Structured Representations in Visual Cognition

Qi-Yang Nie

München 2017

Date of Oral Examination: December 6th, 2017

Structured Representations in Visual Cognition

Inaugural-Dissertation
zur Erlangung des
Doktorgrades der Philosophie an der
Ludwig-Maximilians-Universität
München

vorgelegt von

Qi-Yang Nie

aus München

München, November 2017

Referent:	PD. Dr. Markus Conci
Koreferent:	Prof. Dr. Hermann J. Müller

Table of Contents

Chapter I

Introduction

Visual scenes provide a complex and cluttered input to the visual system, which usually consists of a hierarchical organization, which can be defined as a "multilevel hierarchical structure of parts and wholes" (Palmer, 1977). For instance, a forest is composed of trees, and the trees in turn are composed of parts, e.g. branches, leaves and so forth, illustrating that global wholes and local parts are linked by means of some relational structure.

One of the most enduring issues in vision science concerns the perceptual relations between wholes and their parts. The question is whether processing of the overall structure precedes and determines the processing of the component parts and their properties, or whether the parts are registered first and are then synthesized to form higher global-level objects.

Traditionally, this classical topic can be traced back to the controversy between two schools of perceptual psychology, namely structuralism and Gestalt psychology. The structuralists (e.g., Titchener, 1909; Wundt, 1874) were rooted firmly in British empiricism, with its emphasis on atomism and associative mechanisms, and were influenced by 19th-century physiological view (Wundt, 1874). They held that the basic units of perception are independent local sensations and their physiological counterparts, specific nerve energies. In this view, every sensory whole must be built up from a conglomerate of elementary sensations, and the perception of segregated, organized units corresponding to objects in the physical world is achieved only by associations learned through experience.

The Gestaltists (e.g., Koffka, 1963; Kohler, 1929, 1971; Wertheimer, 1967), on the other hand, argued against both the atomistic assumption and the role of learning in perception, but emphasized the primacy of whole units and the organization into a complete percept. A basic tenet of the Gestalt view is that a specific sensory whole is qualitatively different from the complex that one might predict by considering only its parts. The whole quality is not just one more added element or factor, as was proposed by Ehrenfels's (1890) “Gestalt Qualität”, nor does it arise (through the agency of any auxiliary factor) as a secondary process from the sum of the pieces as such. Instead, it has been suggested that “what takes place in each single part already depends upon what the whole is” (Wertheimer, 1967). Thus, the quality of a part is determined by the whole in which this part is integrated. According to the Gestalt theory, the perception of distinct, organized units is not the product of sensory elements tied together by associative learning but is, instead, an immediate product of electrical field processes in the brain that respond to the entire pattern of stimulation (Wertheimer, 1967).

The basic flavor of the structuralist approach has been retained in many current models of perception, especially models of pattern and object recognition (see Treisman, 1986, for an extensive review). Such analytic models assume that objects are identified, recognized, and classified by detecting combinations of elementary features.

In the last 40 years, the Gestaltist view of perception has nevertheless recaptured the interest of cognitive psychologists (e.g., Beck, 1982; Boff, Kaufman, & Thomas, 1986, Vol. 2; Gopher & Kimchi, 1989; Kubovy & Pomerantz, 1981; Shepp & Ballesteros, 1989). This revival includes work on issues such as configural-superiority effects and global/local processing. It is also expressed in the growing usage of the term wholistic rather than analytic to describe perception (e.g., Uttal, 1988).

The distinction between wholistic versus analytic processing is sometimes referred to as a distinction between top-down versus bottom-up processing (e.g., Kinchla, Solis-Macias, & Hoffman, 1983; Kinchla & Wolfe, 1979). However, the terms top-down and bottom-up processing are often used to refer to the distinction between conceptually driven processing on the one hand and data-driven processing on the other (e.g., Lindsey & Norman, 1977; Rumelhart, 1977). The issue of wholistic/analytic processing is orthogonal to this latter usage of the terms top-down and bottom-up. For instance, whether the processing of the stimulus starts with an analysis of sensory information (i.e., bottom-up processing) or with an internal hypothesis that guides processing (i.e., top-down processing) does not necessarily imply which stimulus aspects will be processed first (see also Kimchi & Palmer, 1982; Navon, 1981b; Pomerantz, 1981; Treisman, 1986).

In the following, this chapter introduces some modern attempts to grapple with the issue of part-whole relationships: configural-superiority (Pomerantz, Sager, & Stoever, 1977) and global precedence (Navon, 1977). I begin with the presentation of the emergent feature (EF) hypothesis and the odd-quadrant paradigm, followed by a brief review of the empirical findings concerning the boundary conditions of the configural-superiority effect, its source and its concurrent brain localization. The subsequent section then focuses on the global-precedence hypothesis and the global/local paradigm that typically employs task, which present single hierarchical objects. I then discuss some issues concerning the interpretation of the global precedence effect, and also propose visual search and change-detection variants of hierarchical object processing, and devise a novel hierarchical stimulus that has representation fidelity in a continuous feature space. I close by briefly considering the implications of these novel approaches for the understanding the perception of hierarchical structure and/or part-whole relationships in object configurations.

Configural superiority

Pomerantz et al. (1977) proposed that Emergent features, or EFs are features that are possessed by wholes — groups of parts — but not by any individual part nor by any single group of parts. Thus, they emerge when parts combine into wholes. Wholes can have fewer or more Gestalt qualities because they possess fewer or more EFs. If a set of trees is closely spaced, proximity and similarity lead them to be perceptually grouped into a whole forest, and that forest has properties (such as density), which is not possessed by any individual tree. If the trees are planted into regularly spaced rows, however, they now potentially gain EFs such as collinearity and symmetry that go beyond the mere clustering of parts into bunches.

Only some EFs give rise to configural superiority effects or CSEs (Pomerantz et al., 1977), which can be used as an index to indicate when wholes are perceived before parts (i.e., when the forest comes before trees). The easiest test for CSEs starts with benchmarking performance in a baseline task of localizing a singleton (or an odd-one-out element) in a search display, for example, finding a single letter B in a display that otherwise consists of As. Then an identical context stimulus, e.g., the letters C, are added to each element so the task is now to locate the sole BC in a field of ACs (see Figure 1A). Normally, adding identical, noninformative context hurts performance because it makes the stimuli more similar (in addition to increasing overall processing load and possibly introducing masking or crowding). This is the case with these letter stimuli: Participants take longer to find the BC in a field of ACs than to find the B in a field of As.

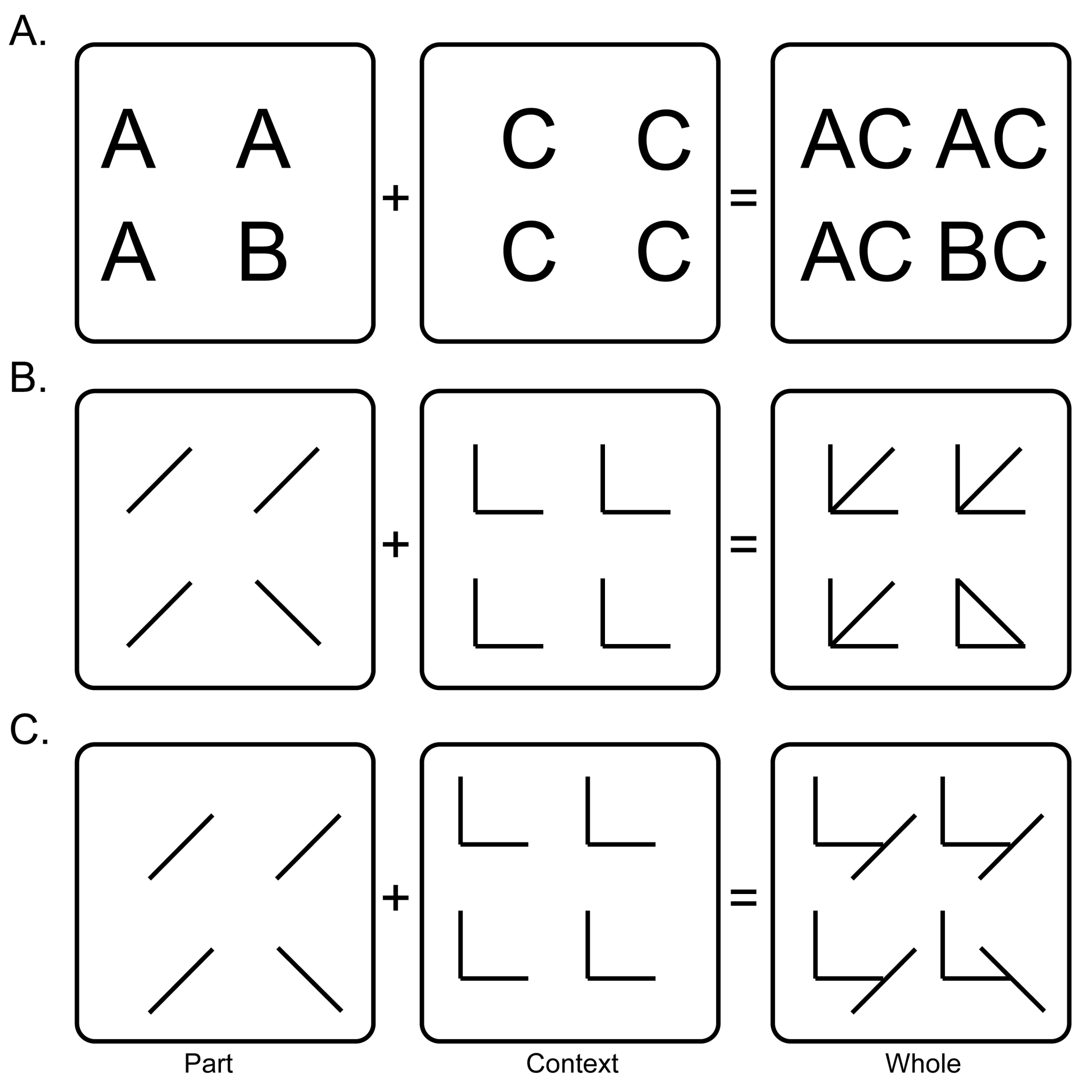

Figure 1. Panel A shows a schematic of the odd-quadrant discrimination task. Participants see either a part or a whole display (the context display is only presented for illustrative purpose). In the example, A, B, and C are placeholders for several possible stimuli. Panel B shows a corresponding example of a stimulus set that leads to a reliable configural superiority effect, whereas panel C illustrates an example that yields a configural inferiority effect (adapted from Pomerantz & Portillo, 2011).

With other parts other than A, B, and C, however, the opposite result can arise, which constitutes evidence for configural superiority. If diagonal line segments and an L-shaped corner are used for A, B, and C so that the diagonals combine with the Ls to form arrows and triangles, perceivers are more than twice as fast to spot the target (see Figure 1B). When these same parts are shifted just slightly in position, however, the CSE is lost (see Figure 1C).

The key factor in obtaining a CSE appears to be the creation of salient, emergent features when the context C is added to the base elements A and B. With the arrows and triangles of Figure 1B, those EFs appear to be closure, number of terminators, and type of intersection. Some of the strongest, most robust CSEs discovered involve topological EFs such as presence versus absence of holes, connectivity, and inside–outside relationships (Chen, 2005). (For more CSEs and a new framework called the theory of basic Gestalts, see Pomerantz & Portillo, 2011.) However, it remains unknown whether CSEs also exist for illusory Gestalts such as Kanizsa figures (Kanizsa, 1979), especially when this kind of illusory figures serve as the targets or distractors in a visual search array. So whether it is distractor inhibition or target facilitation that drives the CSE remains to be tested.

Global precedence

Navon's (1977) global precedence hypothesis states that processing proceeds from global structures towards an analysis of local properties. This hypothesis was formulated within a framework that views a visual object as represented by a hierarchical network with nested relationships. The globality of a visual property corresponds to the level it occupies within the hierarchy: Properties at the top of the hierarchy are more global than those at the bottom, which are in turn more local. Consider a face defined by the spatial relationship between facial components (e.g., eyes, nose, mouth), which are, in turn, defined by relationships among their subparts. The spatial relationship between the components is more global than the specific shapes of the components, and in turn, the relationship between the subparts of a component is more global than the specific properties of the subparts. The global precedence hypothesis claims that the order of processing of an object is from global to local: Global properties of a visual object are processed first, followed by analysis of local properties. It has been tested with hierarchical patterns, in which larger figures are constructed from smaller figures (first introduced by Asch, 1962, and later by Kinchla, 1974, 1977). An example is a set of hierarchical letters: large letters constructed from the same set of smaller letters having either the same identity as the larger letter or a different identity. Hierarchical patterns like these satisfy two conditions, which are critical for testing the global precedence hypothesis (Navon, 1977): First, the global and local structures can be equated in terms of familiarity, complexity, codability, and identifiability, so that they differ only in their level of globality; and second, the two structures can be independent, so that one structure cannot be predicted from the other.

In a popular paradigm, observers are presented with hierarchical stimuli and are required to identify the larger (global) or the smaller (local) letter, in separate blocks of trials. Findings of global advantage — faster identification of the global letter than the local letter and a disruptive influence from irrelevant global conflicting information on local identification (global-to-local interference) — are taken as support for global precedence (e.g., Navon, 1977, Experiment 3). Much subsequent research has concentrated on delineating the boundary conditions of the global advantage and examined whether its locus is perceptual or postperceptual (for reviews, see Kimchi, 1992; Navon, 2003). Results indicate that several factors can modulate global precedence, including overall size, eccentricity, spatial uncertainty, elements' sparseness, number of elements, relative size of elements, figural goodness, exposure duration, and attention allocation (Kimchi, 1992). Research indicates that the global advantage — when it occurs — arises at the perceptual level, although the effect can be magnified by postperceptual, response-related processes (Miller & Navon, 2002).

Overall, a global advantage is usually observed with the typical hierarchical stimuli used in the global/local paradigm to the limits of visibility and visual acuity. Nonetheless, the fact that a global advantage is obtained only under certain conditions suggests that global precedence is not a universal law. Two main issues have been raised concerning the interpretation of global advantage. One issue concerns the hierarchical patterns that are the cornerstone of the global/local paradigm. Hierarchical patterns provide an elegant control for many intervening variables while keeping the hierarchical structure transparent, but the local elements of the hierarchical patterns do not really form the parts of the whole (Kimchi, 1992; Navon, 2003). Furthermore, it has been argued that the local elements in the Navon type of hierarchical patterns function merely as placeholders (Pomerantz, 1983) or serve just to define texture (Kimchi & Palmer, 1982; Pomerantz, 1983; but see Navon, 2003). If so, the local elements may not be represented as figural units, and consequently, faster identification of the global form may be accounted for not by its level of globality but by a qualitative difference in identification of a figural unit versus a texture element. However, a study on the microgenesis, i.e., the development of global precedence over time (Kimchi, 1998) showed with hierarchical stimuli in a primed matching paradigm that the global form was primed at rather brief exposures, whereas the local elements were primed only at longer exposures, suggesting that the global form is effective already early in the perceptual process, followed by the subsequent individuation of the local elements.

The second issue is that relative size alone rather than globality could explain the global advantage (e.g., Kinchla & Wolfe, 1979; Navon & Norman, 1983). Navon (2003) argued that globality is inherently confounded with relative size — it is a fact of nature that relative size is "an inherent concomitant of part–whole relationship." This is indeed the case if global properties are properties of a higher-level unit. Yet, if global properties depend on the relationship between the elements, as the theoretical motivation for the global precedence hypothesis implies (e.g., Navon, 1977, 2003), then the essential difference between global properties and component properties is not in their relative size but their relative position in the object hierarchy. For example, to distinguish the "squareness" from its component vertical and horizontal lines or the "faceness" from its facial components based only on their relative sizes would miss the point (Kimchi, 1992).

The vast majority of results demonstrate that perceptual processing can proceed from global structuring towards analysis of local properties under certain conditions (hence, global precedence). Further findings also suggest that there are different kinds of wholes with different kinds of parts and part–whole relationships. Consider a face with its eyes, nose, and mouth, versus a wall of bricks. Both are complex visual objects — wholes — but the eyes, nose, and mouth of a face are its parts, whereas the bricks in the wall are mere constituents. It is therefore possible that global precedence characterizes the course of processing of some wholes but not of others. Importantly, most of these paradigms have presented observers with single hierarchical objects. Accordingly, evidence for global precedence in these configurations usually reflects differences in processing between the hierarchical levels of a stimulus that is currently in the focus of attention. However, global precedence may also, at least partially, occur for non-attended objects (Paquet & Merikle, 1988) at preattentive stages of processing (Mattingley, Davis, & Driver, 1997; Conci et al., 2009). With multiple hierarchical stimulus configurations – as, for example, in visual search or change detection tasks – the question is not whether focal attention is set in accordance with the different hierarchical object levels, but whether both the guidance of attention and working memory maintenance of preattentive object feature are sensitive to differences between global and local representations.

Finally, the type of compound letters (Navon, 1977) and composite shapes (Kimchi & Palmer, 1982) often used to examine the global/local structure of visual perception cannot be used to probe the fidelity of vWM representations because they only allow for discrete changes (e.g., from a triangle to a square at either global or local levels). It remains to be seen whether this limitation can be overcome with a novel stimulus that permits continuous changes to be implemented at both global and local levels.

Overview of the current study

The aim of dissertation is to determine attention and memory functions reflect our structured visual environment. For instance, computer algorithms have provided a useful analogy for thinking about cognition, and in the same manner can object or scene structures provide a useful analogy for thinking about the structured representations of attention and visual memory. The dissertation builds on the Gestalt tradition that attention and (short-term) memory systems are hierarchically organized: it is a hierarchical process with a default global state that explains much of its usual work, but which can be observed, retained, and adapted by task demands. The default global state of selection and maintenance is well described by an evidence accumulation process operating over the visual units by means of attention. Structured representations (global/local) in the current attentional state can also be transferred in to subsequent instances, such as in cross-trial priming, and the underlying dynamics of attention and implicit short-term memory are dissociable or independent from each other. Visual memories can be observed through global/local object structure, a maintenance process that tracks the structure of a memorandum as it is no longer available in visual space. And the default memory state can be controlled and adapted by task demands, which reflects flexible maintenance in accordance with the task goals.

The experiments presented in the following four chapters of this cumulative dissertation describe a series of empirical studies that used reaction time (RT) and psychophysical methods along with modeling approaches to investigate how structured representations are reflected in higher order cognitive processes.

Chapter 2 describes an experimental study that was designed to address how grouping by closure in target and distractor objects influences the configural superiority effect (CSE, Pomerantz & Portillo, 2011). We employed a visual search task with Kanizsa-type figure layouts that contained object parts or corresponding wholes. Our results replicated the typical pattern of a CSE, with detection of a configural whole being more efficient than detection of a corresponding part-level target. Moreover, we found in two experiments a much more pronounced CSE and enhanced processing of configural objects when grouping by closure was presented in distractors rather than in the target. This suggests that object integration is not per se modulated by a more efficient detectability of a grouped target, but rather that grouping operations primarily affect the inhibition of distractor configurations. Additional drift-diffusion model analyses of our data revealed that efficient distractor inhibition and closure in particular expedite the rate of evidence accumulation. Altogether, these results suggest that the configural superiority effect is governed by the inhibition of distractor configurations.

Next, Chapter 3 of this thesis presents a set of four experiments, designed to address how the global/local structure of objects affects both mechanisms of attentional selection and (implicit) short-term memory. To this end, we again employed a visual search paradigm with hierarchical Navon letters as targets and nontargets. A series of analyses were specifically dedicated to isolate critical stages in global/local object processing. Our findings depict a robust global precedence effect that manifested in overall faster RTs and shallower RT/set size functions for global, as compared to local, target search, showing that the preattentive guidance of attention is biased towards global levels. Interestingly, short-term memory as measured via intertrial priming also revealed a strong global-level bias, which could be specifically attributed to the allocation of focal attention to the target for its identification, that is, a tentative mechanism to match the selected object with a memorized target template. We further demonstrate that these results cannot be explained by differences in object size and crowding strength between global and local targets. Additional experiments addressed the stability of the global precedence effect, showing that long-term environmental contingencies (e.g., given prevalent local targets) cannot revert the global bias to effectively search for frequent local targets, suggesting that global precedence overall occurs rather automatically. Moreover, when varying global/local target prevalence throughout the experiment, attentional selection was dynamically adjusted according to the prevailing target level, but priming remained stable, indicating that attention and (short-term) memory sources of global/local processing are linked yet show dissociable underlying dynamics.

Chapter 4, then continues to describe a series of four experiments, designed to address how the hierarchical structure of objects is represented in visual working memory. We employed a change detection paradigm with hierarchical (global/local) shapes. Our findings depict a robust advantage in detecting a global-level change over local-level changes. Interestingly, when the similarity between individual objects was systematically varied at a global or a local level, performance of both global and local change detections declined only for globally similar objects, but not for locally similar objects, demonstrating that global ensemble (i.e., summary) representations influence mnemonic precision. Moreover, this global precedence effect in memory was found not to be modulated by variations of the encoding durations and mostly cannot be explained by saliency differences between global and local object levels, suggesting that the effect arises primarily during the retention phase, i.e., it is independent from stimulus processing.

Finally, Chapter 5 reports an experimental study that investigated how hierarchical object structure is represented in visual working memory (vWM), and whether these structured representations adapt to varying task demands. To this end, we developed a novel hierarchical, textured stimulus with global and local orientations, and applied it to both change-detection and continuous-report tasks. Overall, our findings revealed a reliable benefit in storing global object levels in vWM. Moreover, we found a consistent influence of task demands on these structured representations. Specifically, global precedence was reduced when the orientation change magnitudes to be detected (at both hierarchical levels) were small as compared to when they were large. Furthermore, when observers were asked to report the exact global/local orientation (requiring high fidelity at both levels), the emphasis on mnemonic precision actually engendered a reversal of global into local precedence – indicating that a local bias can manifest when detailed information is required to be retained.

In summary, this dissertation poses that structured representations govern selective attention and visual working memory, suggesting that both systems are hierarchically organized (Brady, Konkle, & Alvarez, 2011; Nie, Müller, & Conci, 2017). In this light, issues in selective attention and working memory are seen as that of attending and remembering information in hierarchically structured visual environments, maintaining relevant objects in a visual scene over an immediate duration, and then later selecting and accessing the specific object level in order to perform a given task. This dissertation combines the cognitive psychology and computational approaches to explore the rules by which hierarchical attention and memory systems operate. It entails considering the space of selection and maintenance strategies and how that space is constrained by experiments on the default state, structure, and operation principles of selective and internal attention. One such constraint is global/local object structure.

References

Asch, S. E. (1962). A problem in the theory of associations. *Psychologische Beiträge*, *6*, 553–563.

Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and representation in perception* (pp 285-317). Hillsdale, NJ: Erlbaum.

Boff, K. R., Kaufman, L., & Thomas, J. P. (Eds.). (1986). *Handbook of perception and human performance* (Vol. 2). New York: Wiley.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 4-4.

Chen, L. (2005). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.

Conci, M., Böbel, E., Matthias, E., Keller, I., Müller, H. J., & Finke, K. (2009). Preattentive surface and contour grouping in Kanizsa figures: Evidence from parietal extinction. *Neuropsychologia*, *47*(3), 726-732.

Ehrenfels, C. von. (1890). Ober Gestaltqualitaten [On Gestaltqualitaten]. *Vieneljahrschrift fuer Wissenchaftliche Philosophic*, *14*, 249-292.

Gopher, D., & Kimchi, R. (1989). Engineering psychology. *Annual Review of Psychology*, *40*, 431-455.

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

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, *112*, 24–38.

Kimchi, R. (1998). Uniform connectedness and grouping in the perceptual organization of hierarchical patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1105–1118.

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 521-535.

Kinchla, R. A. (1974). Detecting target elements in multi-element arrays: A confusability model. *Perception & Psychophysics*, *15*, 149–158.

Kinchla, R. A. (1977). The role of structural redundancy in the perception of visual targets. *Perception & Psychophysics*, *22*, 19–30.

Kinchla, R. A., Solis-Macias, V, & Hoffman, J. (1983). Attending to different levels of structure in a visual image. *Perception and Psychophysics*, *33*,1-10.

Kinchla, R. A, & Wolfe, J. M. (1979). The order of visual processing: "Top down," "bottom up" or "middle-out." *Perception and Psychophysics*, *25*, 225-231.

Koffka, K. (1963). *Principles of Gestalt psychology*. New York: Harcourt, Brace & World.

Kohler, W (1929). *Gestalt psychology*. New York: Liveright.

Kohler, W (1971). Human perception. In M. Henle (Ed. and Trans.), *The selected papers of Wolfgang Kohler* (pp. 142-167). New York: Liveright.

Kubovy, M., & Pomerantz, J. R. (Eds.). (1981). *Perceptual organization*. Hillsdale, NJ: Erlbaum.

Lindsey, P. H., & Norman, D. A. (1977). *Human information processing*. San Diego, CA: Academic Press.

Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science*, *275*(5300), 671-674.

Miller, J. and Navon, D. (2002). Global precedence and response activation: evidence from LRPs. *The Quarterly Journal of Experimental Psychology: A, Human Experimental Psychology*, *55*(1), 289–310.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.

Navon, D. (1981). The forest revisited: More on global precedence. *Psychological Research*, *43*, 1-32.

Navon, D. (2003). What does a compound letter tell the psychologist's mind? *Acta Psychologica*, *114*, 273–309.

Navon, D., & Norman, J. (1983). Does global precedence really depend on visual angle? *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 955–965.

Nie, Q.-Y., Müller, H. J., & Conci, M. (2017). Hierarchical organization in visual working memory: From global ensemble to individual object structure. *Cognition, 159*, 85-96.

Paquet, L., & Merikle, P. M. (1988). Global precedence in attended and nonattended objects. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 89-100.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 141-179). Hillsdale, NJ: Erlbaum.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General*, *112*, 516–540.

Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic Gestalts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1331–1349.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 422-435.

Shepp, B. E., & Ballesteros, S. (Eds.). (1989). *Object perception: Structure & process*. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. (1977). *Introduction to human information processing*. New York: Wiley.

Titchener, E. (1909). *Experimental psychology of the thought process*. New York: Macmillan.

Treisman, A. (1986). Properties, parts, and objects. In K.. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 35:1-70). New York: Wiley.

Uttal, W R. (1988). *On seeing forms*. Hillsdale: NJ: Erlbaum.

Wertheimer, M. (1967). Gestalt theory. In W D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 1-11). New York: Humanities Press.

Wundt, W (1874). *Grundzuge der physiologischen psychologie* [*Principles of Physiological Psychology*]. Leipzig, Germany: Engelmann.

Chapter II

Inhibition drives configural superiority of illusory Gestalt:

Combined behavioral and drift-diffusion model evidence

Abstract

Illusory Kanizsa figures demonstrate that a perceptually completed whole is *more* than the sum of its composite parts. In the current study, we explored part/whole relationships in object completion using the configural superiority effect (CSE) with illusory figures (Pomerantz & Portillo, 2011). In particular, we investigated to which extent the CSE is modulated by closure in target and distractor configurations. Our results demonstrated a typical CSE, with detection of a configural whole being more efficient than the detection of a corresponding part-level target. Moreover, the CSE was more pronounced when grouped objects were presented in distractors rather than in the target. A follow-up experiment systematically manipulated closure in whole target or, respectively, distractor configurations. The results revealed the effect of closure to be again stronger in distractor, rather than in target configurations, suggesting that closure primarily affects the inhibition of distractors, and to a lesser extent the selection of the target. In addition, a drift-diffusion model analysis of our data revealed that efficient distractor inhibition expedites the rate of evidence accumulation, with closure in distractors particularly speeding the drift towards the decision boundary. In sum, our findings demonstrate that the CSE in Kanizsa figures derives primarily from the inhibition of closed distractor objects, rather than being driven by a conspicuous target configuration. Altogether, these results support a fundamental role of inhibition in driving configural superiority effects in visual search.

Introduction

Whilst part of what we perceive comes through our sense from the object before us, another part (and it may be the larger part) always comes out of our own head.

-- William James (1890, p. 103)

Understanding how the retinal images of our complex visual world are translated into integrated and coherent object representations was recognized as a central challenge by Gestalt theory (Wertheimer, 1912). A major question in this context is how the visual system combines fragments into wholes despite adverse luminance gradients and partial occlusions of the underlying scene structure. Solving this problem, by means of perceptual organization, is a fundamental function of the visual system. A number of 'laws' have been proposed describing the organizational ('grouping') principles based on which the visual system structures our environment, including grouping based on proximity, closure, and symmetry (Wagemans, Elder, et al., 2012a).

Empirical research has shown that the laws of grouping as described initially on the basis of phenomenological observations are essential for object recognition (Lowe, 1987). For example, parsing retinal images through mechanisms of perceptual organization may result in ordered scene representations where fragments are assigned unambiguously to a given object and each object can be segregated from elements that belong to other objects and the background. Such structured representations are achieved even when distinctive and continuous borders between objects are lacking. For instance, 'Kanizsa figures', such as the Kanisza square depicted in Figure 1C, demonstrate that mechanisms of visual completion can give rise to the impression of an illusory object – that is, in the example, a relatively bright central square with sharp boundaries emerges that appears to occlude the (adjacent) circular inducer elements (Kanizsa, 1955) – even though this percept has no direct physical correspondence in the retinal image (Murray & Herrmann, 2013, for a review). Original Gestalt theory claimed that closure, rather than just being a cue for grouping, is a major determinant of what constitutes a complete form (Koffka, 1935). More recently, Elder and Zucker (1993, 1994) proposed that the most important role of closure may be to relate a 1-D contour to a corresponding 2-D shape – which was supported by the finding that small changes in closure can yield large changes in shape discriminability. In this view, emergent properties of illusory figures may reflect the degree to which grouping by closure yields a global form (Wagemans et al., 2012a; Kogo, Strecha, Van Gool, & Wagemans, 2010; Kogo & Wagemans, 2013). It should be noted that in the example of the Kanizsa square, the closed shape is not part of the actual (physical) stimulus arrangement, but is rather attributed to the emergent, illusory square – that is, it actually constitutes some form of "implied closure". Figure 1 illustrates that implied closure of the emergent figure can be varied systematically by changing the configuration of the pacman inducers. Moreover, along with an increase in closure (from Figure 1A to 1C), the emergent shape exhibits a concurrent increase in the extent to which precise bounding contours are perceived based on grouping by collinearity.

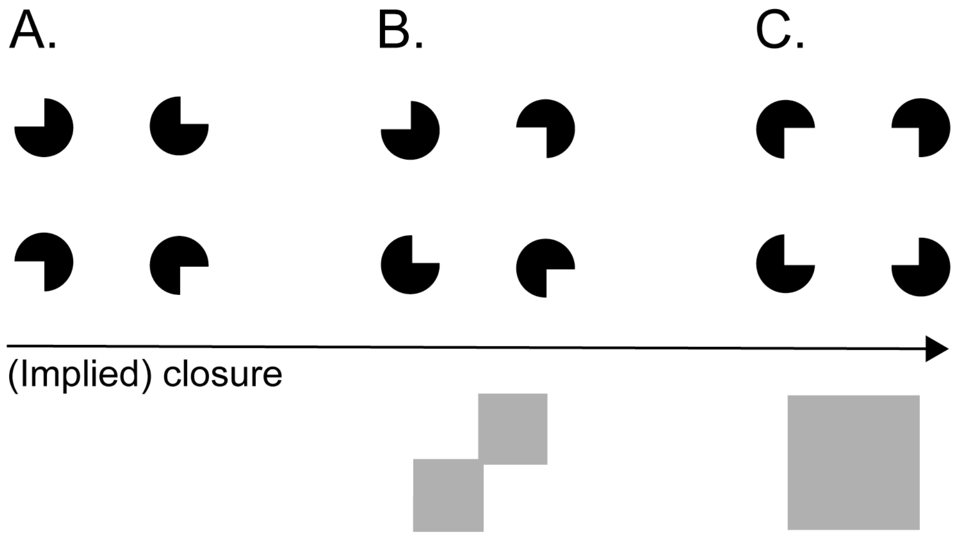

Figure 1. Implied closure in emergent shape configurations. Panels A to C illustrate that a systematic (i.e., inward facing) arrangement of pacman inducers can modify the amount of closure in the emergent (illusory) figure until a 'complete' Kanizsa square (C) is rendered. Each stimulus configuration shows an arrangement of inducers (top) together with a schematic illustration of the corresponding emergent shape representation (bottom).

Despite the complex percepts that arise from illusory figures, arguably, such integrated objects are nevertheless rendered by preattentive coding mechanisms (Davis & Driver, 1994; see also Gurnsey, Humphrey, & Kapitan, 1992). For instance, Davis and Driver (1994) used a visual search task with a Kanizsa square as the target and comparable configurations (that did not give rise to an illusory figure) as nontargets. Davis and Driver found that search for an illusory target figure could be performed 'efficiently', that is, the reaction times (RTs) taken to respond to the presence of the target were independent of the number of configurations presented in the search display (the 'display size'). Subsequent studies, by Conci and colleagues (Conci, Gramann, Müller, & Elliott, 2006; Conci, Müller, & Elliott, 2007a, 2007b; Conci, Töllner, Leszczynski, & Müller, 2011), showed efficient search for illusory figures to primarily rely on grouping by closure; that is, search efficiency, reflecting how readily focal attention is allocated to the target, was primarily determined by the degree of closure provided in the target and distractor configurations. By contrast, search efficiency was found to be unrelated to the contour information, that is, the degree to which emergent shapes are constructed on the basis of grouping by collinearity (Conci et al., 2006, 2007a, 2009; Donnelly, Humphreys, & Riddoch, 1991). Thus, converging evidence from studies that employed Kanizsa-type stimuli suggests that closed object configurations are particularly effective in guiding search at preattentive stages of processing (Conci et al., 2011; Conci et al., 2009; Stanley & Rubin, 2005).

A related paradigm designed to examine the effectiveness of the emergent properties of grouping was introduced by Pomerantz and colleagues (Eidels, Townsend, & Pomerantz, 2008; Pomerantz, 2003; Pomerantz & Portillo, 2011; Pomerantz & Pristach, 1989; Pomerantz, Sager, & Stoever, 1977; Wagemans, Feldman, et al., 2012b). Their 'Configural Superiority Effect' (CSE) typically shows that RTs to localize a target amongst distractors can be significantly faster when 'irrelevant' context parts are added to an item so as to elicit the percept of a complete figure. Figure 2A illustrates a schematic example of the odd-quadrant task typically employed to investigate the CSE. Participants are asked to determine which one of four presented elements is different (e.g., element B) from the other, homogenous distractors (e.g., element A). Then, an additional, 'task-irrelevant' context item (e.g., element C) is added to all objects, now producing novel stimulus pairs (e.g., BC and AC). While this irrelevant context C does not convey any task-relevant information per se, in certain cases, stimuli will group together to form a perceptual 'Gestalt' – with one such configuration providing salient information as to what constitutes the target, thus producing a CSE (see Figure 2B for a typical example). For such configurations, detection (and localization) of the novel, composite target becomes significantly easier (relative to the non-composite target), as evidenced by faster RTs and increased accuracy.

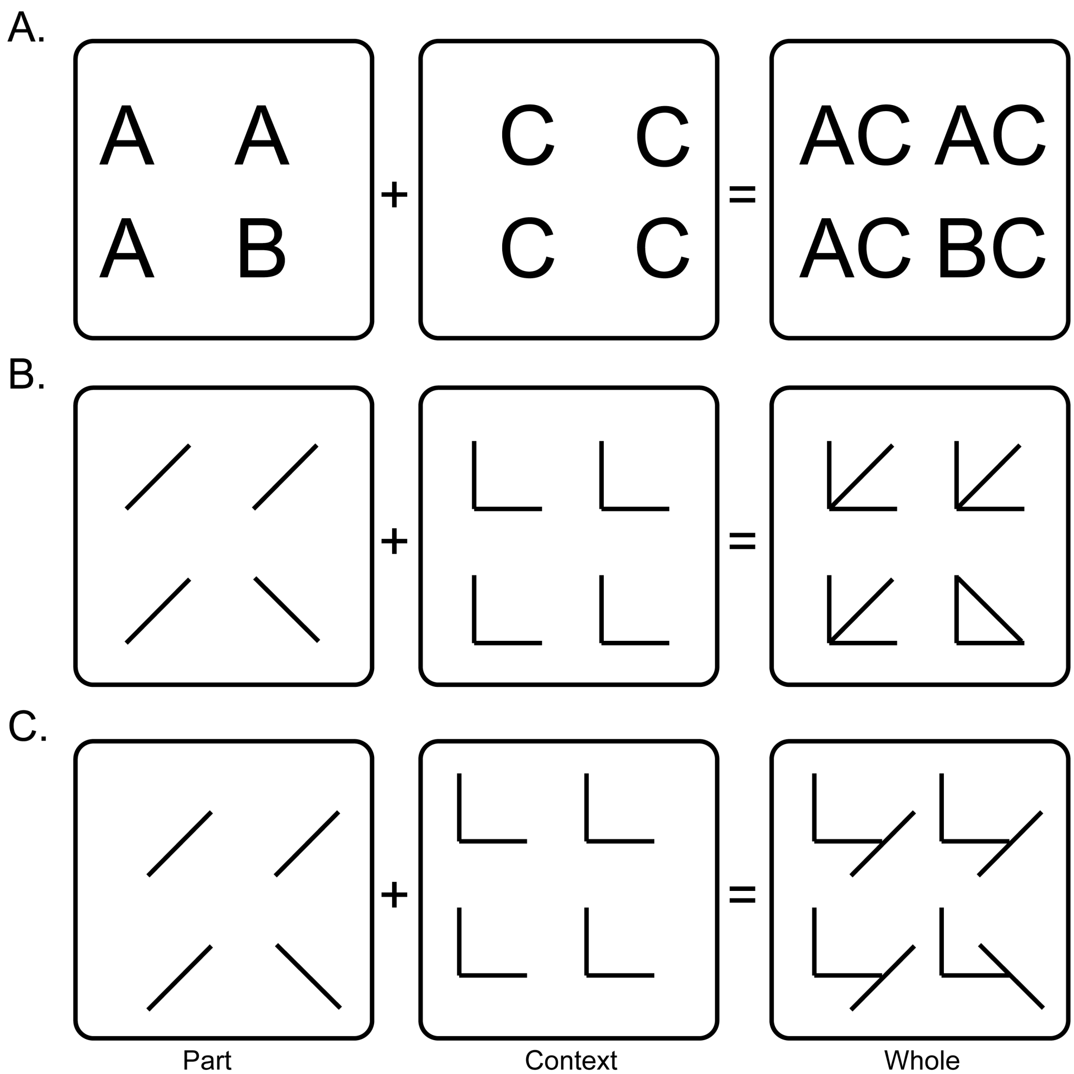

Figure 2. Panel A shows a schematic of the odd-quadrant discrimination task. Participants see either a part or a whole display (the context display is only presented for illustrative purpose). In the example, A, B, and C are placeholders for several possible stimuli. Panel B shows a corresponding example of a stimulus set that leads to a reliable configural superiority effect, whereas panel C illustrates an example that yields a configural inferiority effect (adapted from Pomerantz & Portillo, 2011).

The CSE has been used to illustrate the major role of perceptual grouping for the extraction of basic ‘Gestalts’, or emergent features (Pomerantz & Portillo, 2011). CSEs have been reported for a variety of stimulus configurations. In one prototypical case, additional pacman inducers were presented that, in this variant, combined to form a non-square whole (target) amongst Kanizsa square (distractor) configurations, relative to a part condition that presented incomplete objects consisting of only two pacman inducers (see Figure 3B). In general agreement with the findings from visual search paradigms (Davis & Driver, 1994), presentation of whole Kanizsa figures led to a reliable CSE. In many other cases, though, adding contextual information dilutes the differences between the two elements A and B, making it harder to discern the presence of the composite stimulus BC amongst stimuli AC, as compared to discerning stimulus B amongst stimuli A alone. Moreover, adding a context may also increase the total processing load, as well as increasing the chances of ‘crowding’, or observers may tend to attend to the wrong element (Pomerantz et al., 1977). This is referred to as ‘Configural Inferiority Effect’ (CIE; see Figure 2C for an example), because the composite (whole) is significantly harder to discriminate than the corresponding part elements.

Consistent with the behavioral evidence on the CSE, a recent fMRI study suggests that the ventral visual pathway, in particular the Lateral Occipital Complex (LOC), is involved in the configural processing of emergent features (Kubilius, Wagemans, & Op de Beeck, 2011). Using a localization task (see Figure 2B), this study showed that decoding of neuronal responses in LOC, but not in the primary visual cortex (V1), was better able to predict the location of the odd item when processing wholes, whereas area

V1 (but not LOC) was a better predictor of the position of the odd element when processing parts. This pattern supports the idea that Gestalt configurations may emerge at a relatively higher level of visual processing, with the processing of parts and wholes being related to distinct areas, or stages, in the visual processing hierarchy.

The aim of the present study was to further explore the crucial processes that determine the CSE. For instance, reliable CSEs have been reported for a variety of stimulus configurations, thus providing evidence for the idea that perceptual grouping generates emergent features that allow for an efficient extraction of a given target configuration. However, these studies have – to our knowledge – not investigated in detail whether the detection of a configural target is enhanced because of emergent properties of the target (thus facilitating target detection), or due to emergent features in distractors (i.e., permitting more efficient distractor suppression). On the basis of these considerations, we set out to specifically test and compare how grouping in targets and distractors modulates the CSE.

To this end, Experiment 1 employed a variant of a CSE paradigm presenting circular pacman inducer elements that potentially combine to form an illusory Kanizsa figure (i.e., Pomerantz & Portillo, 2011). The experiment consisted of two task sessions: observers were required to detect either a closed target among open nontargets (Figure 3A) or an open target among closed nontargets (Figure 3B). Importantly, the target could be presented within either a 'Part' or a 'Whole' display (Figure 3, left and right panels, respectively). Comparisons of the two possible target configurations permit us to examine whether the CSE with illusory figures can be related to grouping by closure in targets and/or distractors (Figures 3A and 3B, respectively). Next, to further disentangle grouping by closure in either targets or distractors, Experiment 2 introduced separate experimental parts that independently manipulated the degree of closure in distractors or, respectivley, in the target (while keeping the target or, respectively, the distractors constant, see Figure 7). This approach allowed examination for the separate, independent contributions of closed configurations in targets and distractors.

Moreover, while previous behavioral studies reported reliable RT effects, it is not clear at which functional level of processing the CSE emerges – that is, whether the CSE can be related to basic levels of information processing or to higher-level, decisional stages. For instance, CSE differences across conditions may reflect differences in the rate at which stimulus information is accumulated (the so-called 'drift rate'), the amount of decisional information required to provide a response (i.e., 'boundary separation'), or other nondecisional factors that influence the response, in particular initial sensory processing ('non-decision time'; see Ratcliff & McKoon, 2008). To our knowledge, there have been no attempts to model perceptual grouping by means of such a diffusion-type modeling approach. Thus, to examine such latent processing stages, we applied a model fitting procedure to the behavioral data using the Hierarchical Drift-Diffusion Model (HDDM; Wiecki, Sofer, & Frank, 2013), which incorporates an estimation of these parameters, in addition to the conventional response latency and accuracy measures.

To preview our main findings, both experiments consistently revealed that the CSE or search for a configural (Kanizsa-type) target are primarily determined by grouping by closure in distractors, but not in the target configuration. This suggests that closed shapes can be more readily rejected (as a result, the target is detected more efficiently). Our modeling results further reveal that this effect of closure in distractors is reflected in the drift rates, that is, faster rates of evidence accumulation to reach a given decision when distractor shapes are bound to a coherent (closed) object. In this view, the CSE is determined by the inhibition of to-be-rejected distractor configurations.

EXPERIMENT 1

Experiment 1 investigated object grouping, that is, grouping by closure in target and distractor configurations, using a visual search task with Kanizsa-type configurations (see Figure 1 for examples, and Pomerantz & Portillo, 2011). The target configuration could be presented either as a whole or as a part configuration (see Figure 3, left and right panels, respectively). Two conditions presented either a closed target among open distractors, or, conversely, an open target among closed distractors (Figure 3, panels A and B, respectively). Differences between targets and distractors were kept constant across wholes and parts such that a given target would always yield an identical feature contrast value relative to the distractors[1]. On the basis of previous findings, we expected faster RTs to whole as compared to part configurations, which would be indicative of a CSE (Pomerantz et al., 1977; Pomerantz & Portillo, 2011).

[1] The stimulus set used in both experiments was carefully controlled in terms of the similarity relations between target and distractors. Nevertheless, it remains possible that some subjective components of similarity (cf., Hout et al., 2016) were not captured by our control of the stimulus parameters.

