, consistent with contextual cueing effects on top-down attention (Chun, 2000). Because it is long lasting, stable contextual effect is potentially more powerful and pervasive, and thus may be partly responsible for our everyday attentional exploration of the world. For example, the words cucumber and emerald seem unrelated, yet after a short-term training in color discrimination, the two become strongly linked so that they prime each other (Yee, Ahmed, & Thompson-Schill, 2012); in other words, attention is drawn to the color feature after the contextual color task, making words sharing the same color more similar. Together, these findings define and enrich the concept of contextual attention; visual experiences, both short term and longer term, can shape attentional priorities and thus how we see the world.

Chapter 6

Visual familiarity induces inhibition through attentional inertia

Repeated encounters with objects imbue them with familiarity, facilitating subsequent attentional selection and visual processing. Many studies show that familiar task irrelevant objects demand or even capture visual attention, causing interference in processing task relevant objects. To isolate visual familiarity from automatic associations with specific meanings and responses, we develop an associative learning procedure wherein color objects become familiar by constantly being associated with the targets but, critically, not with a particular behavioral response. In a subsequent response competition task, we make a counter-intuitive observation: incompatible distractors with familiar colors cause less interference effect on target processing than incompatible distractors with unfamiliar colors, demonstrating an inhibition effect of familiar colors. This inhibition effect generalizes to new shape contexts and is not due to habituation of familiar colors. We propose that attentional modes or sets from familiarization carryover to subsequent tests ("attentional inertia"), determining the fate of task irrelevant familiar distractors.

6.1 Introduction

Our visual experiences shape the way we attend, perceive, and interact with the world. After being frequently exposed to the things in the environment, we become familiar with them.

A hallmark of familiarity is the facilitation of perceptual processing, so that after relocating to a new city, gradually we are able to recognize faces of colleagues, extract meanings from strongly accented conversations, and navigate around the neighborhood without getting lost. Indeed, many studies in visual attention show that familiarity speeds selection of task relevant information (Wolfe, 2001): finding the digital number 2 among number 5s is severely impaired when the numbers are rotated 90° and become unfamiliar

(Wang, Cavanagh, & Green, 1994); finding a target among familiar non-target letters becomes much more effortful when the letters are mirror-reflected and thus unfamiliar (Frith, 1974). Therefore, although not a characteristic of the visual pattern itself, familiarity plays a fundamental role in visual attention by speeding up processing of the visual pattern and prioritizing attentional resource to it. This is perhaps best captured by the classic Stroop effect, where naming the ink color of a word with an incongruent color meaning is much slower and more error-prone than naming the semantic color of the word (Stroop, 1935). Despite strong incentives not to attend to them, task irrelevant word meanings intrude ink color perception because familiarity of words through years of extensive training renders the extraction of word meanings so automatic that attention seems to be captured to the meanings (MacLeod & MacDonald, 2000).

Attentional prioritization to, or even capture by, familiar task irrelevant information seems maladaptive. How come then we look at the piles of books and papers in the clutter office as though air, with our limited attentional resource seemingly intact? This distinction suggests an intrinsic but surprisingly overlooked link of attentional prioritization to familiar distractors and the mode of processing during familiarization. We become familiar with words by continuously associating words with their meanings and further reinforcing these associations by using them. During the Stroop task, this mode of meaning extraction carries over to the ink color naming even though it is clearly detrimental to the task at hand, causing interference. On the other hand, we become familiar with the items cluttering our office by mere exposure, without associating specialized responses to them. During our routines in the office, this mode of not associating response to clutters carries over so that we can focus on the task at hand with minimal interference from the clutters. We call these carryover effects of attentional modes or sets attentional inertia. The attentional inertia hypothesis thus unifies the aforementioned seemingly conflicting observations, and furthermore makes a counter-intuitive prediction: familiarity can induce inhibition rather than facilitation in visual attention if objects are constantly *not* associated with specific responses during familiarization because the carryover effect of *not* associating the objects with specific responses inhibits subsequent responses to them.

Testing this prediction meets two methodological challenges: first, perceptual familiarity is usually confounded by association links with specific meanings and responses (Lupyan & Spivey, 2008); second, attentional inhibition, though intuitive, is hard to demonstrate directly (Wuhr & Frings, 2008). To overcome the first difficulty, in the familiarization phase, we used an associative learning protocol in which a particular color (red or green, counter-balanced across subjects) was constantly associated with the targets but, critically, not with a particular behavioral response (left rotated letter T or right rotated letter T). As such, unlike previous research where novel shapes become familiar by being associated with specific labels and responses (MacLeod & Dunbar, 1988), in our design the simple features—colors—became familiar yet without being confounded by meanings and responses during associative learning. To overcome the second difficulty, we investigated attentional inhibition of irrelevant objects within single trials rather than relied on sequential effects. Sequential effects as used in previous research, including inhibition of return (Posner & Cohen, 1984) and negative priming (Tipper, 1985), measure the aftereffects of inhibition, wherein evidence of performance impairments in processing certain information is used to infer that the representation or processing of that information must be inhibited on the last trial or during the last time point, and are subject to alternative accounts without evoking inhibition, such as retrieval of episodic memory (Neil & Mathis, 1998; Neill & Valdes, 1992; Neill, Valdes, Terry, & Gorfein, 1992). Our task took advantage of response competition between task relevant information and task irrelevant information, so that inhibition of irrelevant information should lead to a reduction in the interference effect (Wuhr & Frings, 2008). Specifically, the display in the response competition task of Experiment 1 consisted of a smaller circle or square (the target) within a bigger circle or square (the distractor); the task was to discriminate the shape of the target while ignoring the distractor (Experiment 2 used new shapes—diamond and square). Critically, the target color and the distractor color were the familiar color and the unfamiliar color, respectively, or the reverse, randomized across trials. If the familiar color was indeed inhibited, in incongruent trials (where the target and the distractor were of different shapes, thus evoking response competition), one would expect to see faster response time when the target was the unfamiliar color and the distractor was the familiar color than the reverse; however, if the familiar color was

facilitated, in incongruent trials, one would expect to see faster response time when the target was the familiar color and the distractor was the unfamiliar color than the reverse. Our results are consistent with the inhibition account and cannot be explained by alternative accounts based on faster responses to unfamiliar color or slower responses to familiar colors (e.g. habituation of familiar colors). Moreover, the results hold even when the shapes used in the familiarization phase were distinct from the shapes used in the test phase, demonstrating inhibition of familiar colors independent of shape contexts. These results thus yield support for the attentional inertia hypothesis; carryover of attentional modes may be adaptive, sparing attentional resource for processing novel and urgent information and acquiring new skills.