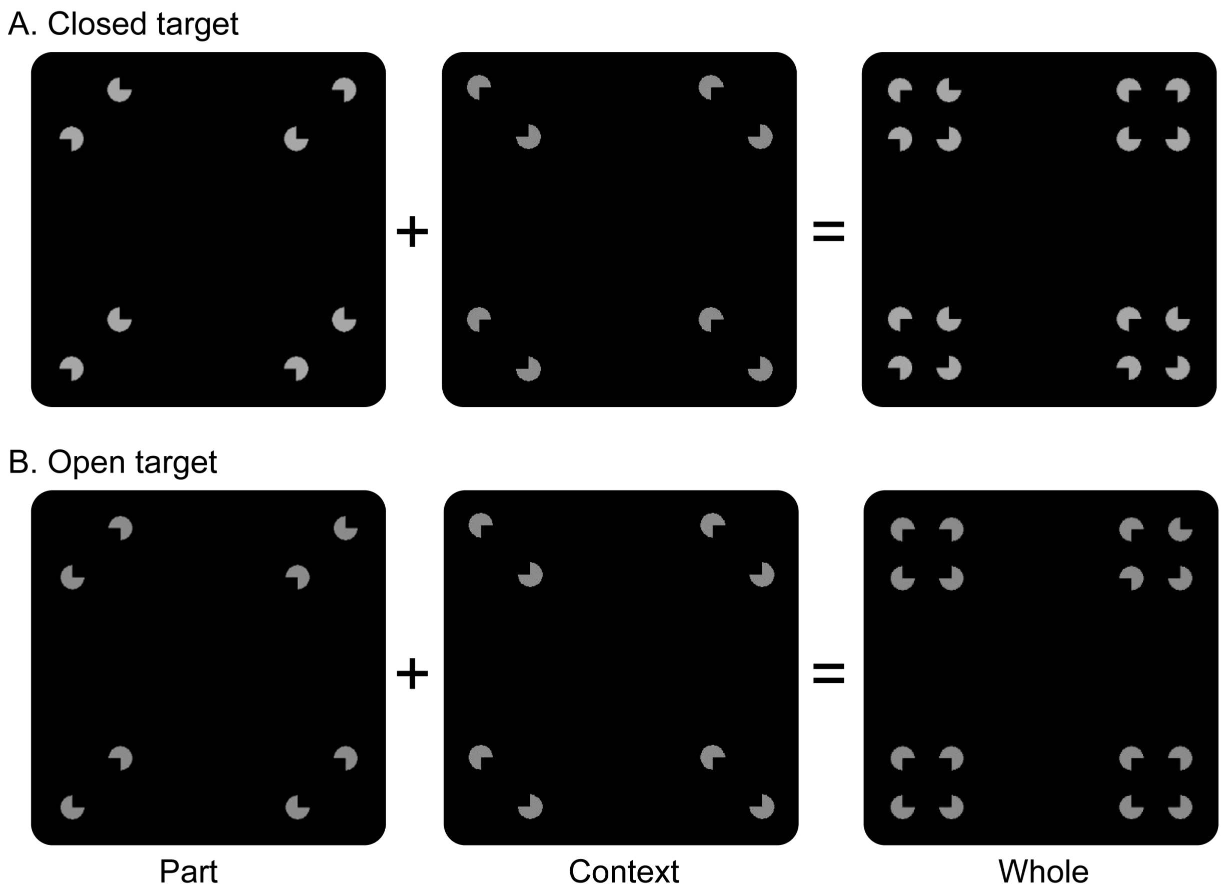

Figure 3. Example search displays in Experiment 1. (A) In the closed target condition, a closed target was presented among open distractors. (B) In the open target condition, the assignment of targets and distractors was reversed. Both open and closed targets conditions were presented either as part or as whole displays. Whole displays combined the part display with a non-informative context display, to reveal complete configurations that typically yield a configural superiority effect.

Methods

Participants. Fourteen right-handed observers (10 female; age range: 21 to 28 years; mean age: 23.6 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour. Participants provided written consent to the procedure of the experiment, which was approved by the ethics committee of the Department of Psychology at LMU München, in accordance with the Declaration of Helsinki.

Apparatus and Stimuli. The experiment was conducted on a PC-compatible computer (Dell Inc., Texas, USA) using Matlab routines and Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 22" LCD monitor screen placed at a viewing distance of approximately 57 cm. Stimuli were presented in light gray (8.5 cd/m^2) against a black (0.02 cd/m^2) background. On each trial, four configurations were placed within the quadrants of the screen, 6° from the screen center. Each configuration subtended 2.8° x 2.8° of visual angle and was composed of two or four pacman inducers, with a diameter of 0.93° each.

Configurations could be presented as 'parts' or 'wholes', presenting two or four pacman inducers, respectively. Part configurations presented two pacmen aligned along an imaginary diagonal line across the quadrant (see Figure 3, left panels). Left- or rightward tilt of the diagonal was chosen randomly for a given trial, though with each display presenting all objects in uniform orientation (i.e., for a given display all distractors were identical). Whole configurations presented four pacmen arranged in square form (see Figure 3, right panels). Configurations with pacman inducers rotated such that all indented segments faced the center of the configuration are referred to as 'closed'; by contrast, when two pacmen faced outwards, the resulting configurations are referred to as 'open'.

Within a given trial, only whole or only part configurations were displayed. Two types of trials were possible: On target-present trials, one target configuration was presented among three distractor configurations, with either a closed target among open distractors, or an open target among closed distractors. On target-absent trials, all four configurations on a given trial were identical. Figure 3 presents examples of target-present displays. The figure illustrates how whole configurations were created by combining a given part display with an additional, 'uninformative' context display.

Design. A three-factors within-subjects design was used. The independent variables were target presence, configuration, and target closure. Target presence had two levels: target present and target absent. For target-present trials, there was always one configuration that differed from the other three, whereas for target-absent trials, all four configurations were the same. Targets appeared with equal probability at the four possible display locations, with target location varying randomly across trials. The second variable, configuration, also had two levels: whole and part, denoting whether a given display consisted of stimulus arrangements made up of four or two pacman inducers, respectively (see Figure 2 and the descriptions above for further details). The third variable, target closure, again had two levels: closed and open (see Figure 3A and 3B, respectively), denoting whether a given target could be grouped to form a closed shape or not. Closed targets were presented with open distractors, and open targets with closed distractors.

Procedure. Participants were comfortably seated in a dimly lit, sound-attenuated room. The experiment was divided into two consecutive sessions that either presented closed or open targets (with order of presentation counterbalanced across observers). Each session started with 48 practice trials for participants to become familiar with the task. Then, in each session, 256 experimental trials were presented in four blocks of 64 trials each, with randomized order of the factors target presence and configuration. There were 64 trials for each factorial combination.

Each trial started with the presentation of a central fixation cross for 500 ms. Subsequently, a search display was presented until the observer's response. Participants responded with a speeded target-present versus target-absent response via mouse keys[2]. The response mapping (i.e., left/right-hand responses to target presence/absence) was counterbalanced across participants. In case of an erroneous response, feedback was provided by an alerting message (a red minus sign) that was presented for 1000 ms in the center of the screen. Each trial was separated from the next by an interval of 500 ms, presenting a blank screen.

Results

Response accuracy. Overall, performance was very accurate, with an average of 94% correct responses. Figure 4A depicts the accuracy data (percentage of correct responses), which were examined by a 2x2x2 repeated-measures analysis of variance (ANOVA) with the factors target presence (present, absent), configuration (whole, part), and target closure (closed, open). We additionally report the estimated Bayes factor (BF) for all significant results, as revealed by a comparable Bayesian ANOVA using JASP (Love et al., 2015). The Bayes factor gives the ratio with which the alternative hypothesis is favored over the null hypothesis (i.e., larger BFs argue in favor of the alternative hypothesis; see Dienes, 2011, for an overview). The accuracy ANOVA revealed the main

[2] It should be noted that CSE tasks usually employ a quadrant localization task (Pomerantz et al., 1977) whereas here we used a detection task. This slight change of the paradigm was implemented in order to apply diffusion modeling to the data (which requires two response alternatives). However, both types of task are usually highly comparable (e.g., Green 1992).

effects of both configuration (wholes vs. parts: 95% vs. 92%, $F(1,13) = 5.38$, $p = .037$, $\eta^2 = .29$, BF = 3.09) and target closure (closed vs. open: 92% vs. 95%, $F(1,13) = 18.64$, $p < .001$, $\eta^2 = .59$, BF = 3.22) to be significant. Importantly, the two-way interaction between configuration and target closure was also significant, $F(1,13) = 11.26$, $p = .005$, $\eta^2 = .46$, BF = 5.51. Post-hoc comparisons revealed a CSE in accuracy: there was a reliable difference in response accuracy only for open targets (5.5%; $t(13) = 3.8$, $p = .002$, $d = 1.01$, BF = 19.7), but not for the closed targets (-0.3%, $t(13) = -.47$, $p = .65$, $d = -0.13$, BF = 0.3). Neither the main effect of nor any interactions involving the factor target presence were significant (all ps > .4, η^2s < .05, BFs < 0.3). This pattern of results suggests that a CSE in accuracy was evident only for open targets (among closed distractors), without a comparable facilitatory effect for closed targets (among open distractors). Moreover, the CSE in accuracy was found to be independent of target presence.

Reaction times. Mean RTs for each observer were calculated excluding error responses and RTs deviating by more than three standard deviations from the mean of each condition. 7.7% of all trials, on average, were excluded by this outlier criterion (Experiment 2 yielded comparable exclusion rates). Mean RTs were again entered in a 2x2x2 repeated-measures ANOVA with the factors target presence (present, absent), configuration (whole, part), and target closure (closed, open). Figure 4B depicts the RT results. The analysis revealed both the main effect of configuration (wholes vs. parts: 1119 vs. 1325 ms, $F(1,13) = 21.82$, $p < .001$, $\eta^2 = .62$, BF = 3.84×10^{10}) and that of target closure (closed vs. open: 1371 vs. 1078 ms, $F(1,13) = 48.4$, $p < .001$, $\eta^2 = .79$, BF = 3.15×10^{4}) to be significant. Moreover, a significant interaction between target closure and configuration was again found ($F(1,13) = 13.05$, $p = .003$, $\eta^2 = .5$, BF = 20.28). This interaction was owing to a reliable CSE, of 285 ms, for open targets (presented among closed distractors), $t(13) = -7.65$, $p < .001$, $d = -2.05$, BF = 5431.4. By contrast, for closed targets (presented among open distractors), the CSE (of 108 ms) was not significant, $t(13) = -1.48$, $p = .16$, $d = -0.4$, BF = 0.66. Again, there was no main or interaction effect that involved target presence (all ps > .48, η^2s < .04, BFs < 0.16), mirroring the pattern in the accuracy data. This pattern of results shows, as above, that the CSE was particularly pronounced for closed distractors, without any substantial contribution arising from target presence and/or target closure.

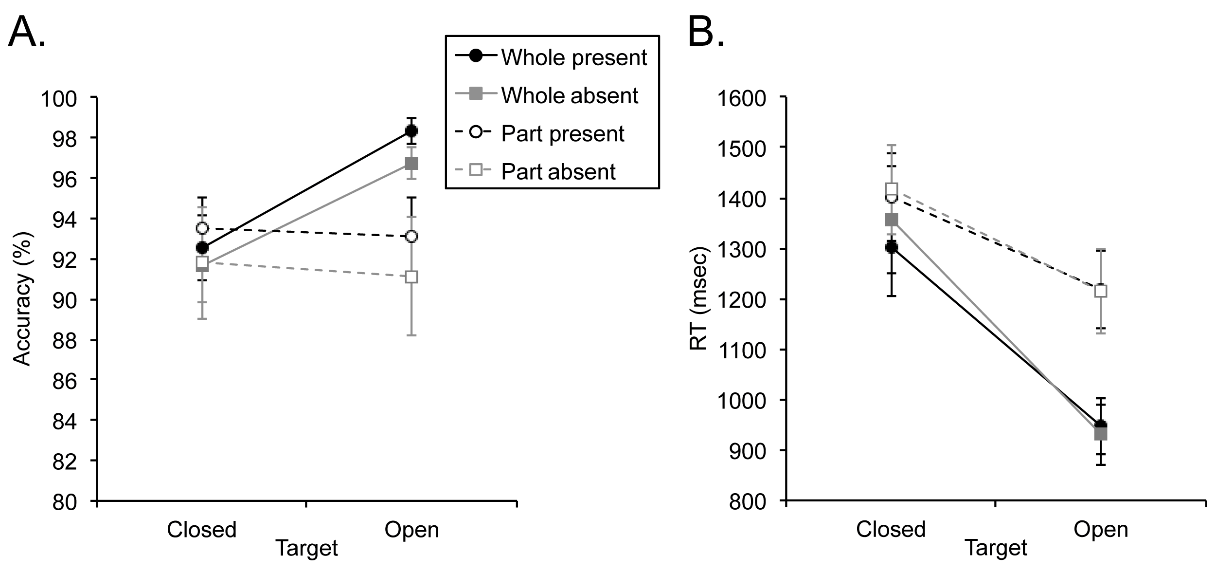

Figure 4. Mean accuracy (A) and mean RTs (B) in Experiment 1 presented as a function of target closure (closed vs. open) for the factorial combinations of configuration (whole, part) and target presence (present, absent). The error bars represent ±1 standard error of the mean.

Hierarchical Drift-Diffusion Modeling

In a subsequent step, a drift-diffusion modeling approach was applied to further demarcate task-critical stages determining the CSE. We used the Hierarchical Drift-

Diffusion Modeling approach (HDDM; Wiecki et al., 2013) to (i) apply a model fitting procedure and (ii) extract model parameters of the best-fitting model for further analysis. Theoretically, the diffusion model specifies decision processes with two possible outcomes (e.g., deciding between target presence and absence) as being inherently noisy, with information being accumulated over time. It permits the extraction of three parameters relating to, respectively: (1) information accumulation, which can be interpreted as a general measure of sensivity to the relevant configurations (the 'drift rate' parameter, v); (2) a decision threshold reflecting the amount of information required to trigger the corresponding response (the 'boundary separation' parameter, a); and (3) a mean 'non-decision' time parameter (T_{er}), which refers to the time taken by the sensory encoding of the information plus the time required for executing the motor response (Ratcliff & McKoon, 2008). It should be noted that motor responses can be assumed to reflect a constant process on all types of trials (as they are issued on every single trial); accordingly, potential differences in non-decision times could be taken to reflect exclusively the stage(s) of initial sensory processing.

HDDM constitutes a recently developed hierarchical Bayesian estimation of drift-diffusion parameters based on the RT distributions of both correct and incorrect responses, allowing for a simultaneous extraction of individual and group parameters. Fits to individual participants are constrained by the group distribution but can deviate from this distribution to a certain extent reflecting individual variability. To compare choice RTs in the CSE, eight different models were investigated, where the three parameters of interest (v, a, T_{er}) were either fixed or allowed to vary across the eight model variants (Table 1). For each model, there were 20,000 samples generated from the posterior probabilities, where the first 2000 samples were discarded. Of the remaining 18,000 samples, every fifth sample was saved, resulting in a trace of 3600 samples. The best model to describe the data across the eight conditions was selected on the basis of the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002), reflecting the best trade-off between the quality of fit and model complexity. To evaluate model performance, posterior predictives generated by the winning model were plotted on top of the observed correct and incorrect RT distributions for each participant. Figure 5 represents an example of one representative participant.

As depicted in Table 1, this model selection procedure showed the best fit when all three parameters (drift rate v, boundary separation a, nondecision time T_{er}) were allowed to vary (model 1, printed in bold), corresponding to a full drift-diffusion model. Next, each parameter of this best fitting model was then entered into a 2x2x2 repeated-measures ANOVA with the factors target presence, configuration, and target closure, as for the above analyses.

Table 1. Model Selection with HDDM in Experiment 1. A lower value of the deviance information criterion (DIC) indicates a better balance between model fit and complexity. v = drift rate; a = boundary; T_{er} = nondecision time.

Model	Free to vary	DIC
1	**v, a, T_{er}**	**7074.9**
2	v, T_{er}	7308.8
3	v, a	7422.5
4	a, T_{er}	7555.0
5	T_{er}	8009.6
6	a	8104.4
7	v	8281.5
8	Fix all	10766.3

First, analysis of the *drift rates* revealed significant main effects for configuration (wholes vs. parts: 2.15 vs. 1.71, $F(1,13) = 17.09$, $p = .001$, $\eta^2 = .57$, BF = 1.61×10^{14}) and target closure (closed vs. open: 1.61 vs. 2.27, $F(1,13) = 153.4$, $p < .001$, $\eta^2 = .92$, BF = 4.6×10^{9}). These main effects indicate that the rate of evidence accumulation was faster for wholes relative to parts, and for open relative to closed targets. There was also a configuration by target closure interaction ($F(1,13) = 82.21$, $p < .001$, $\eta^2 = .86$, BF = 2.78×10^{7}). Post-hoc paired t-tests showed the CSE to be significant only for open targets ($t(13) = 9.78$, $p < .001$, $d = 2.61$, BF = 6.42×10^{4}; whole vs. part: 2.77 vs. 1.77), but not for closed targets ($t(13) = -0.88$, $p = .4$, $d = -0.24$, BF = 0.38; whole vs. part: 1.53 vs. 1.65). As can be seen from Figure 6A, a CSE in drift rates was evident in the open, but not in the closed, target condition.

Next, a repeated-measures ANOVA of the *decision thresholds* revealed only the interaction between target presence and target closure to be significant ($F(1,13) = 5.06$, $p = .042$, $\eta^2 = .28$, BF = 0.093). Post-hoc paired t-tests showed the main effect of target closure to be marginally significant for target absent trials ($t(13) = 1.78$, $p = .099$, $d = 0.48$, BF = 0.94; open vs. closed: 2.76 vs. 2.54), but not for target present trials ($t(13) = -1.25$, $p = .23$, $d = -0.33$, BF = 0.52). Thus, open distractor configurations tended to require more evidence to be accumulated than closed configurations in order to reach the target-absent decision boundary (Figure 6B).

Finally, a repeated-measures ANOVA of the *nondecision times* yielded significant main effects of both configuration ($F(1,13) = 22.23$, $p < .001$, $\eta^2 = .63$, BF = 2.28×10^{4}) and target closure ($F(1,13) = 34.03$, $p < .001$, $\eta^2 = .72$, BF = 1.67×10^{5}). As can be seen from Figure 6C, wholes were encoded faster than the corresponding parts (494 vs. 633

ms) and the same was true for closed distractors (open targets) versus open distractors (closed targets) (closed vs. open distractors: 500 vs. 627 ms); that is, sensory encoding of stimulus information was actually more efficient with both larger amounts of physical stimulation and with closed configurations. There were no further significant effects (ps > .23, η^2s < .11, BFs < 0.6).

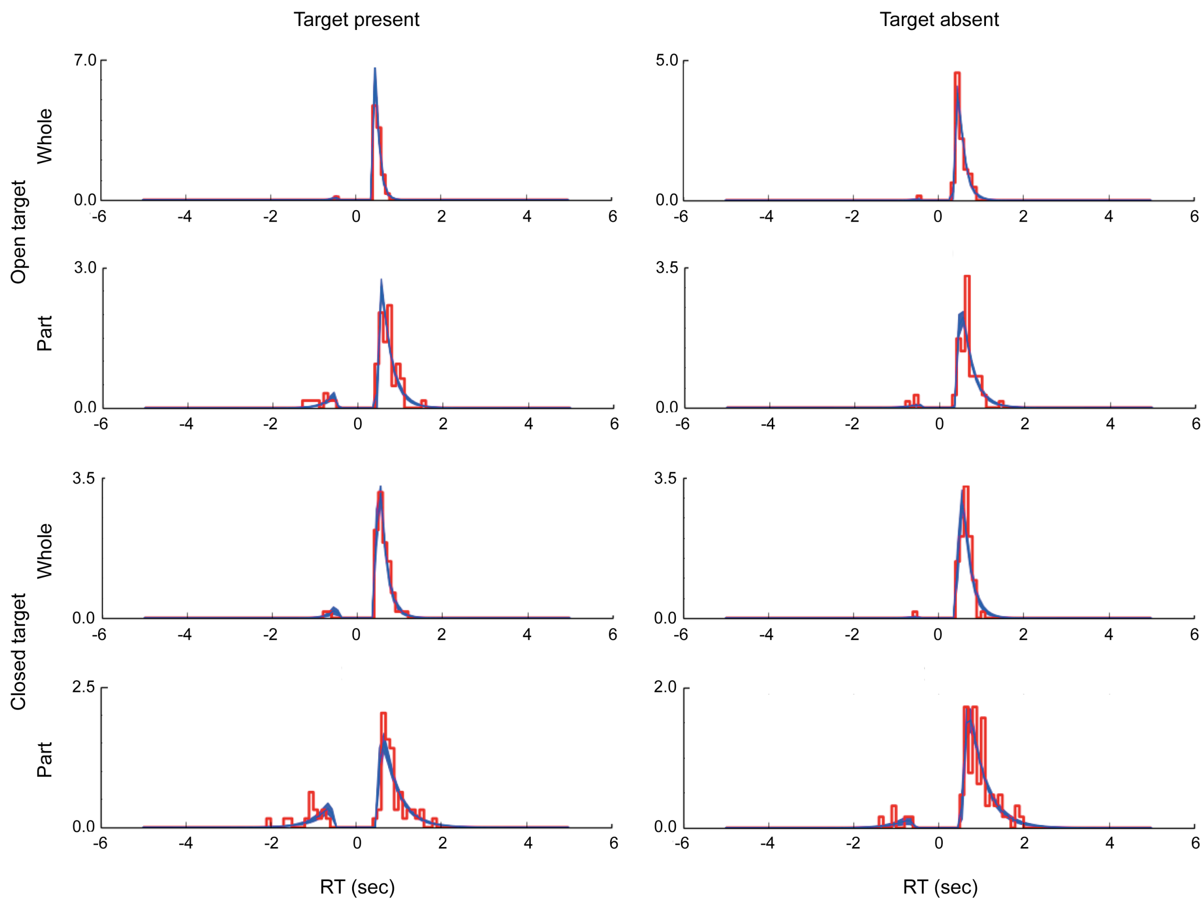

Figure 5. Examples of the posterior predictive distribution as extracted from the optimal HDDM (blue lines), and the respective empirical normalized RT distributions from one representative participant in Experiment 1 (red lines). Each panel depicts the distributions for separate conditions in the experiment. Errors have been mirrored along the x-axis to display correct and incorrect RT distributions in one plot (positive and negative values, respectively).

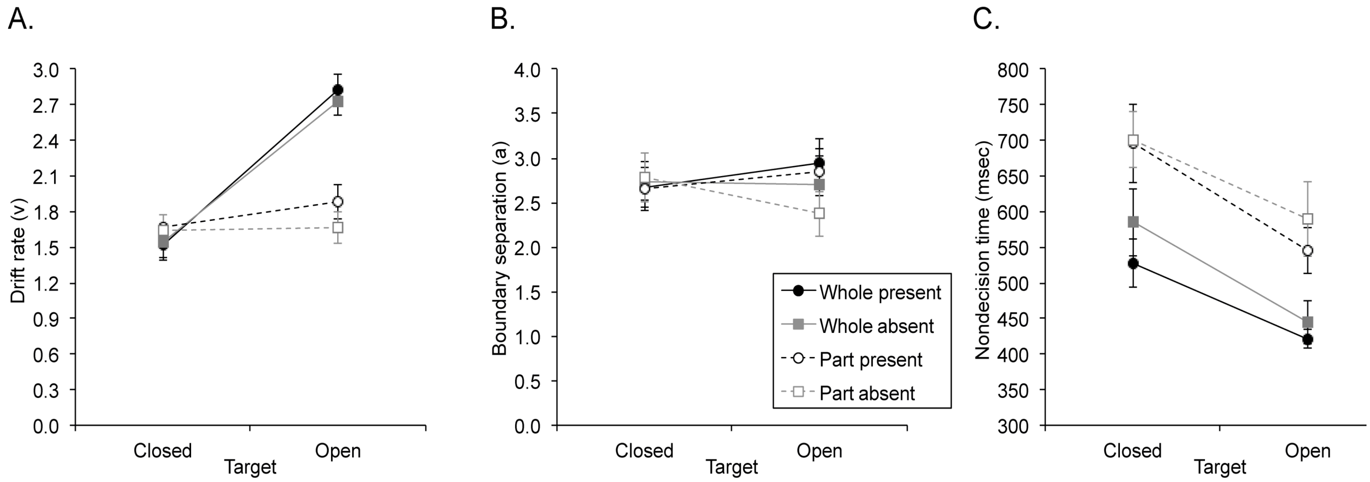

Figure 6. Mean hierarchical drift-diffusion parameters (A: drift rate; B: boundary separation; C: nondecision time) in Experiment 1. All parameters are presented as a function of target closure (closed vs. open) for the factorial combinations of configuration (whole, part) and target presence (present, absent). The error bars represent ±1 standard error of the mean.

Discussion

The results of Experiment 1 replicated previous findings of a CSE with illusory figures (Pomerantz et al., 1977; Pomerantz & Portillo, 2011). Overall, wholes were detected 180 ms faster than the corresponding parts, demonstrating that a given configuration can be processed faster than its constituent elements. Importantly, however, this overall pattern was differentially influenced by target closure: a much larger CSE manifested with an open target (presented among closed distractors), as compared to a closed target (among open distractors; CSEs of 285 [108] ms for open [closed] targets, respectively), indicating that the magnitude of the CSE is modulated by the degree of closure in distractors. Notably, the fact that a robust CSE was obtained only in the condition in which a closed Kanizsa square served as the distractor (but not when the target was a closed Kanizsa square, in which case the CSE was not reliable) would suggest that the emergence of the CSE is primarily associated with the suppression of (closed) distractors, rather than selection of a (closed) target. Moreover, target-absent trials were overall comparable to target present trials, further suggesting that closure is primarily modulating the efficient rejection of a given distractor configuration.

In addition, the drift-diffusion model analysis further identified specific processing stages associated with this CSE-related influence of distractors. The parameter estimates obtained indicate that the initial visual encoding processes, reflected by the *nondecision times*, were affected by object closure, illustrating that closed configurations were encoded more efficiently; however, they were also influenced by the amount of visual stimulation provided – as evidenced by the faster processing of wholes as compared to the corresponding parts (Figure 6C). A difference between closed and open configurations in the CSE was revealed only for subsequent processing stages reflected in the *drift rates*, with faster rates of evidence accumulation for wholes, relative to parts, with open targets [and closed distractors], as compared to closed targets [and open distractors] (Figure 6A). This pattern mirrors that of the CSE in the RT data (Figure 4B), suggesting that efficient rejection of closed distractors can expedite the accumulation of decision-critical evidence in favor of target presence. Next, the analysis of the *decision thresholds* (Figure 6B) revealed somewhat larger thresholds to reach an open (vs. closed) decision boundary on target-absent trials, but no such difference on target-present trials. In sum, the hierarchical drift diffusion modeling disclosed distinctive dynamics at different processing stages: initial stimulus encoding was expedited with both larger amounts of visual stimulation and closed configurations presented, whereas a difference that reflected the (differential) CSE in open and closed targets was manifest at the subsequent stage of evidence accumulation only. Finally, the decision threshold tended to be (marginally) higher for closed distractors (but only when there was no target).

EXPERIMENT 2

Experiment 1 revealed a CSE that was primarily related to the processing (i.e., to the rejection) of closed distractors, manifesting in terms of both expedited RTs and the speed of evidence accumulation. Experiment 2 was designed to systematically examine the independent contribution of grouping by closure to the two (related) processes of target detection and distractor rejection. To this end, in Experiment 2, we only presented complete (whole) stimulus configurations that varied with regard to the amount of closure in either targets or distractors. There were two task sessions: First, in the '*distractor rejection task*', the distractors could be closed or open configurations and the target was held constant, presenting a 'mixed' configuration that was equally similar to both types of distractors (see Figure 7A). Second, in the '*target detection task*', the target could be either a closed or an open configuration, whereas distractors were constant, always presenting a mixed configuration (see Figure 7B). Therefore, these two tasks permit us to quantify closure (closed vs. open) separately in targets and distractors, and further to differentiate its relative contributions to target detection and distractor rejection. On the basis of Experiment 1, we expected that this manipulation would engender a more robust 'closure effect' in distractors than in targets.

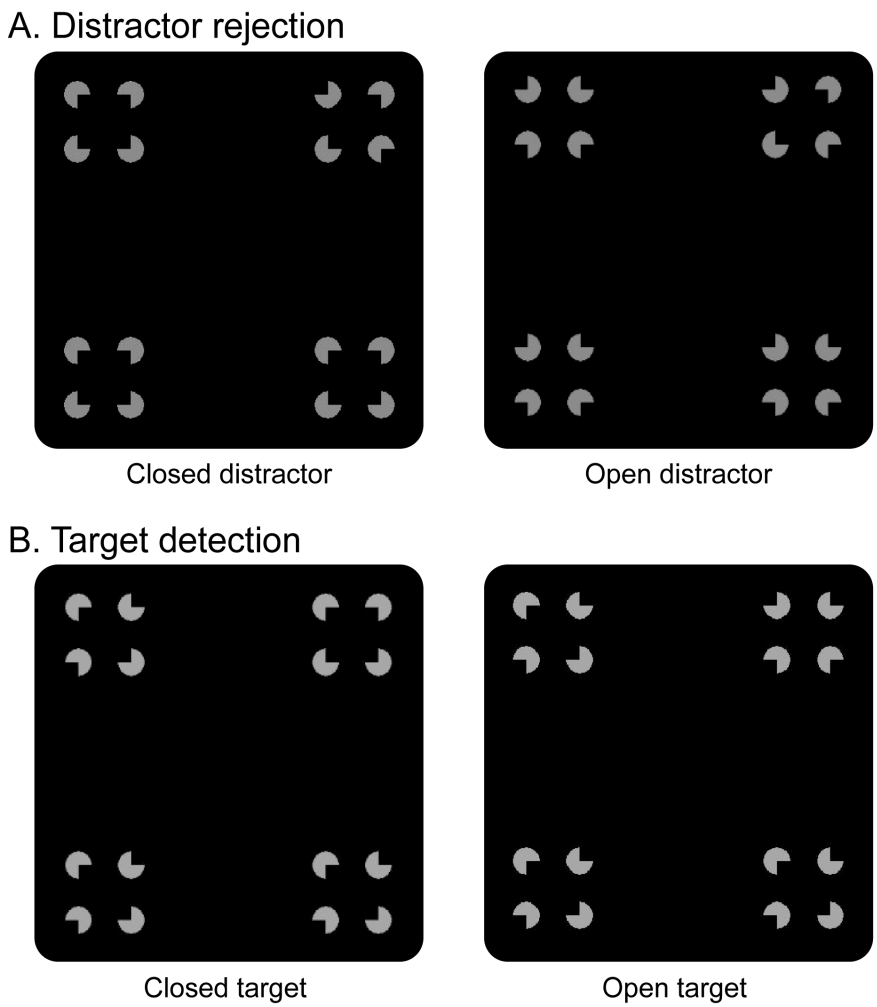

Figure 7. Example search displays in Experiment 2. (A) In the distractor rejection task, closure in distractors was varied while keeping the target constant. (B) In the target detection task, distractors were constant but the target varied in terms of grouping by closure. Note that in all possible displays, the feature contrast between a given target and distractor configuration was the same, i.e., targets and distractors differed from each other to the same extent.

Methods

Participants. Fourteen right-handed observers (10 female; age from 20 to 32; mean age = 25.6 years) with normal or corrected-to-normal visual acuity participated in Experiment 2, receiving course credits or payment of 8 Euro per hour.

Apparatus and Stimuli. All methodological details were essentially the same as in Experiment 1, except that, in Experiment 2, only whole configurations were presented – in three variants: they were (i) arranged to form a closed shape (i.e., a Kanizsa square), or (ii) depicted a mixed arrangement (with two diagonally opposing pacmen facing inwards and the other two pacmen facing outwards), or (iii) could be presented to form an open, symmetric shape (with all four pacmen oriented outwards). Figure 1 presents examples of the closed, mixed, and open configurations (see also Figure 7 for example displays). As in Experiment 1, all distractors in a given search display were identical, homogeneous shapes. Note that the open configuration as previously used in Experiment 1 is now, in this variant of the task, referred to as 'mixed' configuration.

Design and Procedure. As in Experiment 1, the task in Experiment 2 was to detect a target that differed from the other configurations, and to respond with a speeded target-present or -absent response (with response mappings counterbalanced across observers).

The experiment consisted of two different halves, presented to observers in counterbalanced order: In one half of the experiment, the composition of the distractors was varied and the target remained constant throughout – so as to test the efficiency of rejecting closed or open distractors. Thus, in this part of the experiment, observers were required to detect a 'mixed' target among (variably across trials) either 'open' or 'closed' distractors. In the second half of the experiment, in turn, the target was varied and the distractors remained constant – to test the efficiency of detecting closed or open targets. This part of the experiment always presented 'mixed' configurations as distractors and observers were required to either detect an 'open' or a 'closed' target configuration. Figure 7 presents examples of closed and open target-present trials for variations of both distractors (Figure 7A) and targets (Figure 7B). In each part of the experiment, target-present/-absent and closed/open configurations were presented in random order across trials. Targets were randomly assigned to one of the four display quadrants. There were 64 trials for each factorial combination. Each half of the experiment started with one practice block of 48 trials and was followed by 4 experimental blocks of 64 trials each.

Results

In order to directly compare the effects of grouping by closure in the processing of distractors and targets, the accuracies (and RTs) of target-present closed distractor conditions were subtracted from those in the corresponding open distractor conditions for each participant, thus providing a measure of the 'closure effect' in distractors (i.e., the benefit in accuracy and RTs for closed relative to open distractors). The same subtraction procedure was also applied to the closed and open target conditions. For statistical analysis, closure effects in targets and distractors were compared in a series of paired t-tests. Additional one-sample t-tests were employed to further investigate whether the obtained closure effects differed significantly from zero. Additional analyses of the mean target-present and -absent RTs and response accuracies (i.e., without applying this subtraction procedure) are presented in an Appendix.

<u>Response accuracy</u>. A paired-sample t-test on the closure effect in the percentage of correct responses between the experimental halves related to distractor rejection and target detection, respectively, revealed no significant difference (6.7% vs. 5.8%, respectively; $t(13) = -0.21$, $p = .84$, $d = -0.06$, $BF = 0.28$; see Figure 8A). Moreover, only the closure effect in distractors, but not that in targets, was significantly smaller than zero

($t(13) = 2.74$, $p = .009$, $d = 0.73$, BF = 7.28, and $t(13) = 1.53$, $p = .075$, $d = 0.41$, BF = 1.27, respectively), suggesting more accurate responses in rejecting closed than open distractor configurations, which is consistent with the pattern of the CSE in accuracy as obtained in Experiment 1.

Reaction times. The same analysis procedure for the closure effect as above was applied. This analysis showed that the closure effect in distractors was significantly larger than that in targets (382 vs. 149 ms; $t(13) = 2.63$, $p = .02$, $d = 0.7$, BF = 3.13; see Figure 8B), though the effects were significantly larger than zero in both cases ($ts(13) > 2.33$, $ps < .02$, $ds < 0.62$, BFs > 3.92). This indicates that closure facilitated both the detection of a (closed) target and the rejection of (closed) distractors, with closure in distractors yielding larger benefits for search performance – a finding again consistent with the results obtained in Experiment 1.

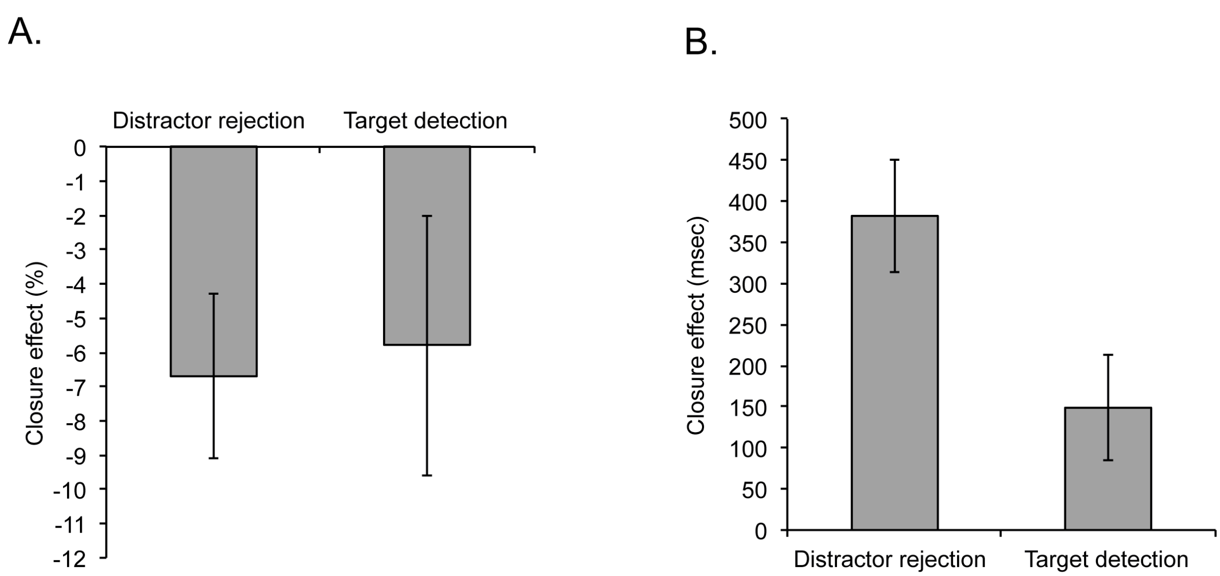

Figure 8. Behavioral results from Experiment 2. (A) Mean Closure effect in Accuracy (mean accuracies for open minus closed configurations), and (B) mean RT Closure effect

(mean RTs for open minus closed configurations) for variations of the distractors and the target, respectively. The error bars represent ±1 standard error of the mean.

Hierarchical Drift-Diffusion Modeling

As in Experiment 1, the HDDM modeling was applied to the data in order to identify the effect-critical stages of processing. The initial model-fitting procedure again supported a model variant where the three parameters (drift rate v, decision threshold a, nondecision time T_{er}) were all allowed to vary across conditions (see Table 2, model 1, printed in bold), thus, optimally predicting the observed RTs. Figure 9 represents an example model fit for one representative participant. As for the RTs, to assess the magnitude of the closure effect, open minus closed distractor conditions (difference) scores for the various parameters as estimated by the best-fitting models were examined by statistical analyses.

Table 2. Model Selection with HDDM in Experiment 2.

		DIC	
Model	Free to Vary	Distractor rejection	Target detection
1	**v, a, T_{er}**	**3581.7**	**6447.9**
2	v, T_{er}	3628.7	6628.2
3	v, a	3730.9	6632.3
4	a, T_{er}	3908.7	6673.6
5	T_{er}	4083.0	6782.6
6	a	4073.8	7124.1
7	v	3895.2	7536.8
8	Fix all	4772.7	8177.6

First, the closure effect on the *drift rates* was computed. Note that for drift rates, more negative values correspond to a benefit for the closed configuration (whereas positive values would denote a cost), that is, the polarity of the effect is reversed for this diffusion parameter (relative to the pattern in RTs). A comparison of the drift rates between distractor and target processing revealed a significant difference (-1.03 vs. -0.41; $t(13) = -2.26$, $p = .04$, $d = -0.6$, $BF = 1.81$; see Figure 10A), revealing a benefit of closure in the rate of evidence accumulation, which was particularly strong for distractor-related processing as compared to a weaker effect for target-related processing. In addition, both distractor- and target-related closure effects in drift rates showed a (marginally) significant difference from zero (distractor: $t(13) = -4.9$, $p < .001$, $d = -1.31$, $BF = 224.5$; target: $t(13) = -1.66$, $p = .06$, $d = -0.44$, $BF = 1.52$), indicating that the speed of evidence accumulation was overall faster for closed than for open configurations.

Next, the analysis of the closure effect on the *decision thresholds* revealed no significant results (all $ps > .2$, $ds < 0.22$, $BFs < 0.36$; Figure 10B). This pattern indicates that the amount of decisional information required for rejecting closed distractors was comparable to that for rejecting open distractors.

Finally, the closure effect on *nondecision times* showed no significant difference between distractor- and target-related processing (132 vs. 149 ms, respectively; $t(13) = -0.32$, $p = .75$, $d = -0.09$, $BF = 0.28$; see Figure 10C), suggesting that the duration of stimulus encoding was equivalent for comparisons of closure in distractor and target configurations. Both distractor- and target-related closure effects in non-decision times were significantly larger than zero ($ts(13) > 3.38$, $ps < .002$, $ds > 0.9$, $BFs > 20.4$), indicating that stimulus encoding of closed objects was more efficient than that of open configurations, irrespective of whether targets or distractors were varied.

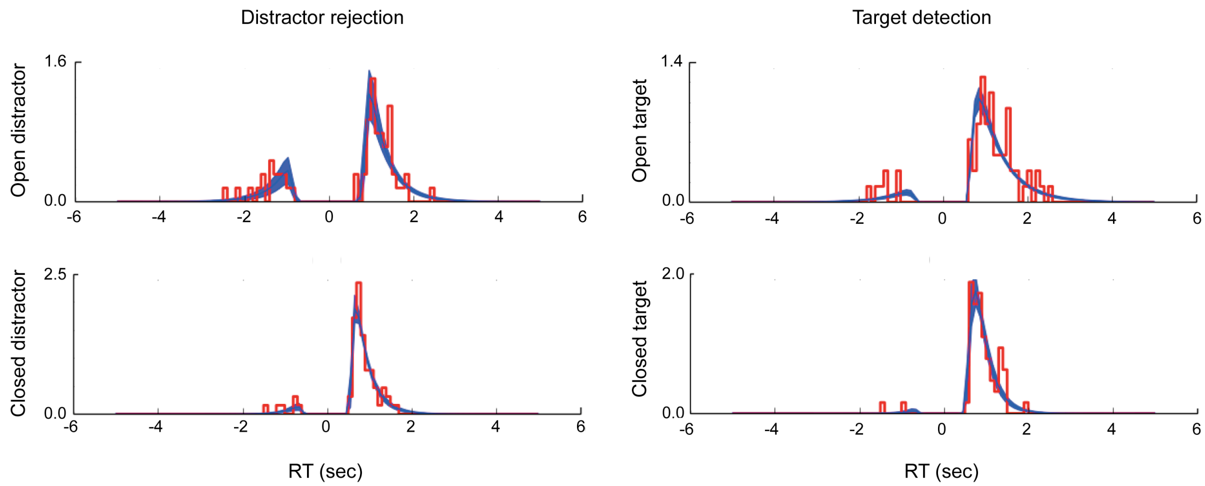

Figure 9. Examples of the posterior predictive distribution as extracted from the optimal HDDM (blue lines), and respective empirical normalized RT distributions from one representative participant in Experiment 2 (red lines). Each panel depicts the distributions for separate (target-present) conditions in the experiment. Errors have been mirrored along the x-axis to display correct and incorrect RT distributions in one plot (positive and negative values, respectively).

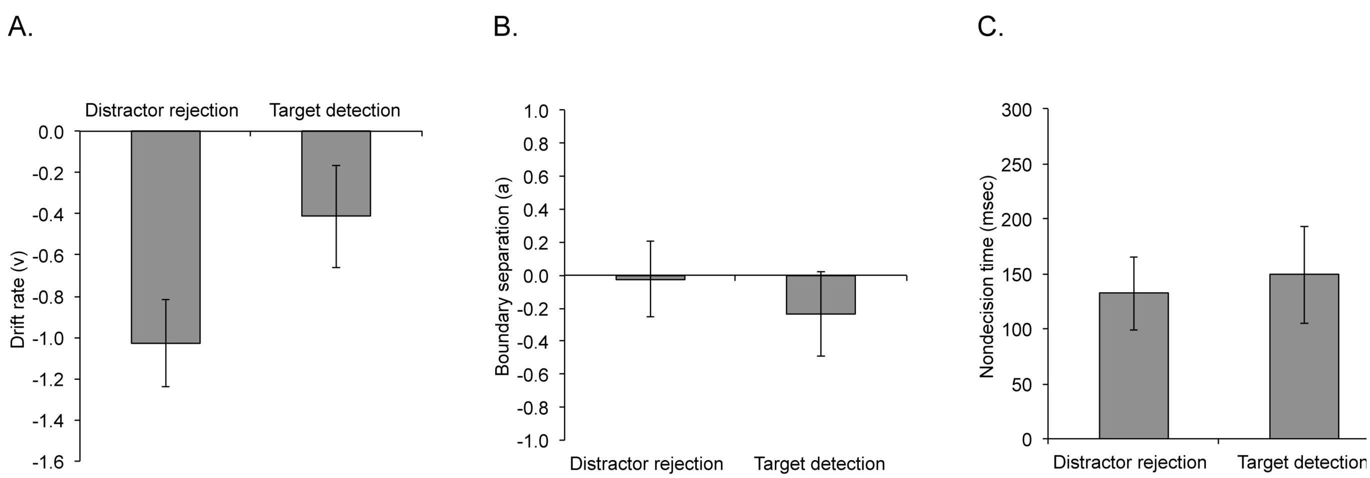

Figure 10. Closure effect (open configurations minus closed configurations) for the mean hierarchical drift-diffusion parameters (A: drift rate; B: boundary separation; C: nondecision time) in Experiment 2 for the distractor rejection and the target detection task (comparing the respective target-present trials). The error bars represent ±1 standard error of the mean.