6.2 Methods

Subjects and apparatus. Fourteen subjects (8 females; mean age 21.0) with normal or corrected-to-normal vision from the University of Minnesota community participated in Experiment 1, sixteen in Experiment 2 (10 females; mean age 20.3), and twenty five in Experiment 3 (17 females; mean age 20.2) in return for money or course credit. The experiments were conducted in accordance with the IRB approved by the Committee on Human Research of University of Minnesota and the Declaration of Helsinki.

The stimuli were presented on a black-framed, gamma-corrected 22-inch CRT monitor (model: Hewlett-Packard p1230; refresh rate: 100 Hz; resolution: 1024 × 768 pixels) using MATLAB Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Subjects sat approximately 57 cm from the monitor with their heads positioned in a chin rest in an almost dark room.

Stimuli and procedure. To train subjects to maintain stable fixation, before the main experiment, subjects took part in a fixation training session, in which they viewed a square patch of black and white noise that flickered in counterphase (i.e., each pixel alternated between black and white across frames). Each eye movement during the viewing would lead to perception of a flash, and subjects were asked to maintain fixation using the flash perception as feedback (i.e. to minimize the perception of flashes).

The procedure in both experiments comprised a familiarization phase through associative learning (30 practice trials in 1 block and 480 formal trials in 8 blocks) and a

test phase (12 practice trials and 144 formal trials combined in 2 blocks). For both experiments, each trial in the learning phase included the following sequence: after 400, 500, or 600 ms (randomly selected for each trial) presentation of a white fixation dot (diameter: 0.31°; luminance: 80.2 cd/m^2) on a black background, six homogeneous shapes, either circles or squares (randomly selected for each trial; circle diameter: 1.8°; square size: 1.6°; width: 0.12°; colors: red, green, blue, yellow, cyan, pink, and gray) for Experiment 1 and circles only for Experiment 2, equidistant from the fixation cross (shape center to fixation distance: 3°) with six white letters (size: 0.8° × 0.8°; luminance: 80.2 cd/m2) within were presented on six corners of an imaginary hexagon for 1000 ms or until response. The target letter was a T rotated 90° or 270º; the distractors were five Ls rotated 0°, 90°, 180°, or 270° (all with different rotation degrees except for two items, randomly selected for each trial). Subjects were asked to discriminate whether the letter T was rotated leftward or rightward and "respond as quickly as possible while minimizing errors." Critically, though not explicitly told to the subjects, the target letter was consistently presented within the red shape or the green shape, counterbalanced across subjects. The associated color was thus the familiar color and the other color was the unfamiliar color. If subjects did not respond within the maximal duration of the letters, the trial was considered incorrect; feedback was provided at the fixation: a minus sign for 1200 ms after each incorrect response, a plus sign for 400 ms after each correct one. The feedback sign was followed by a blank black screen for 1000 ms.

The test phase used a Stroop-like response competition task. In Experiment 1, the display consisted of a smaller circle (diameter: 0.78°; width: 0.08°) or square (diameter: 0.66°; width: 0.08°) within a bigger circle (diameter: 1.8°; width: 0.16°) or square (side: 1.8°; width: 0.16°) centered on the fixation for 2000 ms or until response. The task was to discriminate whether the smaller shape (target) was a circle or a square while ignoring the bigger shape (distractor).

Critically, the target color and the distractor color were the familiar (associated) color and the unfamiliar (unassociated) color, respectively, or the reverse, randomized across trials. Experiment 2 was similar, except that the two shapes were square and diamond (square rotated 45°) and thus were distinct from the shape (circle) used during the familiarization phase.

Experiment 3 comprised Experiment 1-control (8 practice trials and 144 formal trials combined in 2 blocks) and Experiment 2-control (8 practice trials and 144 formal trials combined in 2 blocks), which were similar to their respective test phases, except that only one shape, either the smaller one or the bigger one, was presented one at a time with either the familiar (associated) color or the unfamiliar (unassociated) color; shapes and colors were randomized across trials. Seven of the 14 subjects took part in Experiment 1-control; all of the 16 subjects and two additional subjects took part in Experiment 2-control.

6.3 Results

Experiment 1: Inhibition of a familiar color through associative learning. During the initial familiarization phase, subjects searched for a rotated letter T (target) among rotated letter Ls (distractors; Figure 6-1A). Importantly, for half of the subjects, targets were always presented within the green shapes and distractors were presented within shapes of various other colors including red; hence, the associated color—green—was attended and became familiar to the subjects without being associated with a specific label or response, and other colors, red included, were not searched for and received much less attention. To control for possible differences in physical salience, for the other half of the subjects, red was the associated, familiar color and green was an unassociated and, relative to red, unfamiliar control color. Circle and square shapes were randomized across trials.

During the test phase, a response competition task was used to probe attentional inhibition, wherein subjects discriminated an inside shape as a circle or square (target) while ignoring an outside circle or square (distractor). Thus, the target and the distractor could be of the same shape (congruent condition) or different shapes (incongruent condition). Importantly, the target was of the familiar color (familiar target) and the distractor was of the unfamiliar control color (unfamiliar distractor), or the reverse (unfamiliar target circumscribed by familiar distractor), randomized across trials. A repeated-measures ANOVA on reaction times (RTs) revealed a significant interaction between shape compatibility and color familiarity ($F(1, 13) = 7.83$, $P = 0.015$, $\eta_p^2 = 0.376$; Table 6-1). Specifically, RTs were faster for unfamiliar targets surrounded by

familiar distractors than the reverse in the incongruent condition ($t(13) = -3.18$, $P = 0.007$, Cohen's $d = -0.85$) but not in the congruent condition ($t(13) = 0.51$, $P = 0.616$, $d = 0.14$), revealing an inhibition effect of familiar distractors. This inhibition effect in the incongruent condition could not be attributed to faster RTs to unfamiliar targets than familiar targets, or faster RTs to unfamiliar distractors than familiar distractors caused by habituation of the familiar colors, as no difference in RTs was found in the congruent condition even though the same familiarity relationships applied. A repeated-measures ANOVA on error rates revealed the same pattern as RTs, with a significant interaction between shape compatibility and color familiarity ($F(1, 13) = 9.56$, $P = 0.009$, $\eta_p^2 = 0.424$; Table 6-1). Specifically, error rates were significantly lower for unfamiliar targets surrounded by familiar distractors than the reverse in the incongruent condition ($t(13) = -2.99$, $P = 0.010$, $d = -0.80$) but not in the congruent condition ($t(13) = 0.32$, $P = 0.757$, $d = 0.08$).