Discussion

Experiment 2 revealed a more robust closure effect for distractors as compared to targets, replicating the general pattern of effects observed in Experiment 1. However, closure nevertheless also influenced the efficiency of target detection, albeit to a smaller extent. Overall, participants were more efficient both in rejecting closed distractors and in detecting the closed target, possibly because the Kanizsa square combines both closure and symmetry, while the open configuration is only symmetric but lacks closure and is, thus, less conspicuous.

The results of the hierarchical drift-diffusion modeling further demonstrated the underlying sources of the observed RT effects. Differences in closure between targets and distractors were not reflected in nondecision times (Figure 10C) or in the boundary separation parameter (Figure 10B). Closed configurations afforded overall more efficient sensory encoding than open configurations, but there was no difference when comparing target- and distractor-related processing. Moreover, decision thresholds were comparable across open and closed configurations. Only the drift rates (Figure 10A) showed a differential effect between distractor- and target-related processing, thus mirroring the RT pattern (Figure 8D). This means that evidence accumulation was faster to reach the decision boundary for closed distractors than for open distractors. The same pattern was also observed when comparing closed and open targets, but the respective differences were substantially smaller.

General Discussion

The current study aimed at elucidating how visual grouping by closure in targets and distractors contributes to the emergence of an 'illusory Gestalt'. To this end, two experiments were conducted employing a visual search task that presented variants of Kanizsa figures (Kanizsa, 1955), either inducing a 'part' or a 'whole' configuration with variations in grouping by closure. In Experiment 1, we found a robust CSE, that is, overall faster responses (by 180 ms) to wholes as compared to parts. Moreover, configural superiority was modulated by closure: detection of open targets (among closed distractors) showed a larger CSE than detection of closed targets (among open distractors; mean CSEs of 285 and 108 ms, respectively), with results being comparable for target-present and -absent trials. A diffusion model analysis on these data indicated that the observed CSE emerged at the stage of evidence accumulation. That is, a difference between closed and open configurations was revealed in the drift rate parameter, with faster evidence accumulation for wholes relative to parts with open targets (closed distractors), as compared to closed targets (open distractors). This pattern shows that the CSE in Experiment 1 primarily derived from processes related to the extraction of information to reach a decision. This process of information accumulation in turn seems to be particularly related to the suppression of closed, that is, well-grouped (distractor) configurations.

Next, in Experiment 2, we further investigated the role of grouping by closure, now systematically varying closure in targets and distractors independently of each other (using displays with whole-configurations only). Our analyses were primarily devised to compare the effect of closure in both targets and distractors, with closure quantified by subtracting search RTs for closed from RTs for open configurations. The results revealed a more robust effect of closure in distractor configurations as compared to targets (382 and 149 ms, respectively). Moreover, the enhanced closure effect in distractors was again reflected in the speed of evidence accumulation (the drift rate parameter). This analysis indicates that participants were overall faster to accumulate evidence for closed as compared to open configurations, but this benefit of closure was particularly pronounced with closure of distractor configurations, as compared to a much smaller effect with closure in the target configuration.

Taken together, the current results significantly extend previous studies on the CSE (Pomerantz et al., 1977; Pomerantz & Portillo, 2011) by showing that detection of a target configuration is facilitated primarily by the successful inhibition of distractors, with a considerably smaller role for target-related processing. While configural target processing may modulate search performance (Conci et al., 2011), in fact, we found no evidence of a reliable contribution of the target configuration to the CSE in Experiment 1. This suggests that configural superiority is not related to the emergence of an integrated object that matches a target description, or 'template', held in visual short-term memory. Such target templates are thought to have a privileged status, top-down biasing visual coding processes towards target-defining features (Olivers, Peters, Houtkamp, & Roelfsema, 2011). However, the current experiment yielded little evidence that the template status of the target is enhanced by object closure. Rather, the effect of the grouped configuration was particularly related to the distractors, suggesting that grouping by closure permitted more efficient suppression of task-irrelevant distractor configurations. One reason for the stronger effect of closure in distractors than in the target might simply derive from the fact that, in the typical CSE paradigm, there are more (most often three) distractors as compared to only a single target. In fact, visual search experiments show that, as set size increases, grouped distractors usually bring about a strong modulation of search efficiency (Conci et al., 2007a, 2007b), suggesting that the benefit of grouping in distractors increases as the number of candidate target configurations becomes larger (see also below and Humphreys & Müller, 1993). In this view, the 'emergence' of a configural target thus appears to be a by-product of the efficient suppression of a grouped array of distractors.

The critical stage that determined the observed pattern of the CSE was related to processes of evidence accumulation (as evidenced by the modulation of the drift-rate parameter), with closure in distractors speeding the rate of evidence accumulation. We propose that the emergence of a configural target from its constituent parts derives from the inhibition of distractor configurations. From this perspective, changes in the drift rate parameter are attributable to attentional control settings engaged in the inhibition of task-irrelevant objects, which are especially sensitive to the 'objecthood' (brought about by grouping mechanisms) in distractor arrangements (Kimchi, Yeshurun, & Cohen-Savransky, 2007).

The CSE has primarily been explained in terms of the Theory of Basic Gestalts (Pomerantz & Portillo, 2011), assuming a major role of perceptual grouping for the extraction of basic 'Gestalts', or emergent features, and treating such completed objects as the building blocks for perceptual organization. At the core of the theory is the formation of a Gestalt in a given object configuration, which permits faster and more efficient search for emergent features (that arise from the combination of parts into wholes on the basis of grouping) as compared to the corresponding basic features (i.e., properties of the parts, such as line orientation or color). Classical models of visual search (Treisman & Gelade, 1980; Wolfe, 2007) can usually not account for the CSE, and adding a uniform, non-informative context to search items would not normally be expected to improve performance (but rather only increase processing load). Context is usually added to all search items, but the relative contribution of targets and distractors in the build-up of emergent features has, to the best of our knowledge, not been investigated. In this regard, the current experiments reveal a preferential contribution of Gestalt formation to the CSE in visual search, which arises foremost from the distractors and only to a lesser extent from target-related processing.

Recent evidence suggests that a dedicated brain region in the ventral visual pathway, the LOC, may be particularly related to the processing of configurations, that is, emergent features (Kubilius et al., 2011). The authors showed that LOC (versus V1) was better able to predict the processing of wholes, whereas area V1 (versus LOC) better predicted the processing of part configurations. This pattern, showing processing of parts and wholes in distinct areas of the visual processing hierarchy, supports the idea that Gestalts may emerge only at a relatively high level of visual processing (beyond V1). Note that LOC has also been implicated in the processing of objects in general (Grill-Spector, Kourtzi, & Kanwisher, 2001) and illusory figures in particular (e.g., Bakar, Liu, Conci, Elliott, & Ioannides, 2008), for various tasks. In the light of our findings, the differences in the neuronal responses in LOC, as revealed by Kubilius et al. (2011), would appear to reflect the processing (in particular: the suppression) of distractor wholes, rather than the emergence of a configural target, thus resulting in a behavioral CSE.

In line with studies on the CSE, emergent features in illusory figures have been reported to yield efficient visual search performance (Davis & Driver, 1994; Gurnsey et al., 1992). Allocation of attention in search for Kanizsa-type figures is promoted, in particular, by grouping based on closure, that is, rendering a complete-object representation of the whole figure (Conci et al., 2006, 2007a, 2007b) – where implied closure is implicated in extracting a crude 'salient region' that can effectively guide search (for converging behavioral and electrophysiological evidence, see Conci et al., 2006, 2011; Wiegand et al., 2015). In this regard, the current findings suggest that the CSE for illusory figures is primarily related to distractor inhibition rather than target facilitation. Consistent with the present findings, a recent event-related potential (ERP) study has shown that search for a target Kanizsa figure can integrate information about distractors to optimize target selection (Töllner, Conci, & Müller, 2015) – suggesting that some form of distractor template drives top-down (distractor) suppression, thus reducing the distractors' impact on selection. In this view, the (relatively) efficient detection of a search target would be facilitated by the template-based rejection of grouped distractors (Duncan & Humphreys, 1989; see also Humphreys & Müller, 1993, for a computational model of template-based inhibition of distractors). For instance, Humphreys and Müller's model assumes that items (distractors, the target) compete to activate their respective templates, and in this competitive process, similar items (i.e., distractors of which there are multiple instances in the display) have a competitive advantage, that is, their template unit tends to cross the threshold first – upon which the whole set of distractors are 'rejected'. This is an essential component of the model and it might well account for the importance of closure in distractors, if one assumes that closed objects have an advantage in activating the respective template. Thus, these results and theoretical models are in accordance with the present findings, which provided evidence for the inhibition of closed distractor configurations, rather than facilitation of the corresponding targets, being the driving force of the behavioral CSE.

Distractor inhibition may not only operate at the level of grouped, configural objects, but also that of basic features. For instance, a target defined by a simple, salient feature discontinuity (e.g., line orientation) in a field of uniform distractors usually leads to ‘pop-out’ (e.g., Müller, Heller, & Ziegler, 1995). One idea is that pop-out is the result of low-level local ‘iso-feature’ suppression (e.g., Zhaoping & May, 2007), that is, inhibitory interactions among nearby detectors coding similar features, impeding the distractors’ ability to compete for selection and making the odd-ball (unsuppressed) target pop out. Recent ERP evidence suggests that such feature-based attention operates primarily via inhibition of distractor features, rather than activation of target features, at early stages of processing (Moher, Lakshmanan, Egeth, & Ewen, 2014). In this view, efficient detection of a target defined by a feature discontinuity is mediated by the suppression of uniform distractors, with potentially comparable mechanisms as described here for more complex object configurations.

Besides having a bearing on configural object processing and the CSE, the current results may also be seen as constituting a “search asymmetry” (Treisman & Souther, 1985; Treisman & Gormican, 1988; see also Wolfe, 2001). In a typical search asymmetry experiment, one of two stimuli (e.g., the letters Q and O) serves as target and the other as distractors in one condition (e.g., search for the Q amongst O’s), with the target and distractor roles reversed in the other condition (e.g., search for the O amongst Q’s). For this example, it has been shown that it is easier to find a target Q among distractor O's than finding a target O among distractor Q's (Treisman & Souther, 1985). The typical explanation for such an asymmetry is that, in the easier search condition, a distinctive feature (e.g., the stroke of the letter Q) would enable efficient search, while it is more difficult to find a target that is defined by the absence of a distinctive feature (e.g., the target O can be differentiated from the Q's as not having a stroke). The results of Experiment 1 obeys a comparable logic: We find more efficient performance when searching for an open target among closed distractors than when searching for a closed target among open distractors. However, in contrast to standard search asymmetries, this difference in performance does not arise because of a distinctive feature in the target (e.g., an emergent object that arises from grouping by closure), but rather the asymmetry results from the distinctive feature in distractors.

However, there are alternative explanations of search asymmetries in terms of distractor complexity. For instance, Rauschenberger and Yantis (2006) proposed that, in the above example (i.e., more efficient search with a Q target and O distractors than with the reverse assignment), the search asymmtery is caused not by (the presence vs. absence of) a distinctive feature in the target, but rather because O-shaped distractors are less complex stimuli than Q-shaped distractors, modulating search efficiency exclusively via distractor suppression (which is more efficient with less complex stimuli). Our findings lend support to this interpretation: less complex distractors (i.e., closed configurations) afford less effortful search than more complex distractors (i.e., open configurations).

Taken together, the present study points to a more prominent role of illusory Gestalt processing in the inhibition of distractors than previously thought, with implications for paradigms that investigate the role of inhibition in attention and awareness. For instance, in the perception of figure and ground, the assignment of a region in terms of being part of the figure or of the background determines which of the two leads to the prevailing percept – namely, the figure, while the other perceptual interpretation (of the background) is inhibited (e.g., Driver, Baylis, & Rafal, 1992; Roelfsema, 2006; Wagemans et al., 2012a). Moreover, in studies of binocular rivalry, where two incompatible stimuli are presented to each eye simultaneously, one of them will usually be temporarily suppressed in visual awareness, so as to make the other one perceived. Such interocular competition (between rivaling percepts) is solved by means of mutual inhibition enabling a single, coherent percept to emerge at any given moment in time (Kim & Blake, 2005). Thus, the present findings add to the notion that inhibition plays a major role in visual perception, in particular as regards the temporal and spatial filtering of the incoming sensory signals (Moors, Wagemans, van Ee, & de-Wit, in press; Tong, Meng, & Blake, 2011).

Conclusion

The present study reveals a major role of distractor inhibition in driving the emergence of an illusory Gestalt in Kanizsa figures. Our results show that the CSE is more pronounced when an emergent feature (e.g., as defined by closure) characterizes the search distractors rather than the target. Behavioral and drift-diffusion model evidence indicates that, in visual search, the configural superiority effect engendered by illusory figures arises primarily at the stage of evidence accumulation, where decisions are less driven by the conspicuity of the target configurations, but rather by the more effective suppression of grouped distractor configurations.

Appendix

Experiment 2 – Additional analyses on the mean RT and accuracy data

The results of Experiment 2 in the main manuscript present the "closure effect" in the [target-present] RT and accuracy data by subtracting the averages of closed configurations from the corresponding averages of the open configurations. To complement these results, this supplement presents the analyses of both the target–present and –absent conditions without applying a subtraction procedure.

<u>*Distractor rejection task*</u>

<u>Response accuracy</u>. The mean percentage of correct responses from the distractor rejection task was calculated for each observer and variable combination. A 2x2 repeated-measures ANOVA on the percentage of correct responses, with the factors target presence (present, absent) and distractor closure (closed, open), revealed only the main effect of distractor closure (closed vs. open: 96% vs. 89%, $F(1,13) = 9.94$, $p = .008$, $\eta^2 = .43$, $BF_{10} = 177.7$) to be significant (Figure S1A). Neither the main effect of target presence nor the interaction between target presence and distractor closure was significant (all $ps > .35$, $\eta^2s < .07$, $BFs < 0.38$).

<u>Reaction times</u>. An identical analysis was performed on mean RTs in the distractor rejection task (Figure S1A). The analysis again revealed only the main effect of

distractor closure (closed vs. open: 744 vs. 1098 ms, $F(1,13) = 56.2$, $p < .001$, $\eta^2 = .8$, $BF_{10} = 1.21e+10$) to be significant. The main effect of target presence and the interaction between target presence and distractor closure were also not significant (all ps > .12, η^2s < .18, BFs < 0.38), mirroring the pattern in the accuracy data.

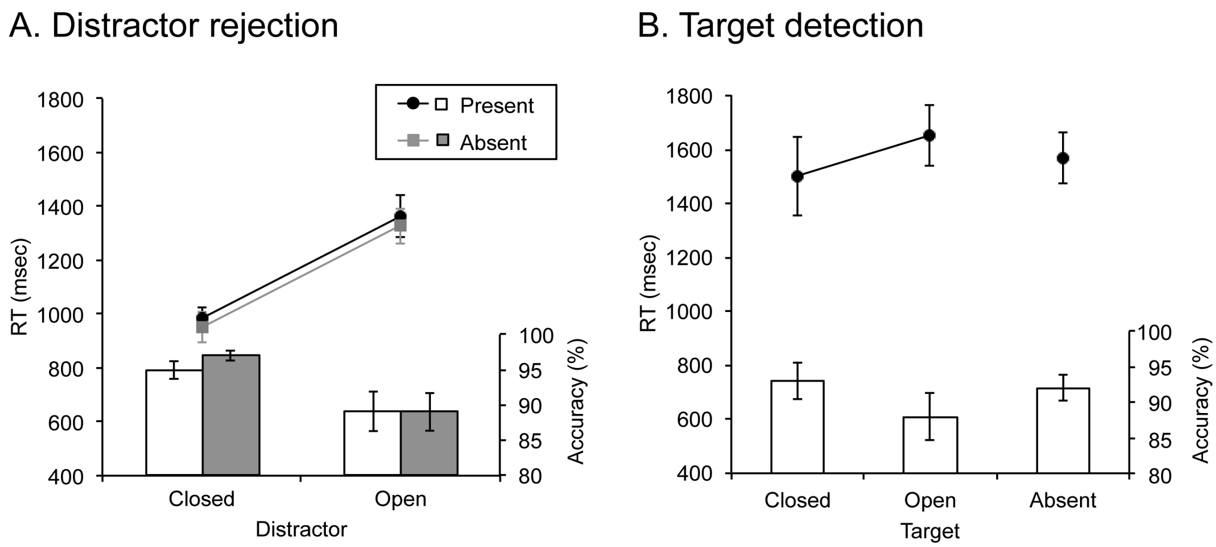

Figure S1. Behavioral results from Experiment 2. (A) Mean RTs (lines) and accuracies (bars) presented as a function of distractor closure for target-present and -absent conditions in the distractor rejection task. (B) Mean RTs (lines) and accuracies presented as a function of target closure (closed, open, absent) in the target detection task.

Target detection task

Response accuracy. Mean percentages of correct responses from the target detection task were calculated for each observer and condition. An one-way ANOVA on the percentage of correct responses, with the factor target closure (closed, open, absent) revealed no significant effect ($F(2,26) = 1.48$, $p = .25$, $\eta^2 = .43$, $BF_{10} = 0.57$; Figure S1B), and all post-hoc pairwise comparisons also showed no significant differences (all ps > .15, ds < .41, BFs < 0.7).

Reaction times. Mean RTs from the target detection task were analyzed similar to the above analysis on accuracies (Figure S1B). The analysis revealed no significant effect ($F(2,26) < 1$, $p = .39$, $\eta^2 = .07$, $BF_{10} = 0.34$). Notably, post-hoc paired t-tests nevertheless showed a significant difference between closed and open targets (closed vs. open: 1500 vs. 1649 ms, $t(13) = -2.33$, $p = .036$, $d = -0.62$, $BF_{10} = 2.01$), but no further significant differences ($ps > .41$, $ds < .23$, $BFs < 0.37$).

In sum, the pattern described here is essentially comparable to the outcomes presented in the main manuscript. Closed distractor configurations led to fewer errors and to faster responses than corresponding open distractor configurations. A comparable, but somewhat less reliable benefit was also revealed in the mean RTs for closed (as compared to open) target configurations.

References

Bakar, A. A., Liu, L., Conci, M., Elliott, M. A., & Ioannides, A. A. (2008). Visual field and task influence illusory figure responses. *Human Brain Mapping, 29*(11), 1313-1326. doi: 10.1002/hbm.20464

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision, 10*(4), 433-436. doi: 10.1163/156856897X00357

Conci, M., Böbel, E., Matthias, E., Keller, I., Müller, H. J., & Finke, K. (2009). Preattentive surface and contour grouping in Kanizsa figures: Evidence from parietal extinction. *Neuropsychologia, 47*(3), 726-732. doi: 10.1016/j.neuropsychologia.2008.11.029

Conci, M., Gramann, K., Müller, H. J., & Elliott, M. A. (2006). Electrophysiological correlates of similarity-based interference during detection of visual forms. *Journal of Cognitive Neuroscience, 18*(6), 880-888. doi: 10.1162/jocn.2006.18.6.880

Conci, M., Müller, H. J., & Elliott, M. A. (2007a). The contrasting impact of global and local object attributes on Kanizsa figure detection. *Perception & Psychophysics, 69*(8), 1278-1294. doi: 10.3758/BF03192945

Conci, M., Müller, H. J., & Elliott, M. A. (2007b). Closure of salient regions determines search for a collinear target configuration. *Perception & Psychophysics, 69*(1), 32-47. doi: 10.3758/BF03194451

Conci, M., Töllner, T., Leszczynski, M., & Müller, H. J. (2011). The time-course of global and local attentional guidance in Kanizsa-figure detection. *Neuropsychologia, 49*(9), 2456-2464. doi: 10.1016/j.neuropsychologia.2011.04.023

Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature, 371*, 791-793. doi: 10.1038/371791a0

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Psychological Science, 6*(3), 274-290. doi: 10.1177/1745691611406920

Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 561-570. doi: 10.1037/0096-1523.17.2.561

Driver, J., Baylis, G. C., & Rafal, R. D. (1992). Preserved figure-ground segregation and symmetry perception in visual neglect. *Nature, 360*(73-75). doi: 10.1038/360073a0

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433-458. doi: 10.1037/0033-295X.96.3.433

Eidels, A., Townsend, J. T., & Pomerantz, J. R. (2008). Where similarity beats redundancy: The importance of context, higher order similarity, and response assignment. *Journal of Experimental Psychology: Human Perception & Performance, 34*(6), 1441-1463. doi: 10.1037/a0012857

Elder, J., & Zucker, S. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Research, 33*(7), 981-991. doi: 10.1016/0042-6989(93)90080-G

Elder, J., & Zucker, S. W. (1994). A measure of closure. *Vision Research, 34*(24), 3361-3369. doi: 10.1016/0042-6989(94)90070-1

Green, M. (1992). Visual search: Detection, identification, and localization. *Perception 21*, 765-777. doi: 10.1068/p210765

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. doi: 10.1016/S0042-6989(01)00073-6

Gurnsey, R., Humphrey, G. K., & Kapitan, P. (1992). Parallel discrimination of subjective contours defined by offset gratings. *Perception & Psychophysics, 52*(3), 263-276. doi: 10.3758/BF03209144

Hout, M. C., Godwin, H. J., Fitzsimmons, G., Robbins, A., Menneer, T., & Goldinger, S. D. (2016). Using multidimensional scaling to quantify similarity in visual search and beyond. *Attention, Perception, & Psychophysics, 78*(1), 3-20. doi: 10.3758/s13414-015-1010-6

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology, 25*(1), 43-110. doi: 10.1006/cogp.1993.1002

James, W. (1890). *The Principles of Psychology* (Vol. II). New York: H. Holt and Company.

Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea [Quasi-perceptual margins in homogeneously stimulated fields]. *Rivista di Psicologia, 49*(1), 7–30.

Kim, C.-Y., & Blake, R. (2005). Psychophysical magic: Rendering the visible 'invisible'. *Trends in Cognitive Sciences, 9*(8), 381-388. doi: 10.1016/j.tics.2005.06.012

Kimchi, R., Yeshurun, Y., & Cohen-Savransky, A. (2007). Automatic, stimulus-driven attentional capture by objecthood. *Psychonomic Bulletin & Review, 14*(1), 166-172. doi: 10.3758/BF03194045

Koffka, K. (1935). *Principles of Gestalt psychology*. London, England: Lund Humphries.

Kogo, N., Strecha, C., Van Gool, L., & Wagemans, J. (2010). Surface construction by a 2-D differentiation–integration process: A neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures. *Psychological Review, 117*(2), 406-439. doi: 10.1037/a0019076

Kogo, N., & Wagemans, J. (2013). The emergent property of border-ownership and the perception of illusory surfaces in a dynamic hierarchical system. *Cognitive Neuroscience, 4*(1), 54-61. doi: 10.1080/17588928.2012.754750

Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2011). Emergence of perceptual Gestalts in the human visual cortex: The case of the configural-superiority effect. *Psychological Science, 22*(10), 1296-1303. doi: 10.1177/0956797611417000

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E.-J. (2015). JASP (Version 0.7). *[Computer software].*

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31*(3), 355–395. doi: 10.1016/0004-3702(87)90070-1

Moher, J., Lakshmanan, B. M., Egeth, H. E., & Ewen, J. B. (2014). Inhibition drives early feature-based attention. *Psychological Science, 25*(2), 315-324. doi: 10.1177/0956797613511257

Moors, P., Wagemans, J., van Ee, R., & de-Wit, L. (in press). No evidence for surface organization in Kanizsa configurations during continuous flash suppression. *Attention, Perception, & Psychophysics*.

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics, 57*(1), 1-17. doi: 10.3758/BF03211845

Murray, M. M., & Herrmann, C. S. (2013). Illusory contours: A window onto the neurophysiology of constructing perception. *Trends in Cognitive Sciences, 17*(9), 471-481. doi: 10.1016/j.tics.2013.07.004

Olivers, C. N., Peters, J., Houtkamp, R., Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15*(7):327-34. doi: 10.1016/j.tics.2011.05.004.

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

Pomerantz, J. R. (2003). Wholes, holes, and basic features in vision. *Trends in Cognitive Sciences, 7*(11), 471-473. doi: 10.1016/j.tics.2003.09.007

Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic Gestalts. *Journal of Experimental Psychology: Human Perception and Performance, 37*(5), 1331-1349. doi: 10.1037/a0024330

Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 635-649. doi: 10.1037/0096-1523.15.4.635

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3*(3), 422-435. doi: 10.1037/0096-1523.3.3.422

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922. doi: 10.1162/neco.2008.12-06-420

Rauschenberger, R., & Yantis, S. (2006). Perceptual encoding efficiency in visual search. *Journal of Experimental Psychology: General, 135*(1), 116-131. doi: 10.1037/0096-3445.135.1.116

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29*, 203-227. doi: 10.1146/annurev.neuro.29.051605.112939

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B, 64*(4), 583-639.

Stanley, D. A., & Rubin, N. (2005). Rapid detection of salient regions: Evidence from apparent motion. *Journal of Vision, 5*, 690-701. doi: 10.1167/5.9.4

Töllner, T., Conci, M., & Müller, H. J. (2015). Predictive distractor context facilitates attentional selection of high, but not intermediate and low, salience targets. *Human Brain Mapping, 36*(3), 935-944. doi: 10.1002/hbm.22677

Tong, F., Meng, M., & Blake, R. (2011). Neural bases of binocular rivalry. *Trends in Cognitive Sciences, 10*(11), 502-511. doi: 10.1016/j.tics.2006.09.003

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136. doi: 10.1016/0010-0285(80)90005-5

Treisman A, & Gormican S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*(1), 15-48.

Treisman. A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*(3), 285-310.

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

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012b). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218-1252. doi: 10.1037/a0029334

Wertheimer, M. (1912). Experimentelle studien über das sehen von bewegung [Experimental studies on the seeing of motion]. *Zeitschrift für Psychologie, 61*, 161-265.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics, 7*. doi: 10.3389/fninf.2013.00014

Wiegand, I., Finke, K., Töllner, T., Starman, K., Müller, H. J., & Conci, M. (2015). Age-related decline in global form suppression. *Biological Psychology, 112*, 116-124. doi: 10.1016/j.biopsycho.2015.10.006

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics, 63*(3), 381-389.

Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99-119). New York: Oxford University Press.

Zhaoping, L., & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology, 3*(4), e62. doi: 10.1371/journal.pcbi.0030062

Chapter III

Searching for the forest or trees:

Attentional zooming and level-specific memory in hierarchical objects

Abstract

Objects can be represented at multiple hierarchical levels, though typically more global levels receive precedence over more local levels. Here, we explored how object hierarchy affects the zooming of attention within and across trials using a novel visual search task with Navon letters as global/local targets and nontargets. Experiment 1 revealed search to be more efficient for targets defined at the global level versus comparable local-level targets. Moreover, a global precedence effect was also evident in cross-trial priming effects: an advantage of level repetitions (vs. changes) occurred only for global targets but not for local targets. Experiment 2 demonstrated that this differential pattern of performance across global/local object levels does not simply result from differences in object size and crowding strength. Then in Experiments 3 and 4, the prevalence of global and, respectively, local targets was manipulated to investigate the stability of the global/local processing asymmetry. When local targets were presented more frequently than global targets (i.e., on 75% of all trials), global precedence was overall reduced and priming occurred at both object levels. Furthermore, when systematically changing the prevalence of global and local targets over the whole experiment, attentional selection exhibited a dynamic adjustment according to the prevailing target level, whereas asymmetric object-level priming remained stable. Taken together, these results revealed a default (global) attentional state that is tuned to both short and long-term environmental contingencies, as global precedence reflects the flexible zooming of attention and leads to a stable global bias of object levels in short-term memory.

Introduction

Our natural environment is complex and must therefore be perceptually structured to maximize the processing efficiency of the visual system. Visual structuring is mainly accomplished by mechanisms of perceptual organization, integrating fragmentary parts to form coherent wholes. Furthermore, natural scenes and objects are normally composed of multiple layers, which can be described at multiple, hierarchical levels of perceptual organization (e.g., a forest has trees, and the trees have leaves). Thus, objects may be represented at different levels in a visual hierarchy, with global representations at the top and more local representations towards the bottom (Kimchi, 1992; Wagemans, Elder, et al., 2012). Such a hierarchical relationship between parts and wholes has also been demonstrated for a variety of composite figures (e.g., Navon, 1977; Pomerantz, Sager, & Stoever, 1977). As depicted in Figure 1, such composite figures consist of elements at a local level of representation (e.g., the letters "H", circular "pacmen", or small squares), which are combined to yield a different object at a global level (e.g., the letter "U", an illusory square, or a big triangle).

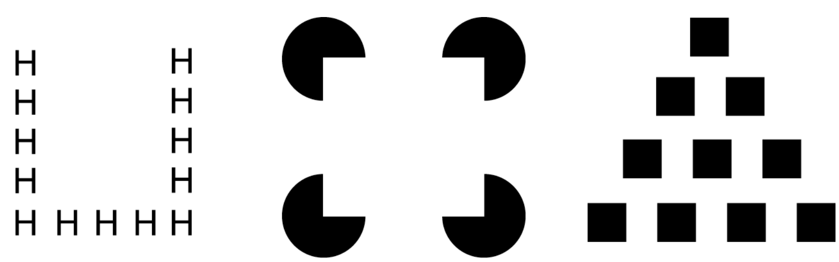

Figure 1. Examples of hierarchical stimulus configurations with global and local levels of representation: Navon letter (left), Kanizsa square (middle), hierarchical shape (right).

Typically, (more) global object levels receive precedence over (more) local levels, indicating that the zooming of attention is set in accordance with the hierarchical structure of visual objects (Navon, 1977, 1981, 2003). For instance, in Navon's basic experiment to study global/local object representations, observers were presented with hierarchical letter stimuli (Figure 1, left) and asked to identify either the global (whole) or the local (component) letters, in separate blocks of trials. These types of configuration usually reveal a global advantage: faster identification of the global than the local letters. Moreover, conflicting information across levels gives rise to a disruptive influence of task-irrelevant global information on local-level identification (global-to-local interference), but no interference of irrelevant local information with global identification (Navon, 1977). Taken together, the global advantage and the global-to-local interference support a general pattern of "global precedence" (e.g., Navon, 1977; Kimchi, 2015): global properties of a visual object are processed first, followed by the analysis of local details (Kimchi, 1992). The Reverse Hierarchy Theory (RHT) proposed by Hochstein and Ahissar (2002) suggests an explanation of global precedence from a neurophysiological perspective. RHT assumes that global information is initially extracted from a feedforward sweep of information processing by high-level cortical mechanisms. Thus, this initial percept represents the global "gist" of the scene, whereas local details become available only subsequently via recurrent feedback connections to lower-level areas (see also Roelfsema, 2006). This recurrent architecture allows flexible processing of local details by operating feedback connections specific to the attentional zooming required for a given task (Roelfsema, 2006). Within this framework, the extraction of global object properties at higher levels precedes the processing of local details.

A number of studies have demonstrated reliable global precedence effects using a variety of stimulus materials, such as traditional Navon letters, Kanizsa figures, abstract hierarchical shapes, faces, and gratings (see Figure 1 for examples, and Dale & Arnell, 2013, for a comparison of various stimulus types). Importantly, most of these paradigms have presented observers with single hierarchical objects. Accordingly, evidence for global precedence in these configurations usually reflects differences in processing between the hierarchical levels of a stimulus that is currently in the focus of attention. However, global precedence may also, at least partially, occur for non-attended objects (Paquet & Merikle, 1988) at preattentive stages of processing (Mattingley, Davis, & Driver, 1997; Conci et al., 2009). With multiple hierarchical stimulus configurations – as, for example, in visual search tasks – the question is not whether focal attention is set in accordance with the different hierarchical object levels, but whether the guidance of attention by preattentive object information is sensitive to differences between global and local representations.

Concerning this question, visual search studies have shown that detection of a global target configuration is more efficient than detection of a local arrangement of items (Conci, Müller, & Elliott, 2007a, 2007b; Conci, Töllner, Leszczynski, & Müller, 2011; Deco & Heinke, 2007; Donnelly, Humphreys, & Riddoch, 1991; Nie, Maurer, Müller, & Conci, 2016; Wagemans, Feldman, et al., 2012), while precedence in visual search is influenced by various perceptual factors, including the size, number, and density of local elements (Enns & Kingstone, 1995). Overall, attentional guidance is based primarily on global object information. For example, in a study of visual search with hierarchical Navon letters as target and nontarget stimuli, Deco and Heinke (2007) found search for a target configuration to be relatively efficient – in terms of the slopes of the functions relating reaction time (RT) to the number of display items – when the nontargets were identical to the target at the local level of representation (mean slope 16 ms/item); by contrast, search was inefficient when the nontargets were identical to the target at the global level (mean slope 26 ms/item). These findings illustrate an effect of global precedence in visual search, suggesting that information at global and local object levels differentially influences search efficiency – that is: attentional guidance. Thus, taken together, studies of hierarchical stimulus processing indicate that the prevailing global object structure is a major determinant of attentional allocation in visual scenes.

Another line of recent research has shown that selection history is also a major factor to determine attentional guidance in visual search, as evidenced by search studies that have examined intertrial "history" effects on search performance (see Kristjánsson & Campana, 2010; Krummenacher & Müller, 2012; Lamy & Kristjánsson, 2013, for reviews). For example, detection of "pop-out" targets singled out by a specific color or, respectively, shape from amongst the nontargets becomes easier on a given trial if the current target is defined by the same feature, or in the same feature dimension (Found & Müller, 1996; Maljkovic & Nakayama, 1994, 1996; Müller, Heller, & Ziegler, 1995), or if the current target appears at the same location (Maljkovic & Nakayama, 1996) as the target on the preceding trial(s) – effects that have been attributed to intertrial "priming". Moreover, such effects have also been shown to operate not only at the level of individual feature dimensions (e.g., color), but also for entire object representations (Kristjánsson, Ingvarsdöttir, & Teitsdöttir, 2008). Taken together, these results suggest that past experience influences the allocation of visual attention, reflecting a kind of top-down bias based on some form of (implicit) short-term memory that buffers successful task settings and supports predictions of likely upcoming events (Müller, Krummenacher, & Heller, 2004).

While global precedence effects indicate that global and local information are processed with asymmetric attentional priorities (with a bias towards the global level), intertrial priming across hierarchical levels of a given stimulus does not appear to be different for global and local object levels. A number of studies have examined level-repetition effects in the Navon task (Hübner, 2000; Lamb & Yund, 1996, 2000; Lamb, London, Pond, & Whitt, 1998; Robertson, 1996; Ward, 1982). These studies, in general, revealed a benefit when a target was repeatedly presented at the same level, with priming effects across levels remaining comparable even when the target configuration and the associated response changed (Filoteo, Friedrich, & Stricker, 2001; Robertson, 1996). That is, when selecting relevant information at a global or, respectively, local level of a given hierarchical stimulus, there is a performance benefit when the level repeats across trials (and, conversely, a cost when the levels switch). However, the magnitude of level repetition benefit is usually comparable for both global and local targets and independent of stimulus- and response-specific factors – suggesting that object-level priming is an automatic bias to sustain the scale of attention from one moment to the next (Robertson, 1996).

In the present study, we explored how object structure influences both attentional selection and memory-based guidance of visual search. To this end, we employed a visual search task presenting hierarchical Navon letters as target and nontarget stimuli (Figure 2A), with observers being required to detect a T-shaped target among L-shaped nontargets. Importantly, the target could be represented either at a global (Figure 2B, left panel) or at a local level (Figure 2B, right panel), with the target-defining level varying randomly across trials. The basic paradigm was initially tested in Experiment 1, where comparisons of the two possible target levels permit us to (i) examine whether the global precedence effect would be replicable in visual search, that is, in terms of search efficiency as assessed by the search RT slopes (i.e., the slopes of the functions relating search RT to set size, the number of display items); and (ii) to assess whether effects of global precedence would also manifest in the pattern of intertrial priming across global and, respectively, local target-object levels. That is, both measures may be related to aspects of attentional guidance (search efficiency) and implicit short-term memory (priming) within one-and-the-same task. Experiment 2 was then performed to investigate whether size differences between stimuli at global and local levels would suffice to explain the current pattern of results. In addition, to further elucidate the stability of global/local object processing in search, Experiments 3 and 4 investigated how global precedence varies with a systematic manipulation of "target prevalence" (e.g., Wolfe & Van Wert, 2010), that is, how global precedence is adjusted when a given target level is presented more frequently than the other level.

EXPERIMENT 1

Experiment 1 was performed to investigate hierarchical object processing in a visual search task, employing Navon letters as target and nontargets (see Figure 2 for examples, and Deco & Heinke, 2007). The target letter could be represented either at the global or the local hierarchical level (i.e., a global target would combine with a local nontarget and vice versa). Differences between targets and nontargets were kept constant across both object levels, such that global and local targets always exhibited identical feature contrasts relative to the nontarget configurations. On the basis of previous findings, we expected faster RTs to global as compared to local targets, which would be indicative of a global precedence effect (Navon, 1977).

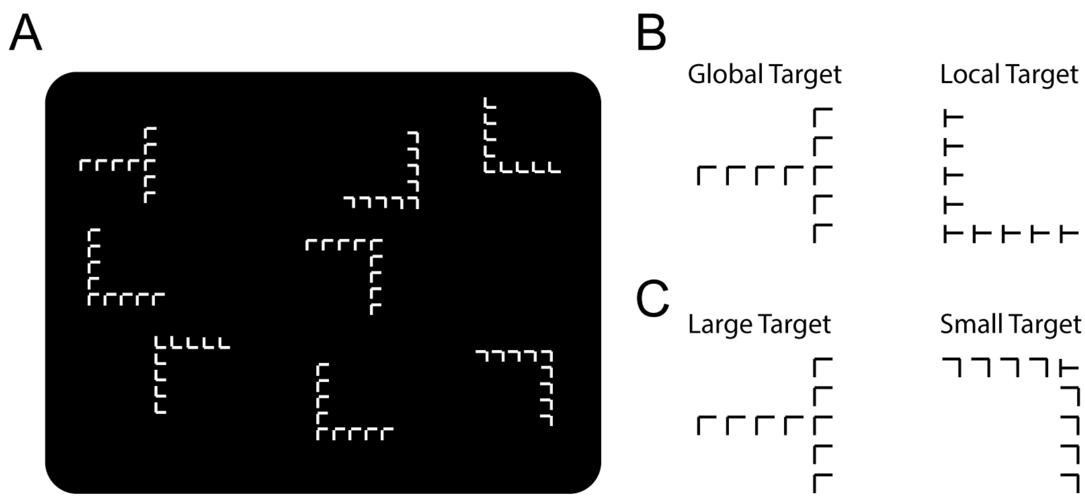

Figure 2. A. Example Navon-letter search display with eight hierarchical objects, presenting a global (large) target among L-shaped nontargets. B. Examples of a global (left panel), and local (right panel) target configuration (with leftward and rightward T orientation, respectively) as used in Experiments 1, 3 and 4. C. In Experiment 2, a large target (left panel) was compared to a small target (right panel, again displaying a leftward and rightward T orientation, respectively).

Methods

Participants. Fourteen observers (5 male; age range: 21 to 31 years; mean age = 26.6 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour.

The sample size was determined a-priori by means of a G*power 3 (Faul, Erdfelder, Buchner, & Lang, 2009) power analysis based on predicted effect size. From the results of previous studies (Nie et al., 2016), we predicted th effect size to be large (d = 0.75, according to Cohen, 1988) for the current experimental design. With 70% power given a .05 significance level, the sample size suggested was approximately 14 observers, which was used for Experiments 1-3.

Apparatus and Stimuli. The experiment was controlled by an IBM-PC compatible computer using Matlab routines and Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were Navon letters (Navon, 1977) presented in gray (8.5 cd/m^2) against a black (0.02 cd/m^2) background on a 17-inch monitor (1024 × 768 pixels screen zooming). Each stimulus consisted of a large (global) letter that subtended 2.6° × 2.6° of visual angle. Large letters were constructed from 9 small identical (local) letters, arranged in an invisible 5 by 5 grid. The small letters covered an area of 0.4° × 0.4°, with a gap of 0.15° between adjacent local letters.

Search arrays consisted of 4, 8, or 12 Navon letters (display size). Search displays were generated by placing one target configuration and 3, 7, or 11 nontargets randomly within the cells of an invisible 8 by 6 matrix (cell size 2.9°). Within each cell, the Navon stimuli were randomly jittered horizontally and vertically within a range of ±0.3°. Nontargets were global Ls made up of local L-shapes rotated randomly in one of four orthogonal orientations. The target was a (global or local) T-shape rotated 90° to either the left or the right. Targets could be defined at the global level (i.e., a global T would be constructed from 9 local Ls), or at the local level (i.e., 9 local Ts were combined to form a global L). Figure 2A shows an example display and corresponding global and local target configurations (Figure 2B).

Trial Sequence. Each trial started with the presentation of a central fixation cross for 500 ms. The fixation cross was followed by the search display, to which observers responded with a speeded response via mouse keys. The task was to search for an oriented T (either global or local) among L-nontargets and to decide as quickly and accurately as possible whether the T was pointing to the left or the right. Displays remained on-screen until a response was recorded. In case of an erroneous response, feedback was provided by an alerting red minus sign ("–") presented for 1000 ms at the screen center. Successive trials were separated by a 500-ms interval.

Design and Procedure. A two-factors within-subjects design was used. The independent variables were target level and display size. Target level could be either global (global T made up of local Ls; Figure 2B, left) or local (local Ts forming a global L; Figure 2B, right). Display size had three levels: 4, 8, or 12 Navon letters in a given search array.

At the beginning of the experiment, participants completed 1 block of 60 practice trials (generated randomly) to become familiar with the task. The formal experiment then presented 540 trials, divided into 9 blocks of 60 trials each.

Results

Response Accuracy. Mean percent correct responses for each observer and variable combination were calculated. Overall, responses were very accurate: 97% correct on average. A repeated-measures analysis of variance (ANOVA) of the percentage of correct responses, with the factors target level (global, local) and display size (4, 8, 12) revealed the main effect of target level to be significant, $F(1,13) = 5.46$, $p < .05$, $\eta_p^2 = .3$: accuracy was slightly (by 2.1%) higher with global relative to local targets. No other effects were significant (all ps > .7).

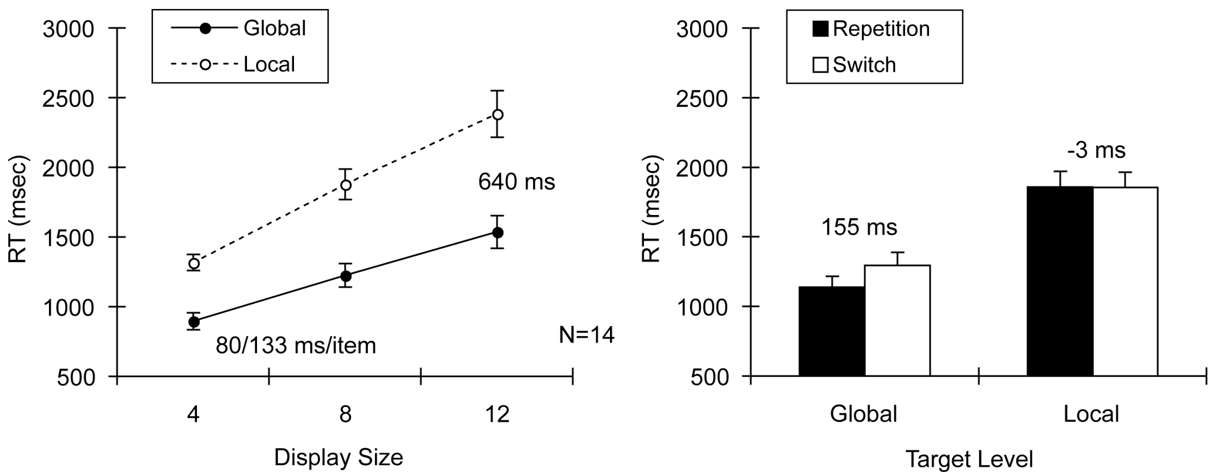

Figure 3. Mean search RTs (left panel) and corresponding intertrial effects (right panel) in Experiment 1. RTs are presented as a function of display size for global and local target levels (left), and as a function of the global/local target level for repetition and switch trials (right). Error bars represent ±1 standard errors of the mean (SEM).

Search RTs. Mean RTs for each observer were calculated excluding error responses and RTs deviating by more than three standard deviations from the mean. 4% of all trials, on average, were excluded by this outlier criterion (Experiments 2-4 yielded comparable exclusion rates). Figure 3 (left panel) presents the mean correct RTs as a function of display size, separately for global and local targets. Individual search RTs were subjected to a repeated-measures ANOVA with the factors target level and display size. This analysis revealed both main effects to be significant: target level, $F(1,13) = 46.1$, $p < .001$, $\eta_p^2 = .78$, and display size, $F(2,26) = 112.7$, $p < .001$, $\eta_p^2 = .9$. Global targets were detected much faster than local targets (mean global precedence effect: 640 ms), and RTs increased as display size increased (by 1048 ms from display size 4 to 12). Moreover, the two-way interaction was significant, $F(2,26) = 11.6$, $p < .001$, $\eta_p^2 = .47$: global targets yielded a smaller increase in search RTs with display size than local targets (search slopes were 80 and 133 ms/item for global and local targets, respectively). To summarize, global search RTs were much faster overall and exhibited shallower slopes than local search RTs, indicating that global search is more efficient.