Because RTs and error rates show the same pattern of results, we combine both measurements into a single index—normalized RTs, mean RTs divided (normalized) by the proportion correct for each condition (Townsend & Ashby, 1983). Figure 6-1B plots the results: normalized RTs were significantly faster for unfamiliar targets surrounded by familiar distractors than the reverse only in the incongruent condition ($t(13) = -5.17$, $P < 0.001$, $d = -1.38$) but not in the congruent condition ($t(13) = 0.60$, $P = 0.557$, $d = 0.16$), resulting in a significant interaction between shape compatibility and color familiarity ($F(1, 13) = 19.06$, $P < 0.001$, $\eta_p^2 = 0.595$). This observation of inhibition of familiar colors is striking considering that the familiar colors were constantly searched for during the familiarization phase. These results thus provide a novel demonstration that familiarity can induce inhibition rather than facilitation when objects are not being associated with specific labels or responses during familiarization, supporting the attentional inertia hypothesis.

Table 6-1. Mean reaction time (in milliseconds) and error rate (in %) in the test phase (the inside target shape and the outside distractor shape were presented simultaneously; the target color and the distractor color were the familiar color and the unfamiliar color, respectively, or the reverse, randomized across trials) of Experiment 1 (same shapes across familiarization and test) and Experiment 2 (different shapes across familiarization and test)

	Target color			
	Experiment 1 (N = 14)		Experiment 2 (N = 16)	
Compatibility	Familiar	Unfamiliar	Familiar	Unfamiliar
Congruent				
Reaction time	452.3 (10.4)	454.4 (11.4)	517.2 (18.5)	520.7 (17.3)
Error rate	3.7 (0.8)	4.0 (0.8)	5.9 (0.9)	6.0 (1.1)
Incongruent				
Reaction time	480.2 (12.0)	469.2 (12.3)	558.8 (17.3)	546.1 (13.5)
Error rate	10.7 (2.1)	6.2 (1.1)	10.9 (1.7)	8.2 (1.5)

Note: For half of the subjects, the familiar (associated) color was red and the unfamiliar (unassociated) color was green; for the other half, it was the reverse.

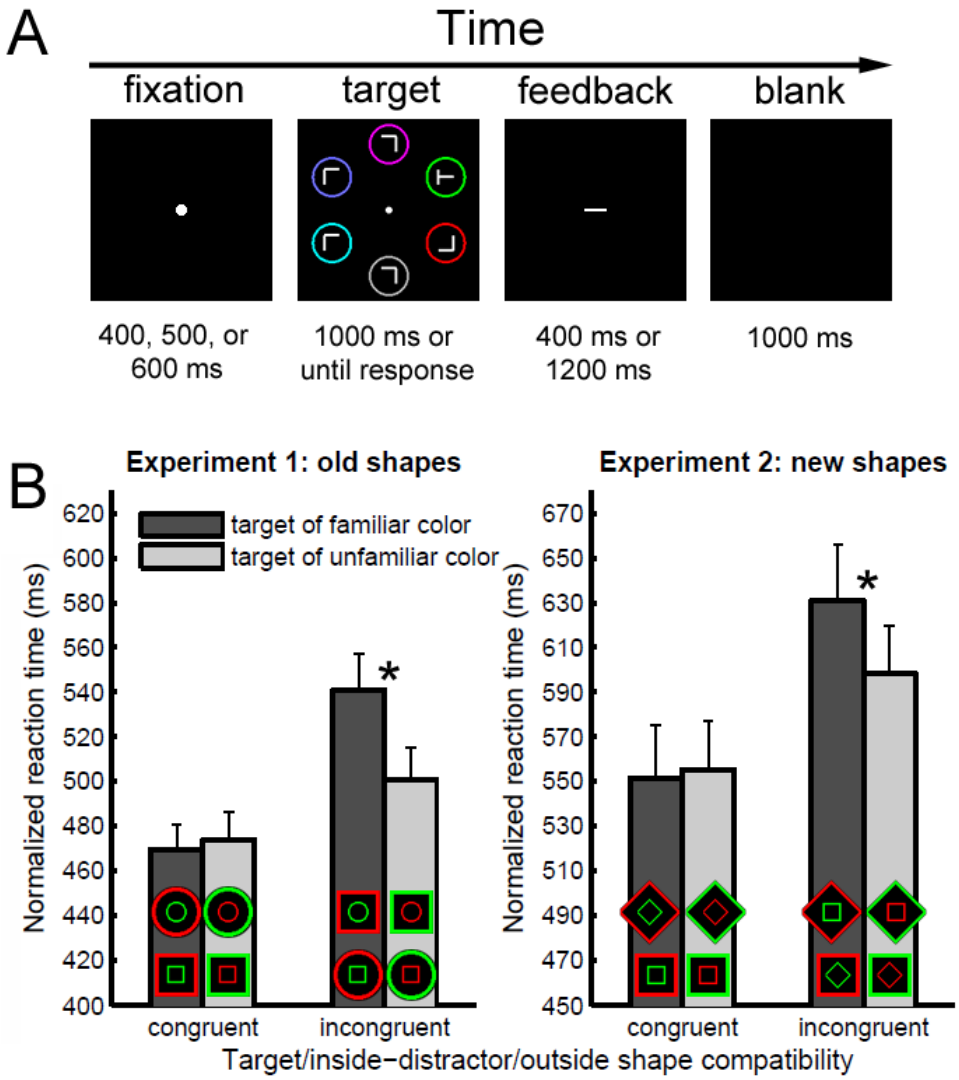


Figure 6-1. Familiarity induces inhibition in visual attention through attentional inertia. (A) Stimuli in the familiarization phase (associative learning): Subjects fixated on a central dot throughout the experiment and were asked to find the rotated letter T among rotated distractor letter Ls and reported whether it was rotated to the left or to the right. In Experiment 1, the letters were within circles or squares, randomized across trials, and the target was consistently within the green circle/square (or red circle/square, counterbalanced across subjects). Experiment 2 used circles only. Colors were thus imbued with familiarity without being associated with specific meanings or responses: for half of the subjects, the familiar (associated) color was red and the unfamiliar (unassociated) color was green; for the other half, it was the reverse. (B) Results from the test phase (Stroop-like response competition; n = 14 in Experiment 1 and n = 16 in

Experiment 2): In Experiment 1 (old shapes), subjects were asked to categorize the smaller shape at the fovea as a circle or square while ignoring the bigger circle or square outside (both shapes were old shapes from the familiarization phase); Experiment 2 (new shapes) used diamond and square, distinct from the shape (circle) used in its familiarization phase. Critically, the target color and the distractor color were the familiar color and the unfamiliar color, respectively, or the reverse, randomized across trials. In incongruent trials (where the target and the distractor were of different shapes), performance was better when the target was of the unfamiliar color and the distractor was of the familiar color than the reverse. Normalized reaction time was the mean reaction time divided (normalized) by the proportion correct for each condition. Error bars: standard error of the mean (SEM). *: statistically significant difference.

Experiment 2: Generalization of inhibition to new shape contexts. The shapes used in the test phase in Experiment 1, circle and square, were those used during the familiarization phase. This raised the question regarding the generalizability of the inhibition effect: to what extent does inhibition of familiar colors depend on the shape context? To address this question, in this experiment tested on new subjects, only one shape—circle—was used during familiarization, and two new shapes—diamond and square—were used during the test. Thus the shapes used in the test, composed of straight lines, were distinct from the circle shape used during familiarization.

As with Experiment 1, RTs and error rates yield the same pattern of results (Table 6-1) and thus, as a single index, normalized RTs were used (Figure 6-1B). Normalized RTs were significantly faster for unfamiliar targets surrounded by familiar distractors than the reverse in the incongruent condition ($t(15) = -3.99$, $P = 0.001$, $d = -1.00$) but not in the congruent condition ($t(15) = 0.35$, $P = 0.731$, $d = 0.09$), resulting in a significant interaction between shape compatibility and color familiarity ($F(1, 15) = 6.18$, $P = 0.025$, $\eta_p^2 = 0.292$). These results thus clearly demonstrate that the inhibition of familiar colors is independent of shape contexts, providing strong support for the attentional inertia hypothesis.

Experiment 3: Inhibition rather than habituation. In both Experiment 1 and 2, no difference was found between familiar target-unfamiliar distractor displays and unfamiliar target-familiar distractor displays in the congruent condition, suggesting that

habituation of familiar colors was unlikely. To further assess whether subjects were slower in responses to familiar colors compared with unfamiliar colors, as the habituation account suggests, here we presented only one shape at a time—either the inside shape or the outside, with either the familiar or the unfamiliar color—and asked subjects to discriminate the shape. A subset of the subjects in Experiment 1 were asked to discriminate circle from square as used in Experiment 1; all the subjects in Experiment 2 and two new subjects were asked to discriminate diamond from square as used in Experiment 2. For the circle and square group, no difference was found between familiar and unfamiliar colors in RTs ($F(1, 6) = 0.18$, $P = 0.690$, $\eta_p^2 = 0.029$; Table 6-2) or error rates ($F(1, 6) = 0.71$, $P = 0.432$, $\eta_p^2 = 0.106$), nor was there a significant interaction between familiarity and size in RTs ($F(1, 6) = 0.34$, $P = 0.583$, $\eta_p^2 = 0.053$) or error rates ($F(1, 6) = 0.62$, $P = 0.462$, $\eta_p^2 = 0.093$). Similarly, for the diamond and square group, no difference was found between familiar and unfamiliar colors in RTs ($F(1, 17) = 0.01$, $P = 0.913$, $\eta_p^2 < 0.001$; Table 6-2) or error rates ($F(1, 17) = 1.00$, $P = 0.332$, $\eta_p^2 = 0.056$), nor was there a significant interaction between familiarity and size in RTs ($F(1, 17) = 0.06$, $P = 0.808$, $\eta_p^2 = 0.004$) or error rates ($F(1, 17) = 0.01$, $P = 0.905$, $\eta_p^2 < 0.001$). The same pattern was found when the data from both groups were combined. There results indicate that subjects were equally fast and accurate in responses to familiar and unfamiliar colors; hence, the inhibition effect observed in Experiment 1 and 2 cannot be attributed to habituation of familiar colors.

Table 6-2. Mean reaction time (in milliseconds) and error rate (in %) in Experiment 3 (either the inside shape or the outside shape was presented, not both, the color of which was either the familiar or the unfamiliar color).

	Target color			
	Circle or square (N = 7)		Diamond or square (N = 18)	
Inside or outside	Familiar	Unfamiliar	Familiar	Unfamiliar
Inside				
Reaction time	450.1 (13.4)	448.8 (11.5)	474.9 (15.2)	473.8 (15.1)
Error rate	2.5 (1.0)	2.8 (1.4)	7.0 (1.2)	6.1 (1.4)
Outside				
Reaction time	428.9 (16.9)	433.6 (14.7)	455.4 (14.2)	455.9 (15.1)
Error rate	2.5 (1.0)	4.7 (1.3)	5.5 (1.0)	4.4 (1.3)

Note: For half of the subjects, the familiar (associated) color was red and the unfamiliar (unassociated) color was green; for the other half, it was the reverse.

6.4 Discussion

Objects in the world can be categorized into either old or new objects. This imprint of familiarity from visual experiences is a core function of memory, enabling the accumulation and coherency of visual experiences throughout lifetime. Visual familiarity is widely assumed to play a fundamental role in perception and action by speeding up attentional selection and processing. The experiments here reported a counter-intuitive observation: after colors were imbued with familiarity through short-term associative learning with targets, incompatible distractors with familiar colors produced less interference effect on target processing in a response competition task than incompatible distractors with unfamiliar colors, indicating an inhibition effect of familiar colors.

This finding of attentional inhibition by visual familiarity prompts us to rethink the role of familiarity in visual attention and perception. The idea that familiar information can be processed more efficiently is evident in many studies in visual search,

as illustrated by the disrupting effects when items are rotated and rendered unfamiliar (Shen & Reingold, 2001; Wang et al., 1994). More dramatically, when familiar information becomes task irrelevant, it may demand attention and cause interference in processing task relevant information. For example, when two Chinese characters differ in the orientation of one stroke, search for one character among tokens of the other character is very efficient for observers unfamiliar with Chinese due to the unique difference in orientation, yet the same task proves to be harder for observers familiar with Chinese, indicating a distracting effect from familiar stimuli (Shen & Reingold, 2001). This interpretation is consistent with a study using schematic facial features: curvature search among simple curves is more efficient when these curves form unfamiliar structures than when they are organized into familiar, schematic facial structures (Suzuki & Cavanagh, 1995). Indeed, research in the Stroop effect provides compelling demonstrations that familiar, task irrelevant words automatically evoke responses, causing interference in ink color naming (Stroop, 1935). These findings thus suggest that familiar, task irrelevant information seems to grab attention automatically. However, this conclusion contradicts other experiences in our daily life, such as our apparent ease at ignoring familiar, irrelevant items in the office, at home, and on the road. We suggest that the key to the resolution of this contradiction is the analysis of the mode of processing during familiarization: human observers become familiar with characters, letters, and faces by continuously associating them with meanings and responses, but become familiar with many everyday stuffs by mere exposure and without associating specific meanings and response to them. We propose that attentional modes or sets during familiarization can carry over to subsequent stages ("attentional inertia"), causing interference effects in Stroop naming and visual search (MacLeod & Dunbar, 1988; Shiffrin & Schneider, 1977). By the very same attentional inertia mechanism, familiarity can induce inhibition rather than facilitation in visual attention when objects are constantly *not* associated with specific responses during familiarization (as we did in the current set of experiments). Our finding of attentional inhibition by visual familiarity is thus consistent with this attentional inertia account. Carryover of attentional modes may be the foundation of automatization in human cognition (Logan, 1988).