Intertrial Effects. To examine whether search for global and local targets is influenced by intertrial contingencies, that is, whether performance on the current trial is affected by the hierarchical level that defined the target on the previous trial, a further 2 × 2 repeated-measures ANOVA of the RTs was performed with the factors target level (global, local) and previous trial (repetition, switch; coding whether the current target level was the same as or different to that on the previous trial); for this analysis, data were collapsed across the different display size conditions. The ANOVA revealed both main effects to be significant: target level, $F(1,13) = 47.3$, $p < .001$, $\eta_p^2 = .78$, and intertrial transition, $F(1,13) = 9.01$, $p < .01$, $\eta_p^2 = .41$. The main effect of target level essentially mirrored the above finding of global precedence. Moreover, search RTs were shorter (by 100 ms) when the target level repeated across trials as compared to when it switched. Importantly, the two-way interaction, was also significant, $F(1,13) = 10.4$, $p < .008$, $\eta_p^2 = .44$. Post-hoc comparisons revealed the level repetition effect to be reliable only for the global target level (155 ms; $t(13) = 4.55$, $p < .02$), but not for the local level (-3 ms, $t(13) = 0.18$, $p = .86$; see Figure 3, right panel). In other words, a level repetition benefit was evident only for global targets, without a comparable facilitatory effect for local targets.

Intertrial Search Slopes. An additional analysis examined whether the above described difference in search efficiency for global and local targets is dependent on intertrial transitions, that is: whether the hierarchical level that defined the target on the previous trial affects the search slope on the current trial. An effect of object hierarchy on the slopes could be taken to be indicative of a preattentive source of the effect (influencing attentional guidance), whereas the lack of a slope difference would suggest that priming effects arise post-selectively, that is, following attentional allocation to the target (Lamy, Carmel, Egeth, & Leber, 2006). To decide between these alternatives, a 2 × 2 repeated-measures ANOVA was performed on the search slopes with the factors target level (global, local) and previous trial (repetition, switch). This analysis revealed only the main effect of target level to be significant, $F(1,13) = 14.4$, $p = .002$, $\eta_p^2 = .53$, with shallower slopes for global than for local targets (80 vs. 133 ms/item), mirroring the above results. Neither the main effect of previous trial nor the interaction between target level and previous trial was significant (all $ps > .6$). This indicates that priming likely occurred subsequent to the allocation of attention. In turn, only the global precedence effect affected slopes – indicative of this effect arising from a preattentive source.

Response Priming. A further analysis was performed to examine whether the motor responses executed on successive trials influences the pattern of intertrial priming, that is, whether performance on a given (current) trial differs depending on whether the response was the same (repeated) or different (switched) relative to the previous trial (see also Robertson, 1996). A 2 × 2 repeated-measures ANOVA on mean priming effects (mean RTs for level switches minus repetitions), with the factors target level (global, local) and response type (repetition, switch), revealed the main effect of target level to be significant, $F(1,13) = 11.5$, $p = .005$, $\eta_p^2 = .47$, with larger priming for global than for local targets (155 vs. -3 ms), mirroring the above results. However, neither the main effect of response type nor the interaction between response type and target level was significant (all $ps > .12$), suggesting that the execution of the response did not influence the pattern of priming effects. Note that analogous analyses conducted in all subsequent experiments also revealed no evidence of response priming contributing to the intertrial priming effects.

Discussion

The results of Experiment 1 replicated previous findings of a global precedence effect (Navon, 1977) in visual search with multiple objects (Deco & Heinke, 2007). Overall, global targets were detected 640 ms faster than local targets, demonstrating a large overall bias of attention towards a global level of representation. In addition, search for global and local targets was also differentially influenced by display size (with search slopes of 80 [133] ms/item for global [local] targets), indicative of attentional guidance being more efficient for global than for local targets.

The results of the intertrial analyses, examining for priming patterns across hierarchical levels, also revealed an RT benefit, of 155 ms, for global target repetitions, but no comparable effect for local target repetitions (-3 ms). This novel, asymmetric pattern of intertrial effects mirrors the above pattern of global precedence: targets defined at the global level enjoy a performance benefit across trials, while there is no comparable effect for local target repetitions. Two additional analyses on the intertrial effects were performed to identify the critical stage of processing at which priming occurs. The first of these analyses showed that search efficiency (measured in terms of search RT slopes) was comparable between object-level repetition and switch trials, suggesting that priming occurred only subsequent to the allocation of attention. The second analysis, performed to examine the role of response repetitions and switches across trials (response priming), likewise revealed no difference between response repetitions and switches, indicating that object-level priming (on the current trial) was independent of the response executed on the previous trial. Taken together, these results suggest that priming in the current experiment occurs subsequent to the allocation of attention to the target (i.e., at a post-selective processing stage), but prior to response selection. Restated, priming relates to some form of top-down bias from (presumably implicit) short-term memory that – across trials – selectively facilitates the identification of global (but not of local) targets.

In summary, the results of Experiment 1 not only replicate previous findings on global precedence (e.g., Navon, 1977; Deco & Heinke, 2007), but also show novel evidence that the asymmetric bias of attention towards the global level is also evident in the pattern of intertrial priming, as illustrated by global priming in particular throughout a series of (randomly intermixed global- and local-target) trials. – Next, in Experiment 2, we investigated whether this pattern of global precedence could be accounted for by object size differences between global and local targets.

EXPERIMENT 2

Experiment 1 provides evidence that global precedence affects both attentional guidance and the pattern of priming across trials – suggesting that the global/local structure in a given display influences how attention is allocated on the current and the subsequent trial(s). However, despite perceptual hierarchy, the results revealed in Experiment 1 might also be explained by assuming that global targets are simply more salient than corresponding local targets. A global target is rendered by a single, large-scale configuration, whereas the local target consists of many, yet smaller items. Thus, the differential search efficiency might simply be attributable to a difference in object size and/or the degree of crowding at a given target level, with target salience being more influenced by crowding for local than for global targets. Such potential confounds owing to object size and/or crowding could alternatively (at least to some extent) account for the outcome of Experiment 1 (see also Navon, 1981; Kimchi, 1992).

Experiment 2 was performed to examine these alternatives, while keeping the basic set of stimuli comparable. In Experiment 2, performance for a large, global target (identical to the global target in Experiment 1; see Figure 2C, left) was compared to a new small target (see Figure 2C, right): a single, local-level "T" shape presented in the corner of a given large L-configuration, together with another 8 small distractor L's in a given configuration. Thus, the small target was shorter in the number of local "T"'s than the local target in Experiment 1 (in terms of task-critical information being provided by only 1 of 9 local items in the small target configuration, as compared to 9 of 9 local items in the local target configuration of Experiment 1), though with a roughly comparable level of crowding strength (determined by the distance of the target to the neighboring distractor items). If global precedence is determined primarily by relative object size and crowding, then the small target in Experiment 2 should give rise to either a comparable or even a larger precedence effect than the local target in Experiment 1.

Methods

Experiment 2 presented a large target (identical to Experiment 1), which was compared to a single, small target (see Figure 2C, left and right panels, respectively). For the small target, a single T letter was placed in the corner of a large L-shaped configuration among 8, neighboring small L letters. All other experimental parameters were the same as in Experiment 1. A new group of fourteen observers (5 female; age from 18 to 31 years; mean age = 20.4 years) performed the experiment. All participants had normal or corrected-to-normal visual acuity, and received course credits or payment of 8 Euro per hour.

Results

Response Accuracy. Responses were again very accurate, with 99% correct overall. A repeated-measures ANOVA on the percentage of correct responses with the factors target size (large, small) and display size (4, 8, and 12) revealed the main effect of target size to be significant, $F(1,13) = 5.76$, $p = .032$, $\eta_p^2 = .31$: accuracy was slightly (by 0.6%) higher with large relative to small targets. No other effects were significant (all ps $> .5$).

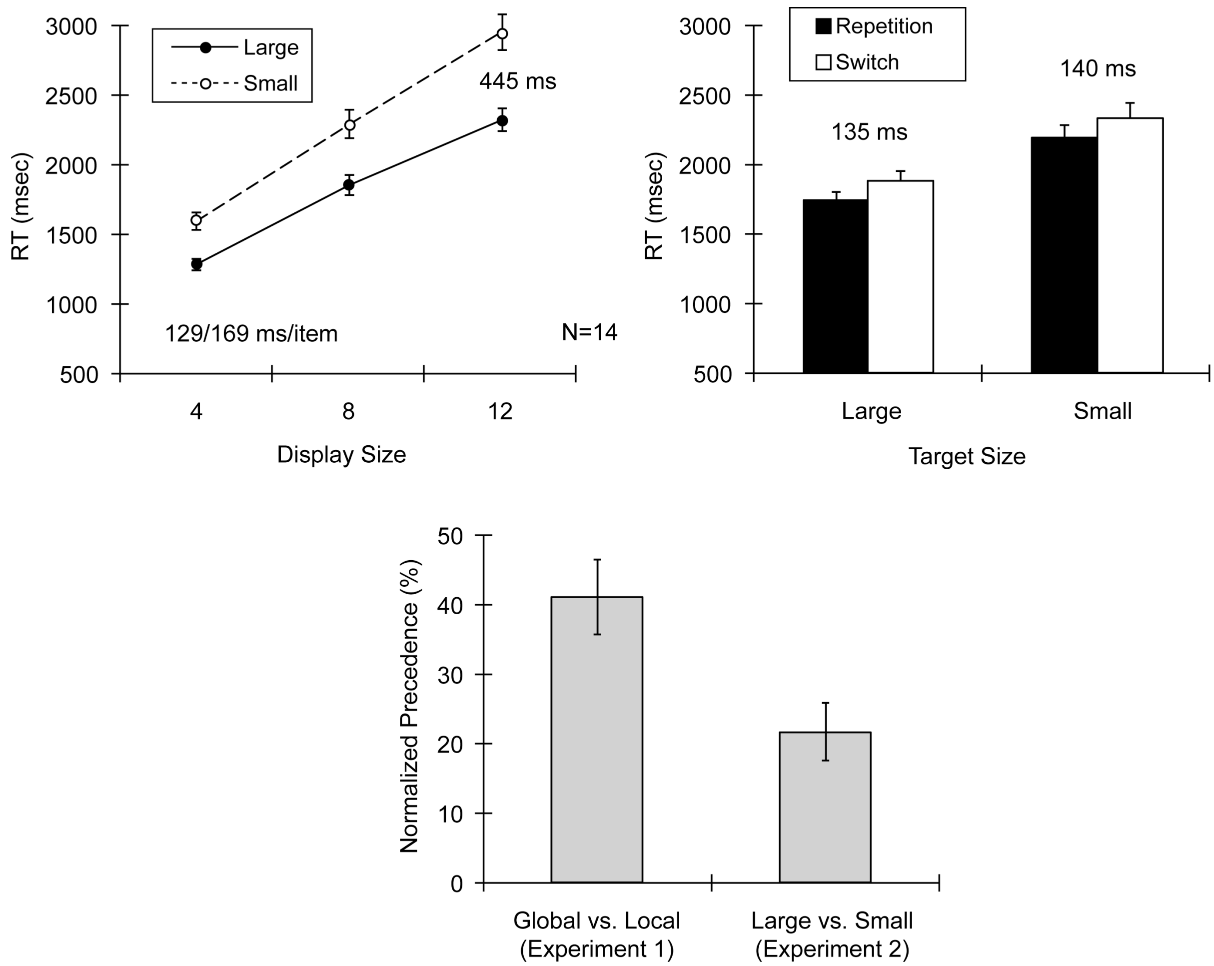

Figure 4. Mean search RTs (top left panel) and corresponding intertrial effects (top right panel) in Experiment 2. RTs are presented as a function of display size for large and small target sizes (top left), and as a function of target size for repetition and switch trials (top right). The graph in the bottom panel shows the normalized precedence effect in Experiment 1 (global vs. local) and Experiment 2 (large vs. small). Error bars represent ±1 SEM.

Search RTs. Figure 4 (left panel) presents the mean correct RTs as a function of display size, separately for large and small targets. A repeated-measures ANOVA of the search RTs, with the factors target size and display size, revealed significant main effects of target size, $F(1,13) = 25.0$, $p < .001$, $\eta_p^2 = .66$, and display size, $F(2,26) = 308.78$, $p <$

.001, $\eta_p^2 = .96$. Large targets were detected faster than small targets (mean size-precedence effect: 445 ms), and RTs increased by 1184 ms from display size 4 to 12. As in Experiment 1, the two-way interaction was significant, $F(2,26) = 7.38$, $p = .003$, $\eta_p^2 = .36$, that is, the per-item search rate was faster for large than for small targets (129 vs. 169 ms/item), though with overall reduced efficiency (i.e., steeper slopes) compared to Experiment 1 (80 vs. 133 ms/item), potentially due to inefficient search for the small target which impaired the overall search performance in Experiment 2.

In a subsequent step, a normalized precedence effect was computed [RT(local[small])-RT(global[large])/RT(mean)] to take overall group differences in the mean RTs into account. The normalized precedence effects were then compared between Experiments 1 and 2. An independent-samples t-test revealed a significant difference, $t(26) = 2.87$, $p = .008$, with the global/local precedence effect being larger in Experiment 1 (41%) than the large/small precedence effect in Experiment 2 (22%; see Figure 4, bottom panel). Thus, search for a large target was more efficient than search for a small target in Experiment 2, but the relative size of this difference was greatly reduced, by about half, when compared to Experiment 1~~.~~

Intertrial Effects. Next, to examine the pattern of intertrial priming, the RTs were subjected to a repeated-measures ANOVA with the factors target size (large, small) and previous trial (repetition, switch; i.e., same or different target size on consecutive trials). This analysis revealed both main effects to be significant: target size, $F(1,13) = 24.29$, $p < .001$, $\eta_p^2 = .65$, and previous trial, $F(1,13) = 10.5$, $p < .01$, $\eta_p^2 = .45$. The main effect of target size mirrors the above difference between large and small targets. In addition, search RTs were 138 ms shorter for cross-trial target size repetitions as compared to switches. However, in contrast to Experiment 1, the two-way interaction was far from significance, $F(1,13) = 0.005$, $p > .9$, $\eta_p^2 = .00$ – that is, there was a size repetition benefit, of a comparable magnitude, for both small and large targets (see also Figure 4, right panel).

Intertrial Search Slopes. A follow-up analysis examined whether the object size that defined the target on the previous trial affected the search slope on the current trial. An effect of object size on the slopes would be indicative of a preattentive source of the effect (influencing attentional guidance), whereas lack of a slope difference would suggest that priming effects arise following attentional allocation to the target (Lamy et al., 2006). A target size target size (large, local) × previous trial (repetition, switch) ANOVA on the search slopes revealed only the main effect of target size to be significant, $F(1,13) = 7.95$, $p = .015$, $\eta_p^2 = .38$, with shallower slopes for large than for small targets (129 vs. 169 ms/item), mirroring the above results. Neither the main effect of previous trial nor the interaction between target size and previous trial was significant (both $ps > .09$) – pointing to a post-selective locus of the symmetric size-based priming effects.

Discussion

An increase in the object size difference between large and small targets in Experiment 2 confirmed that size matters in visual search: large targets were detected more efficiently than small targets, though – in contrast to Experiment 1 – intertrial priming occurred to the same extent for both target sizes. However, the effect of size-based guidance in Experiment 2 was substantially reduced compared to the difference between hierarchical levels in Experiment 1 (22% vs. 41%, respectively), even though the basic display arrangement and the degree of crowding remained the same. In addition, the lack of an asymmetry in the pattern of large/small-target intertrial priming in Experiment 2 (as compared to a substantial difference between global and local target priming in Experiment 1) suggests a qualitative difference between these two search variants. The current search task is a size singleton search task, in which priming happens whenever a target size repeats (vs. switches; Maljkovic & Nakayama, 1994) along the scale of object size, indicating that intertrial priming is size-invariant in size singleton search, even though the small target is crowded with nontarget, local items. In this view, search defined by a rather strict global/local object hierarchy modulates performance to a larger extent than search on the basis of mere variations in object size, with a pattern of global precedence being evident in global/local search slopes and the associated priming effects (but less so for size-defined targets). This suggests that the global precedence effect in Experiment 1 is, at least to a substantial extent, determined by the relational structure as given by the representation of a whole versus their component parts, while variations in object size with comparable crowding strength as Experiment 1 reveal a much smaller influence.

Next, Experiment 3 was performed to investigate the stability of the global precedence effect – that is, whether a global bias could be reversed into a local bias.

EXPERIMENT 3

Experiment 1 revealed a global precedence effect, manifesting in terms of both prioritized attentional guidance and intertrial priming, which cannot readily be accounted for by relative size and/or crowding differences between target types (Experiment 2). Experiment 3 was designed to further examine whether the preference for global-level processing can be reduced, or even reversed into a bias favoring a finer (local) level. To this end, target prevalence was manipulated in Experiment 3 such that local targets were more likely to occur than global targets, increasing the behavioral relevance of the local-level (and decreasing that of the global-level) representation. This could engender a local bias, facilitating the processing of local targets.

Methods

Experiment 3 was essentially identical to Experiment 1, except that target prevalence was manipulated, presenting a local target configuration on 75% (and a global target on 25%) of all trials. Participants initially completed 1 block of 60 practice trials (with 15 [45] trials containing a global [local] target, respectively), followed by 720 experimental trials (180 [540] global [local] search targets). The experiment was divided into 12 blocks of 60 trials each. Fourteen new observers (3 male; age range: 20 to 32 years; mean age = 27.3 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour.

Results

Response Accuracy. Responses were very accurate overall with 98% of correct responses. A repeated-measures ANOVA on percent-correct responses with the factors target level (global, local) and display size (4, 8, and 12) revealed no significant effects (all ps $> .17$).

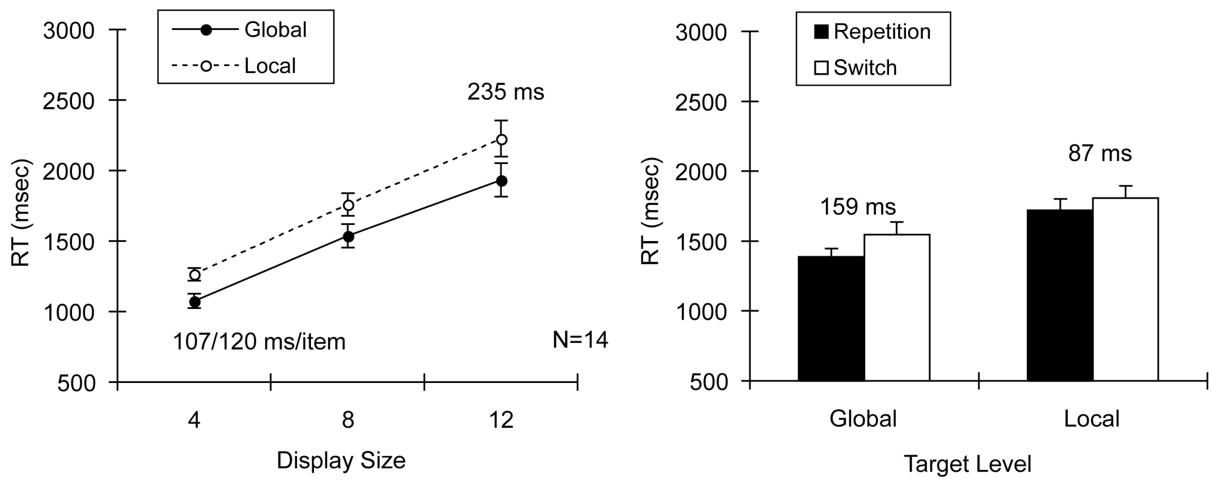

Figure 5. Mean search RTs (left panel) and corresponding intertrial effects (right panel) in Experiment 3 (with 75% local target prevalence). RTs are presented as a function of display size for global and local target levels (left), and as a function of the global/local target level for repetition and switch trials (right). Error bars represent ±1 SEM.

Search RTs. Figure 5 (left panel) presents the mean correct RTs as a function of display size, separately for global and local targets. A repeated-measures ANOVA of the search RTs, with the factors target level and display size, revealed both main effects to be significant: target level, $F(1,13) = 12.7$, $p < .01$, $\eta_p^2 = .5$, and display size, $F(2,26) = 166.6$, $p < .001$, $\eta_p^2 = .93$. Global targets were again detected faster than local targets (mean global precedence effect: 235 ms), and RTs increased by 940 ms from display size 4 to 12. Importantly, in contrast to Experiment 1, the two-way interaction was not significant, $F(2,26) = 1.12$, $p > .3$, $\eta_p^2 = .08$, that is, search efficiency in Experiment 3 was similar for global and local targets (with slopes of 107 and 120 ms/item, respectively). In a subsequent step, the overall global precedence effect was compared between Experiments 1 and 2: an independent-samples t-test revealed a significant difference, $t(26) = 3.18$, $p < .01$, with the global precedence effect being larger in Experiment 1 than in Experiment 2 (640 vs. 235 ms). Thus, global search RTs were still shorter than local RTs in Experiment 2, even though the overall global precedence effect was reduced by the introduction of a local bias in Experiment 2. Moreover, both target levels exhibited comparable slopes, indicative of essentially comparable attentional guidance for global and local targets.

Intertrial Effects. Next, the RTs were subjected to a repeated-measures ANOVA with the factors target level (global, local) and previous trial (repetition, switch; i.e., same or different target level on consecutive trials) to examine the pattern of intertrial effects. This analysis revealed both main effects to be significant: target level, $F(1,13) = 21.3$, $p < .001$, $\eta_p^2 = .62$, and intertrial transition, $F(1,13) = 10.5$, $p < .01$, $\eta_p^2 = .45$. The main effect of target level mirrors the above global precedence effect. In addition, search RTs were 114 ms shorter for cross-trial target level repetitions as compared to switches. However, in contrast to Experiment 1, the two-way interaction was not significant, $F(1,13) = 0.8$, $p > .39$, $\eta_p^2 = .06$, which means that level repetition benefits occurred for both global and local targets to a comparable extent (see also Figure 5, right panel).

Discussion

Experiment 3 revealed a reliable global precedence effect of 235 ms – despite the incentive to search for the local targets. This finding shows that the global attentional bias cannot easily be overcome. However, the introduction of a local bias nevertheless had a clear effect, as demonstrated by the reduction of global precedence in Experiment 3 as compared to Experiment 1. Moreover, there were no significant differences in the search slopes between global and local targets, indicating that, when local targets were made prevalent, attentional guidance was as efficient for local as for global search. Overall, this pattern of results shows that the availability of the local level representation can be increased to a certain extent, without however effectively abolishing the global precedence effect.