Attentional inhibition by visual familiarity is striking considering that the familiar colors are always associated with the targets, so that searching for the familiar colors is a rewarding process and subjects have all the incentive to do so. In previous research, researchers attempt to reduce interference effect from familiar information by asking subjects to learn to ignore distractors through response competition or attentional capture. For example, in a counting Stroop task, subjects count the number of items on a display while ignoring the meaning of the items (e.g. "3, 3, 3, 3"); practicing this task reduces the interference effect, with some transfer effects to distractor stimuli sharing the same meaning as the ignored items (Reisberg, Baron, & Kemler, 1980). Similarly, in an attentional capture task, subjects perform a foveal task while learning to ignore peripheral distractors; practicing this task leads to improvements in information filtering, with some transfer effects to other attentional capture displays (Kelley & Yantis, 2009). These tasks are valuable in elucidating the effect of learning on attentional selection, but the effects can be attributed to processes unrelated to active inhibition, such as habituation. In other words, subjects may become habituated to the distracting information, making the distractor less potent in generating interference effect. In our design, we have taken a completely different approach based on the attentional inertia hypothesis: rather than learning to ignore, subjects actually look for the familiar color, but critically, the familiar color is never associated with a specific label or response. The rationale is that the carry-over effect of *not* associating the familiar color with any specific response would inhibit subsequent responses to them even when the subsequent task involves categorization responses, which is indeed what we have observed. Together, these findings suggest multiple routs to reducing distractor interference, including active inhibition, habituation, and familiarization without labeling.

In conclusion, repeated encounters with objects imbue them with a unique psychological characteristic—familiarity. Familiar objects are processed faster and more efficient. When task irrelevant, familiar objects may demand and capture visual attention if they become familiar by being associating with specific meanings and responses, but can actually be inhibited if they become familiar without being associating with specific meanings and responses.

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science, 14*(5), 427-432.

Addams, R. (1834). An account of a peculiar optical phenomenon seen after having looked at a moving body, etc. *London and Edinburgh Philosophical Magazine and Journal of Science, 3rd series, 5*, 373-374.

Almeida, J., Mahon, B. Z., & Caramazza, A. (2010). The role of the dorsal visual processing stream in tool identification. *Psychol Sci, 21*(6), 772-778.

Almeida, J., Mahon, B. Z., Nakayama, K., & Caramazza, A. (2008). Unconscious processing dissociates along categorical lines. *Proceedings of the National Academy of Sciences of the United States of America, 105*(39), 15214-15218.

Altmann, C. F., Bulthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology, 13*(4), 342-349.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences of the United States of America, 108*(25), 10367-10371.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617-629.

Bex, P. J., Metha, A. B., & Makous, W. (1999). Enhanced motion aftereffect for complex motions. *Vision Research, 39*(13), 2229-2238.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(43), 77-80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143-177.

Blake, R., Tadin, D., Sobel, K. V., Raissian, T. A., & Chong, S. C. (2006). Strength of early visual adaptation depends on visual awareness. *Proc Natl Acad Sci U S A, 103*(12), 4783-4788.

Boi, M., Ogmen, H., & Herzog, M. H. (2011). Motion and tilt aftereffects occur largely in retinal, not in object, coordinates in the Ternus-Pikler display. *Journal of Vision, 11*(3).

Boi, M., Ogmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision, 9*(13), 5 1-11.

Boi, M., Vergeer, M., Ogmen, H., & Herzog, M. H. (2011). Nonretinotopic exogenous attention. *Current Biology, 21*(20), 1732-1737.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745-4765.

Burr, D. C., & Morrone, M. C. (2011). Spatiotopic coding and remapping in humans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 366*(1564), 504-515.

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Research, 51*(13), 1484-1525.

Cavanagh, P., Holcombe, A. O., & Chou, W. (2008). Mobile computation: spatiotemporal integration of the properties of objects in motion. *Journal of Vision, 8*(12), 1 1-23.
Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences, 14*(4), 147-153.
Chou, W. L., & Yeh, S. L. (2012). Object-based attention occurs regardless of object awareness. *Psychon Bull Rev, 19*(2), 225-231.
Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*(5), 170-178.
Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28-71.
Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10*, 360-365.
Chun, M. M., & Johnson, M. K. (2011). Memory: enduring traces of perceptual and reflective attention. *Neuron, 72*(4), 520-535.
Cichy, R. M., Chen, Y., & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *Neuroimage, 54*(3), 2297-2307.
Cook, E. P., & Maunsell, J. H. (2004). Attentional modulation of motion integration of individual neurons in the middle temporal visual area. *J Neurosci, 24*(36), 7964-7977.
Corballis, M. C., & Beale, I. L. (1976). *The psychology of left and right*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201-215.
Cornelissen, F. W., Wade, A. R., Vladusich, T., Dougherty, R. F., & Wandell, B. A. (2006). No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex. *Journal of Neuroscience, 26*(14), 3634-3641.
Culham, J., Nishida, S., Ledgeway, T., Cavanagh, P., von Grunau, M., Kwas, M., et al. (1998). Higher order effects. In G. Mather, F. A. J. Verstraten & S. Anstis (Eds.), *The Motion-Aftereffect: A Modern Perspective* (pp. 85-124). Cambridge MA: MIT.
d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience, 10*(2), 249-255.
Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychol Sci, 15*(8), 559-564.
Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature, 371*(6500), 791-793.
Davis, G., & Driver, J. (1997). Spreading of visual attention to modally versus amodally completed regions. *Psychological Science, 8*(4), 275-281.
Davis, G., & Driver, J. (1998). Kanizsa subjective figures can act as occluding surfaces at parallel stages of visual search. *Journal of Experimental Psychology-Human Perception and Performance, 24*(1), 169-184.

De Weerd, P., Gattass, R., Desimone, R., & Ungerleider, L. G. (1995). Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. *Nature, 377*(6551), 731-734.

Dennett, D. C. (1992). Filling in versus finding out: A ubiquitous confusion in cognitive science. In H. L. Pick Jr., P. van den Broek & D. C. Knill. (Eds.), *Cognition: Conceptual and methodological issues*. Washington, DC: American Psychological Association.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America, 93*(24), 13494-13499.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn Sci, 11*(8), 333-341.

Dodd, M. D., & Wilson, D. (2009). Training attention: Interactions between central cues and reflexive attention. *Visual Cognition, 17*(5), 736-754.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science, 11*(6), 467-473.

Driver, J., Baylis, G. C., Goodrich, S. J., & Rafal, R. D. (1994). Axis-based neglect of visual shapes. *Neuropsychologia, 32*(11), 1353-1365.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509-540.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*(5040), 90-92.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology, 48*, 269-297.

Eimer, M. (1997). Uninformative symbolic cues may bias visual-spatial attention: Behavioral and electrophysiological evidence. *Biological Psychology, 46*(1), 67-71.

Enns, J. T., & DiLollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science, 8*(2), 135-139.

Enns, J. T., Lleras, A., & Moore, C. M. (2010). Object updating: a force for perceptual continuity and scene stability in human vision. In R. Nijhawan (Ed.), *Space and Time in Perception and Action* (pp. 503-520). Cambridge, UK: Cambridge University Press

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat Neurosci, 8*(10), 1380-1385.

Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision, 8*(7), 2 1-9.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47.

Fiorani Junior, M., Rosa, M. G., Gattass, R., & Rocha-Miranda, C. E. (1992). Dynamic surrounds of receptive fields in primate striate cortex: a physiological basis for perceptual completion? *Proceedings of the National Academy of Sciences of the United States of America, 89*(18), 8547-8551.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience, 6*(6), 555-556.
Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci, 12*(6), 499-504.
Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General, 134*(4), 521-537.
Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance, 18*(4), 1030-1044.
Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception and Psychophysics, 65*(7), 999-1010.
Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495.
Frith, U. (1974). A curious effect with reversed letters explained by a theory of schema. *Perception and Psychophysics, 16*(1), 113-116.
Galfano, G., Rusconi, E., & Umilta, C. (2006). Number magnitude orients attention, but not against one's will. *Psychonomic Bulletin & Review, 13*(5), 869-874.
Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience, 28*(15), 3988-3999.
Gerrits, H. J., & Vendrik, A. J. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. *Experimental Brain Research, 11*(4), 411-430.
Gibson, B. S., & Egeth, H. (1994). Inhibition of return to object-based and environment-based locations. *Perception and Psychophysics, 55*(3), 323-339.
Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron, 9*(1), 1-13.
Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences of the United States of America, 93*(2), 615-622.
Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin, 132*(6), 920-945.
Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*(4), 480-506.
Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
Gregory, R. L. (1972). Cognitive contours. *Nature, 238*(5358), 51-52.
Gregory, R. L. (1997). *Eye and brain : the psychology of seeing* (5th ed.). Princeton, N.J.: Princeton University Press.
Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10-11), 1409-1422.
Gross, C. G., & bornstein, M. H. (1978). Left and right in science and art. *Leonardo, 11*, 29-38.
Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics, 55*(1), 48-121.

Halgren, E., Mendola, J., Chong, C. D. R., & Dale, A. M. (2003). Cortical activation to illusory shapes as measured with magnetoencephalography. *Neuroimage, 18*(4), 1001-1009.

Hayhoe, M., Lachter, J., & Feldman, J. (1991). Integration of form across saccadic eye movements. *Perception, 20*(3), 393-402.

He, D., Kersten, D., & Fang, F. (2012). Opposite Modulation of High- and Low-Level Visual Aftereffects by Perceptual Grouping. *Current Biology*.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359*(6392), 231-233.

Hegde, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas v2 and v4. *Cerebral Cortex, 17*(5), 1100-1116.

Hock, H. S., Romanski, L., Galie, A., & Williams, C. S. (1978). Real-world schemata and scene recognition in adults and children. *Memory & Cognition, 6*(4), 423-431.

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience, 4*(2), 127-128.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science, 12*(5), 360-365.

Horwitz, G. D., & Hass, C. A. (2012). Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing. *Nature Neuroscience*.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5*, 94-100.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194-203.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 54*(4), 1105-1124.

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science, 18*(4), 349-355.

Jordan, H., & Tipper, S. P. (1999). Spread of inhibition across an object's surface. *British Journal of Psychology, 90*, 495-507.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology, 24*(2), 175-219.

Kanai, R., & Tsuchiya, N. (2012). Qualia. *Current Biology, 22*(10), R392-396.

Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception*. New York: Praeger.

Kelley, T. A., & Yantis, S. (2009). Learning to attend: Effects of practice on information selection. *Journal of Vision, 9*(7).

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognit Psychol, 23*(2), 141-221.

Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nature Reviews Neuroscience, 7*(3), 220-231.

Kourtzi, Z., Tolias, A. S., Altmann, C. F., Augath, M., & Logothetis, N. K. (2003). Integration of local features into global shapes: monkey and human FMRI studies. *Neuron, 37*(2), 333-346.

Kristjansson, A., Mackeben, M., & Nakayama, K. (2001). Rapid, object-based learning in the deployment of transient attention. *Perception, 30*(11), 1375-1387.
Kristjansson, A., & Nakayama, K. (2003). A primitive memory system for the deployment of transient attention. *Perception and Psychophysics, 65*(5), 711-724.
Lak, A. (2008). Attention during adaptation weakens negative afterimages of perceptually colour-spread surfaces. *Canadian Journal of Experimental Psychology, 62*(2), 101-109.
Lalanne, C., & Lorenceau, J. (2006). Directional shifts in the barber pole illusion: effects of spatial frequency, spatial adaptation, and lateral masking. *Visual Neuroscience, 23*(5), 729-739.
Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6*(5), 541-567.
Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35*(5), 1346-1358.
Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proc Natl Acad Sci U S A, 98*(4), 1907-1911.
Leventhal, A. G., & Schall, J. D. (1983). Structural basis of orientation sensitivity of cat retinal ganglion cells. *Journal of Comparative Neurology, 220*(4), 465-475.
Levick, W. R., & Thibos, L. N. (1982). Analysis of orientation bias in cat retina. *J Physiol, 329*, 243-261.
Li, W., & Gilbert, C. D. (2002). Global contour saliency and local colinear interactions. *Journal of Neurophysiology, 88*(5), 2846-2856.
Li, W., Piech, V., & Gilbert, C. D. (2008). Learning to link visual contours. *Neuron, 57*(3), 442-451.
Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation, 10*(4), 903-940.
Lin, J. Y., Franconeri, S., & Enns, J. T. (2008). Objects on a collision path with the observer demand attention. *Psychological Science, 19*(7), 686-692.
Lin, J. Y., Murray, S. O., & Boynton, G. M. (2009). Capture of attention to threatening stimuli without perceptual awareness. *Current Biology, 19*(13), 1118-1122.
Lin, Z., & He, S. (2009). Seeing the invisible: the scope and limits of unconscious processing in binocular rivalry. *Progress in Neurobiology, 87*(4), 195-211.
Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, 240*(4853), 740-749.
Lleras, A., & Moore, C. M. (2003). When the target becomes the mask: using apparent motion to isolate the object-level component of object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 29*(1), 106-120.
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492-527.
Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research, 32*(2), 263-273.
Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology, 18*(10), R410-R412.

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science, 21*(5), 682-691.
MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 126-135.
MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383-391.
Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory and Cognition, 22*(6), 657-672.
Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception and Psychophysics, 58*(7), 977-991.
Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 529-540.
Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 275*(942), 483-519.
Marr, D. (1982). *Vision*. Cambridge, MA: MIT Press.
Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 200*(1140), 269-294.
Marr, D., & Ullman, S. (1981). Directional Selectivity and Its Use in Early Visual Processing. *Proceedings of the Royal Society of London Series B-Biological Sciences, 211*(1183), 151-&.
Mather, G., Verstraten, F., & Anstis, S. (1998). *The Motion Aftereffect: A Modern Perspective*. Cambridge, MA: The MIT Press.
Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science, 275*(5300), 671-674.
Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience, 6*(5), 450-452.
McClelland, J. L. (2010). Emergence in Cognitive Science. *Topics in Cognitive Science, 2*(4), 751-770.
Melcher, D., & Colby, C. L. (2008). Trans-saccadic perception. *Trends in Cognitive Sciences, 12*(12), 466-473.
Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience, 6*(8), 877-881.
Mendola, J. D., Dale, A. M., Fischl, B., Liu, A. K., & Tootell, R. B. (1999). The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. *Journal of Neuroscience, 19*(19), 8560-8572.
Meng, M., Remus, D. A., & Tong, F. (2005). Filling-in of visual phantoms in the human brain. *Nat Neurosci, 8*(9), 1248-1254.
Meng, X., Mazzoni, P., & Qian, N. (2006). Cross-fixation transfer of motion aftereffects with expansion motion. *Vision Research, 46*(21), 3681-3689.

Michotte, A., Thines, G., & Crabbe, G. (1964). *Les complements amodaux des structures perceptives*. Louvain, France: Studia Psycologica Louvain, Publications Universitaires de Louvain.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia, 46*(3), 774-785.

Minsky, M. (1975). The psychology of computer vision. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211-277): McGraw-Hill.

Mitchell, C. J., & Le Pelley, M. E. (2010). *Attention and associative learning: from brain to behaviour*. New York: Oxford University Press.

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science, 9*(2), 104-110.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6*(2), 182-189.

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration Without Awareness: Expanding the Limits of Unconscious Processing. *Psychological Science, 22*(6), 764-770.

Mulckhuyse, M., Talsma, D., & Theeuwes, J. (2007). Grabbing attention without knowing: Automatic capture of attention by subliminal spatial cues. *Visual Cognition, 15*(7), 779-788.

Murray, M. M., Foxe, D. M., Javitt, D. C., & Foxe, J. J. (2004). Setting boundaries: Brain dynamics of modal and amodal illusory shape completion in humans. *Journal of Neuroscience, 24*(31), 6898-6903.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 99*(23), 15164-15169.

Murray, S. O., Schrater, P., & Kersten, D. (2004). Perceptual grouping and the interactions between visual cortical areas. *Neural Networks, 17*(5-6), 695-705.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29*(11), 1631-1647.

Neil, W. T., & Mathis, K. M. (1998). Transfer-inappropriate processing: Negative priming and related phenomena. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 38, pp. 1-44). San Diego, CA: Academic Press.

Neill, W. T., & Valdes, L. A. (1992). Persistence of negative priming: Steady state or decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 565-576.

Neill, W. T., Valdes, L. A., Terry, K. M., & Gorfein, D. S. (1992). Persistence of negative priming: II. Evidence for episodic trace retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 993-1000.

Nishida, S., Ashida, H., & Sato, T. (1994). Complete interocular transfer of motion aftereffect with flickering test. *Vision Research, 34*(20), 2707-2716.

Nishida, S., Watanabe, J., Kuriki, I., & Tokimoto, T. (2007). Human visual system integrates color signals along a motion trajectory. *Current Biology, 17*(4), 366-372.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends Cogn Sci, 11*(12), 520-527.
Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13*(11), 4700-4719.
Olson, C. R. (2003). Brain representation of object-centered space in monkeys and humans. *Annual Review of Neuroscience, 26*, 331-354.
Olson, C. R., & Gettner, S. N. (1995). Object-centered direction selectivity in the macaque supplementary eye field. *Science, 269*(5226), 985-988.
Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature, 409*(6823), 1040-1042.
Palmer, S. E., Neff, J., & Beck, D. (1996). Late influences on perceptual grouping: Amodal completion. *Psychonomic Bulletin & Review, 3*(1), 75-80.
Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of Neurophysiology, 86*(5), 2505-2519.
Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience, 5*(12), 1332-1338.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.
Perna, A., Tosetti, M., Montanaro, D., & Morrone, M. C. (2005). Neuronal mechanisms for illusory brightness perception in humans. *Neuron, 47*(5), 645-651.
Pertzov, Y., Zohary, E., & Avidan, G. (2010). Rapid formation of spatiotopic representations as revealed by inhibition of return. *Journal of Neuroscience, 30*(26), 8882-8887.
Pessoa, L., Thompson, E., & Noe, A. (1998). Filling-in is for finding out. *Behavioral and Brain Sciences, 21*(6), 781-802.
Pillow, J., & Rubin, N. (2002). Perceptual completion across the vertical meridian and the role of early visual cortex. *Neuron, 33*(5), 805-813.
Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *J Exp Psychol Hum Percept Perform, 15*(4), 635-649.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-26.
Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma. & D. G. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes* (pp. 531–556). Hillsdale, NJ: Erlbaum.
Price, N. S., Greenwood, J. A., & Ibbotson, M. R. (2004). Tuning properties of radial phantom motion aftereffects. *Vision Research, 44*(17), 1971-1979.
Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821-824.
Rauschenberger, R., & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature, 410*(6826), 369-372.
Reisberg, D., Baron, J., & Kemler, D. G. (1980). Overcoming Stroop interference: the effects of practice on distractor potency. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 140-150.

Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research, 38*(15-16), 2489-2505.
Ristic, J., Wright, A., & Kingstone, A. (2006). The number line effect reflects top-down control. *Psychonomic Bulletin & Review, 13*(5), 862-868.
Rock, I. (1984). *Perception*. New York: Scientific American Library : Distributed by W.H. Freeman.
Rock, I. (1986). The description and analysis of object and event perception. In K. R. Boff., L. R. Kaufman. & J. P. Thomas. (Eds.), *Handbook of Perception and Human Performance* (Vol. II Cognitive Processes and Performance, pp. 1:71). New York: John Wiley.
Roe, A. W., Lu, H. D., & Hung, C. P. (2005). Cortical processing of a brightness illusion. *Proceedings of the National Academy of Sciences of the United States of America, 102*(10), 3869-3874.
Rudel, R. G., & Teuber, H. L. (1963). Discrimination of direction of line in children. *Journal of Comparative and Physiological Psychology, 56*, 892-898.
Sakuraba, S., Sakai, S., Yamanaka, M., Yokosawa, K., & Hirayama, K. (2012). Does the Human Dorsal Stream Really Process a Category for Tools? *Journal of Neuroscience, 32*(11), 3949-3953.
Sasaki, Y., & Watanabe, T. (2004). The primary visual cortex fills in color. *Proc Natl Acad Sci U S A, 101*(52), 18251-18256.
Schank, R. C. (1975). *Conceptual information processing*. New York: Elsevier.
Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105*(11), 4447-4452.
Seghier, M., Dojat, M., Delon-Martin, C., Rubin, C., Warnking, J., Segebarth, C., et al. (2000). Moving illusory contours activate primary visual cortex: an fMRI study. *Cerebral Cortex, 10*(7), 663-670.
Seghier, M. L., & Vulleumier, P. (2006). Functional neuroimaging findings on the human perception of illusory contours. *Neuroscience and Biobehavioral Reviews, 30*(5), 595-612.
Shen, J., & Reingold, E. M. (2001). Visual search asymmetry: the influence of stimulus familiarity and low-level features. *Perception and Psychophysics, 63*(3), 464-475.
Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 38*(3), 475-491.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.
Shimojo, S., Kamitani, Y., & Nishida, S. (2001). Afterimage of perceptually filled-in surface. *Science, 293*(5535), 1677-1680.
Singh, M. (2004). Modal and amodal completion generate different shapes. *Psychological Science, 15*(7), 454-459.
Smith, A., & Over, R. (1975). Tilt aftereffects with subjective contours. *Nature, 257*(5527), 581-582.

Snowden, R. J., & Milne, A. B. (1997). Phantom motion aftereffects - Evidence of detectors for the analysis of optic flow. *Current Biology, 7*(10), 717-722.
Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: evidence for a dual representation of object shape. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 732-744.
Stanley, D. A., & Rubin, N. (2003). fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron, 37*(2), 323-331.
Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences of the United States of America, 109*(6), E360-367.
Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.
Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron, 49*(6), 905-916.
Sutherland, N. S. (1957). Visual discrimination of orientation and shape by the octopus. *Nature, 179*, 11-13.
Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: an object inferiority effect. *Journal of Experimental Psychology: Human Perception & Performance, 21*(4), 901-913.
Sweeny, T. D., Grabowecky, M., & Suzuki, S. (2011). Awareness becomes necessary between adaptive pattern coding of open and closed curvatures. *Psychol Sci, 22*(7), 943-950.
Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception and Psychophysics, 51*(6), 599-606.
Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77-99.
Tipper, S. P. (1985). The negative priming effect: inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 37*(4), 571-590.
Tipper, S. P., & Behrmann, M. (1996). Object-centered not scene-based visual neglect. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1261-1278.
Tipper, S. P., Driver, J., & Weaver, B. (1991). Object-centred inhibition of return of visual attention. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 43*(2), 289-298.
Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*(4), 753-759.
Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci, 5*, 42.
Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge Cambridgeshire ; New York: Cambridge University Press.
Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist, 47*(7), 862-875.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*(2), 171-178.
Treisman, A. (1999). Solutions to the binding problem: progress through controversy and convergence. *Neuron, 24*(1), 105-110, 111-125.
Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.
Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat Neurosci, 8*(8), 1096-1101.
Tynan, P., & Sekular, R. (1975). Moving visual phantoms: a new contour completion effect. *Science, 188*(4191), 951-952.
Ullman, S. (1976). Filling-in Gaps - Shape of Subjective Contours and a Model for Their Generation. *Biological Cybernetics, 25*(1), 1-6.
Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Mass.: MIT Press.
Umilta, C., Castiello, U., Fontana, M., & Vestri, A. (1995). Object-centred orienting of attention. *Visual Cognition, 2*(2/3), 165-181.
Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4*(2), 157-165.
Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle., M. A. Goodale. & R. J. W. Mansfield. (Eds.), *Analysis of Visual Behavior*. Cambridge, MA: MIT Press.
Verghese, P., & Stone, L. S. (1996). Perceived visual speed constrained by image segmentation. *Nature, 381*(6578), 161-163.
Voytek, B., Soltani, M., Pickard, N., Kishiyama, M. M., & Knight, R. T. (2012). Prefrontal cortex lesions impair object-spatial integration. *PLoS ONE, 7*(4), e34937.
Wade, N. J., & Wenderoth, P. (1978). The influence of colour and contour rivalry on the magnitude of the tilt aftereffect. *Vision Res, 18*, 827-835.
Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron, 56*(2), 366-383.
Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics, 56*(5), 495-500.
Watanabe, J., & Nishida, S. (2007). Veridical perception of moving colors by trajectory integration of input signals. *Journal of Vision, 7*(11), 3 1-16.
Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: an object-superiority effect. *Science, 186*(4165), 752-755.
Weisstein, N., Maguire, W., & Berbaum, K. (1977). A phantom-motion aftereffect. *Science, 198*(4320), 955-958.
Weyl, H. (1952). *Symmetry*. Princeton: Princeton University Press.
Wilson, F. A., Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260*(5116), 1955-1958.
Wolfe, J. M. (2001). Asymmetries in visual search: an introduction. *Perception and Psychophysics, 63*(3), 381-389.
Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24*(1), 11-17, 111-125.

Wuhr, P., & Frings, C. (2008). A case for inhibition: visual attention suppresses the processing of irrelevant objects. *Journal of Experimental Psychology: General, 137*(1), 116-130.

Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research, 48*(20), 2070-2089.

Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the cortical hierarchy: low-level curve adaptation affects high-level facial-expression judgments. *Journal of Neuroscience, 28*(13), 3374-3383.

Yang, Y. H., & Yeh, S. L. (2011). Accessing the meaning of invisible words. *Consciousness and Cognition, 20*(2), 223-233.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (can) prime furiously. *Psychological Science, 23*(4), 364-369.

Zhang, P., Jiang, Y., & He, S. (2012). Voluntary attention modulates processing of eye-specific visual information. *Psychol Sci, 23*(3), 254-260.

Zhaoping, L., & Guyader, N. (2007). Interference with bottom-up feature detection by higher-level object recognition. *Current Biology, 17*(1), 26-31.