The cross-trial contingency analysis once again revealed an RT benefit for global target repetitions (as seen in Experiment 1), but this time an effect of comparable size also emerged for local target repetitions (whereas there was no benefit at all in Experiment 1). Thus, when attention was manipulated to prioritize the local level, attentional settings were likely adjusted to a somewhat finer scale, and priming effects manifested for both global- and local-level targets.

Taken together, the results of Experiment 3, while being consistent with those of Experiment 1, additionally show that a local bias can effectively reduce (though not completely overcome) global precedence, while enabling local-level target priming. – Next, in Experiment 4, we investigated whether changes in global precedence can be dynamically adjusted via a changing global/local bias.

EXPERIMENT 4

Experiment 3 yielded a reduced, but nevertheless reliable global precedence effect even when a local bias was introduced. Next, in Experiment 3, we further investigated whether global precedence can be modulated to adjust to a changing likelihood of global/local targets. To this end, target prevalence was manipulated across three phases of Experiment 4, by first presenting equally frequent global and local targets (phase 1), followed by more frequent local targets (phase 2), and subsequently more frequent global targets (phase 3). This change of target prevalence across the experiment allowed us to examine whether attentional guidance and object-level priming can dynamically adjust to the prevailing target level.

Methods

Experiment 4 was essentially identical to previous experiments, except for a fixed set size of (always) 12 Navon stimuli. Moreover, target prevalence was manipulated across the experiment, presenting equally frequent global (50%) and local (50%) targets in phase 1 (as in Experiment 1). Subsequently, local targets were shown more frequently (i.e., they were presented on 75% of trials) in phase 2 (similar to Experiment 3). Finally, global targets were presented in 75% of all trials in phase 3. Participants initially completed 1 block of 60 practice trials (with 30 [30] trials containing a global [local] target, respectively), followed by each 360 experimental trials in phase 1 (180 [180] global [local] targets), phase 2 (90 [270] global [local] targets), and phase 3 (270 [90] global [local] targets). Note that the overall prevalence of global and local targets remained equal across the entire experiment. The experiment was divided into 18 blocks of 60 trials each. Nineteen new observers (7 male; age range: 19 to 28 years; mean age = 23.8 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour. The sample size was larger in Experiment 4 because of an additional condition (i.e., experimental phase) that each observer completed.

Results

Response Accuracy. Responses were once again very accurate overall: 99% correct, on average. A repeated-measures ANOVA on percent-correct responses with the factors target level (global, local) and phase (1, 2, 3) revealed no significant effects (all ps > .15).

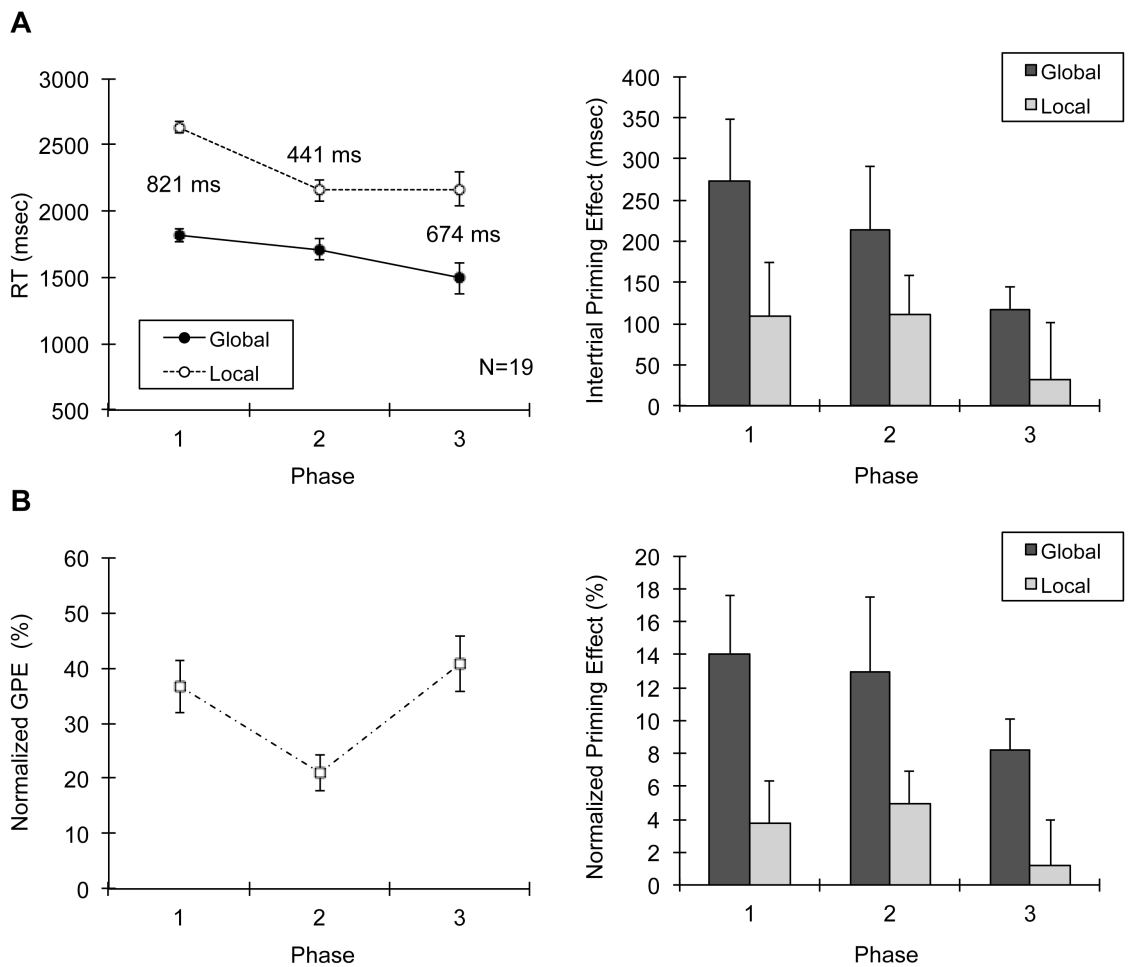

Figure 6. A. Mean search RTs (left panel) and corresponding priming effects (right panel) in Experiment 4 (with variable target prevalence). RTs are presented as a function of the experimental phase for global and local target levels (left). The panel on the right shows priming effects (i.e., intertrial switches minus repetitions) as a function of phase for global and local target levels. Note that global and local targets were equally prevalent in phase 1, followed by phase 2 which presented more prevalent local targets

(75% of trials), and finally more prevalent global targets (75% of trials) in phase 3. B. Left and right panels show the corresponding normalized global precedence effect (GPE) for each phase, and the normalized priming effects, respectively. Error bars represent ±1 SEM.

Search RTs. Figure 6A (left panel) presents the mean correct RTs as a function of phase, separately for global and local targets. Search RTs were analyzed by means of a 2 × 3 repeated-measures ANOVA with the factors target level and phase. This analysis revealed both main effects to be significant: target level, $F(1,18) = 51.2$, $p < .001$, $\eta_p^2 = .74$, and phase, $F(2,36) = 41.04$, $p < .001$, $\eta_p^2 = .7$. Global targets were once again detected faster than local targets (mean global precedence effect: 681 ms). The main effect of phase was due to an overall reduction in response latencies as the experiment progressed (2229, 2049, and 1660 ms in phases 1 to 3, respectively). Moreover, the two-way interaction was significant, $F(2,36) = 14.6$, $p < .001$, $\eta_p^2 = .45$, with mean global precedence effects of 821, 441, and 674 ms for phases 1, 2, and 3, respectively. That is, the global precedence effect became smaller with a local-target bias ($p = .001$, when comparing phases 1 and 2), and larger with a global-target bias ($p = .001$, when comparing phases 2 and 3). Thus, global precedence was modulated according to the prevailing global and local bias in each phase.

Intertrial Effects. Similar to the previous experiments, the intertrial effects were examined by a repeated-measures ANOVA of the RTs with the factors target level (global, local), previous trial (repetition, switch), and phase (1, 2, and 3). The analysis revealed all main effects to be significant: target level, $F(1,18) = 55.4$, $p < .001$, $\eta_p^2 = .76$, intertrial transition, $F(1,18) = 20.7$, $p < .001$, $\eta_p^2 = .53$, and phase, $F(2,36) = 32.5$, $p <$

.001, $\eta_p^2 = .64$. The main effects of target level and phase mirrored the results of the above analysis. The main effect of previous trial again showed (as for the previous experiments) that RTs were faster, by 200 ms, for target level repetitions as compared to switches across trials. In addition, the target level × phase interaction was significant, $F(2,36) = 7.8$, $p = .002$, $\eta_p^2 = .3$, mirroring the results of the above analysis: a modulation of the global precedence effect with a change of the global/local bias. Finally, the target level × previous trial interaction was significant, $F(1,18) = 4.9$, $p = .04$, $\eta_p^2 = .21$, reflecting a larger priming effect for global than for local targets (258 vs. 143 ms; see also Figure 6A, right panel). This pattern is essentially comparable to the results of Experiment 1. No other significant effects were obtained (all ps > .16). Thus, importantly, the intertrial effects were not modulated by the change in global/local bias across the different phases of Experiment 3 (mean priming effects for global/local targets were 273/109, 215/111, and 118/32 ms for phases 1, 2, and 3, respectively; the three-way interaction between target level, intertrial transition, and experimental phases was not significant: $F(2,36) = 0.31$, $p = .74$, $\eta_p^2 = .02$).

Normalized Search RTs. Since RTs were overall reduced by 569 ms during the course of the experiment, the reported analyses could potentially be confounded by this decrease in overall response latencies. To take this into account, a data normalization procedure was again applied computing the normalized global precedence effects relative to the mean RTs in each phase [RT(local)-RT(global)/RT(mean)] (see Experiment 2). These normalized precedence effects were then analyzed by means of a one-way ANOVA with the factor phase, which revealed the phase effect to be significant, $F(2,36) = 22.6$, $p < .001$, $\eta_p^2 = .56$. Overall, global targets were detected faster than local targets (mean normalized global precedence effect across phases: 33%), with the main effect of phase being due to an overall reduction of global precedence in phase 2 (37%, 21%, and 41% in phases 1, 2, and 3, respectively). That is, the global precedence effect became smaller with a local-target bias ($p = .001$, when comparing phases 1 and 2), and recovered with a global-target bias ($p < .001$, when comparing phases 2 and 3; $p > .44$, for comparison between phases 1 and 3; see Figure 6B, left panel). This pattern shows that global precedence dynamically changed according to the prevailing global/local bias for each phase, independently of the overall decrease in RTs.

Normalized Intertrial Effects. A comparable normalization procedure was subsequently also applied to the priming effects relative to the RTs in the global or local target condition [RT(level switch)-RT(level repetition)/RT(global, or local)]. These normalized intertrial effects were again examined by a repeated-measures ANOVA with the factors target level (global, local) and phase (1, 2, and 3). The analysis revealed only the main effect of target level to be significant, $F(1,18) = 10.2$, $p = .005$, $\eta_p^2 = .36$, with more marked priming for global than for local targets (12% vs. 3%; see Figure 6B, right panel). No other significant effects were obtained (all $ps > .25$). This pattern of results confirms the above results in showing a comparable difference between global- and local-level priming irrespective of the changes in prevalence throughout the phases of Experiment 4.

Discussion

Experiment 4 again revealed a robust global precedence effect of 681 ms, even though target prevalence changed throughout the experiment. This indicates that, by default, attention is strongly biased towards the global object representations. Nevertheless, the change of the global/local bias across phases affected the search RTs, as demonstrated by a reduction of the global precedence effect (by 380 ms) with a local-target bias (phase 2), and a subsequent recovery of global precedence (by 233 ms) with a global-target bias (phase 3). A near-identical pattern of effects was also found for the analysis of normalized RTs, which depicted a global precedence of 37% initially, followed by a reduction with the introduction of a local bias (21%) and a recovery with the reintroduction of a global bias (41%). This pattern of results demonstrates a robust *global* attentional state, which can however be adjusted dynamically (i.e., across phases).

While attention was dynamically adjustable, the cross-trial contingency analysis yielded no evidence of an analogous adjustment of priming. As in Experiment 1, priming was more pronounced for global- than for local-level target repetitions. Also, the overall priming effects decreased across the three experimental phases (possibly related to the search RTs becoming progressively faster). However, the asymmetric priming pattern did not change, that is, there was a constant advantage for global targets despite the changes of the global/local bias (and the very same bias remained evident even after the data normalization, i.e., after taking into account the relative decrease in RTs across the three phases of the experiment). This suggests that priming effects might initially be set in phase 1 and are not adjusted afterwards, in the subsequent phases that introduced the bias (by changing global- to local-level target prevalence).

In summary, the results of Experiment 4 show that the settings of global precedence can be adjusted in response to dynamic changes of target prevalence, but priming appears to be set initially and seems not to change afterwards with a change of the global/local bias.

General Discussion

The present study aimed to elucidate how the hierarchical representation of visual objects influences the zooming of attention and its concurrent effect on processing in visual short-term memory (the latter evidenced by intertrial priming). To this end, four experiments presented a novel visual search task with hierarchical objects (Navon, 1977). Observers were required to detect a target among nontargets, with the target object defined at either the global (large-scale) or the local (small-scale) level of representation. In Experiment 1, we found a robust global precedence effect, that is, overall faster responses to global as compared to local targets (by 640 ms), along with higher search efficiencies, that is, shallower search RT/set size slopes for global as compared to local targets (80 vs. 133 ms/item). Experiment 2 then showed that a mere difference in object size between large and small targets can also lead to substantial variations of search guidance; importantly, however, this size difference was substantially reduced by about half when compared to search for a hierarchical object (revealing normalized precedence effects of 22% and 41% in Experiments 2 and 1, respectively). Experiment 3, which returned to the original, hierarchical search task, was designed to to test the stability of global precedence by enhancing the relevance of local target representation. The results showed that an increase in local-target prevalence (i.e., relative local-target frequency) resulted in a numerically smaller, though nevertheless reliable, global precedence effect (of 235 ms). In Experiment 4, the global precedence effect was reliable too, but varied systematically in accordance with changes of target prevalence across three experimental phases. Together, these findings extend the literature on the global precedence effect (e.g., Navon, 1977; Kimchi, 2015) by showing that global object representation can expedite search and enhance search efficiency beyond effects merely attributable to differences in object size and/or crowding strength. Moreover, the findings show that global precedence is sensitive to environmental statistics, in particular: the frequency of local/global target occurrence (see discussion below).

The global precedence effect is typically obtained in paradigms that present displays with single hierarchical objects (Kimchi, 2015). In these task variants, differences between global and local object representations reflect differential processing of hierarchical object levels of a given stimulus that is currently being attended. However, global precedence may also manifest at preattentive stages of processing (Mattingley et al., 1997; Conci et al., 2009). With multiple hierarchical stimuli, visual search studies have shown that detection of a global configuration is more efficient than search for a local item arrangement (Conci et al., 2007a, 2007b; Conci et al., 2011; Deco & Heinke, 2007; Donnelly et al., 1991; Nie et al., 2016). In agreement with these findings, the current results show that attention is allocated predominantly on the basis of the prevailing global object representation, while being less sensitive to local detail.

Intertrial priming of global/local target levels

Our findings not only replicate the basic global precedence effect in a novel visual search task, but also provide new evidence that global/local object levels have differential consequences on selection history. For example, in Experiment 1, there was priming of the repeated global target level from one trial to the next. By contrast, there was no reliable priming effect for local target repetitions. This lack of intertrial priming for local-level repetitions cannot be explained in terms of relative size and/or crowding differences between global and local targets, as Experiment 2 (in which crowding strength in the small target was similar to the local target in Experiment 1) revealed no asymmetry of priming when comparing large versus small targets, suggesting that priming can happen along the scale of object size without any biases. In addition, with an increase in local target prevalence, priming also manifested for local target repetitions in Experiment 3, with effects statistically comparable to priming of the global target level. Finally, in Experiment 4, the prevalence of global and local targets was systematically varied across three experimental phases (starting with equal frequency in phase 1). As in Experiment 1, a more robust priming effect was observed for global than for local target repetitions, which did, however, not change with varying target prevalence. In other words, whereas the global precedence effect varied systematically with the changes in global/local target prevalence across phases, the asymmetric priming effect was unaffected by the changes in global/local target prevalence. This indicates that attention and (short-term) memory sources of global/local processing are dissociable within one and the same search task.

Interestingly, global-level priming – the benefit for global targets on global-level repeat versus switch trials – appeared to be comparable in magnitude across three experiments (155, 159, and 258 ms in Experiments 1, 3, and 4, respectively) and to change relatively little with a manipulation of target prevalence across experiments[3]. By contrast, local-level priming showed a much larger variability, being reliable only when local targets were consistently more prevalent throughout a longer 'phase' of trials (e.g.,

[1]In Experiment 4, a somewhat larger priming effect may have occurred because display size was constantly at 12 items, rather than varying (as in Experiments 1 to 3).

in Experiment 3); also, priming at different size scales appeared to be reliable whenever the size-defined target feature repeated (vs. switched) across trials, indicating that intertrial priming in size singleton search is size invariant (Experiment 2). Overall, these findings suggest that global-level priming operates rather automatically, and that local-level priming is not engaged by default.

Experimental studies that investigated the role of priming in visual search mostly presented objects that, across trials, varied in their simple-feature or feature-conjunction description (see Krummenacher & Müller, 2012; Lamy & Kristjánsson, 2013, for reviews). However, priming not only occurs for individual features and their conjunctions, but also at the level of complete-object representations (Kristjánsson et al., 2008), suggesting that separable features of a given target object are combined to form an integrated memory representation that may modulate preattentive (Müller et al., 2004) and/or postselective (Huang, Holcombe, & Pashler, 2004) stages of visual processing, influencing perceptual processing in the following trial episode(s). The current study extends these findings by showing that hierarchical levels of a given target object are represented asymmetrically in memory: Global object levels lead to effective priming, whereas local object levels are – by default – not facilitated to the same extent across trials. Several analyses performed on the present data were designed to specify the critical stage(s) at which object-level priming occurs. The results showed that the search RT slopes were unaffected by level repetitions versus switches, indicating that priming derives from processes engaged subsequent to attentional allocation at the target location. Moreover, analyses of response priming also revealed no evidence of level repetition benefits being enhanced when the required response repeated versus switched. Taken together, this pattern suggests that object-level priming occurred (i) at some postselective stage of processing and (ii) prior to the selection (and execution of) a motor response. Given this, priming is likely to derive from some form of top-down target template (i.e., short-term memory) that – across trials – selectively facilitates the identification of the target at a global object level. By contrast, local object levels are not represented with comparable consistency across trials (priming of local targets was lower or completely absent and showed a reliable facilitation only when local targets were highly prevalent).

A different line of work also investigated repetitions across levels in the classic Navon global/local paradigm, albeit only employing single hierarchical configurations presented in isolation (e.g., Hübner, 2000; Lamb & Yund, 1996, 2000; Lamb et al., 1998; Robertson, 1996; Ward, 1982). These studies, in general, showed a benefit when a target was repeatedly present at the same level – without major differences in priming across object levels even when the target configuration or the response changed (Filoteo et al., 2001; Robertson, 1996). This lack of a global/local asymmetry seems to be at odds with the current results. We suggest that the crucial difference between these studies is the type of task employed. When attention is already engaged at the location of a hierarchical target object (as in the standard Navon task), then both object levels can be represented across trials, leading to priming for both global and local levels. However, in a search task, attention is typically guided by preattentive object representation, which primarily codes for the global object level (e.g., Conci et al., 2007a; Deco & Heinke, 2007). It appears that this asymmetry between global and local levels not only affects the zooming of attention in search, but also the representation of the corresponding target-critical object level across trials. Thus, some form of target template may affect the zooming of attention across trials, and this target representation is, by default, biased towards global object representation.

Target prevalence in hierarchical search

Our results also show that global precedence changes with the prevailing environmental statistics: In Experiment 3, we introduced a consistent local bias, with local targets occuring in 75% of all trials. This local target *prevalence* effectively reduced the global *precedence* effect by about 60% (when compared to Experiment 1) and the differences in global/local search efficiency, measured in terms of search RT slopes, dramatically disappeared. Despite this large reduction of global precedence, global targets were still prioritized in search (i.e., the direction of the precedence effect did not reverse to favor local targets in Experiment 3). However, the frequent local targets in Experiment 3 did lead to significant local-level priming (comparable in magnitude to global-level priming). This pattern suggests that the default state of global precedence cannot be effectively reversed to facilitate search for prevalent local targets; rather, local-level priming occurs *in addition* to the global-level repetition benefits. Thus, global-level repetitions are encoded automatically, whereas local-level repetitions only take effect when there is a persistent incentive for task engagement (e.g., a consistent local bias in Experiment 3).

The results from Experiment 4 further suggest that systematic changes of target prevalence can lead to concurrent modulations in the size of global precedence during search. However, no comparable modulation occurred for the pattern of priming effects, which suggests that object-level priming is less dynamic than level-specific search guidance. Rather, priming appears to be primarily determined by initial phase and remains stable afterwards with changes of the global/local biases.

In general, targets tend to be easily missed when they appear infrequently across trials (Menneer, Donnelly, Godwin, & Cave, 2010). Search for rare targets is usually associated with a shift of the decision criterion, rather than a loss in perceptual sensitivity (Wolfe & Van Wert, 2010; Menneer et al., 2010). In addition, with two possible targets, a target that is presented with a higher level of prevalence may lead to enhanced performance at the expense of the less prevalent target (Godwin, Menneer, Cave, & Donnelly, 2010). Conversely, in the current experiments, prevalence modulated the efficiency of attentional guidance towards global or local object levels (as reflected in the search RT slopes). Target prevalence may thus be considered as engendering a top-down bias that influences attentional zooming towards global versus local levels.

A theoretical account of the default (global) attentional state

The present results might best be described in terms of a conceptual account of the default (global) attentional state within the framework of Reverse Hierarchy Theory (RHT, Ahissar & Hochstein, 2004). In this view, visual input is initially transmitted in parallel to high levels of processing, which generate a global percept or scene gist that guides attention. Local detail, in turn, is only available via recurrent connections that provide feedback to lower levels of representation in order to obtain a sufficient degree of zooming for extracting finer-grained local object information.

On the basis of RHT, we assume that attention is set by default towards a global level of object representation. Accordingly, initial visual scanning of a given display typically operates at a global level. Consequently, a global target is more conspicuous than a local target – which requires attention to be switched from the default global "comfort-state" to a local state, requiring a resource-demanding "zooming in" on local details (see also Stoffer, 1993). This process of setting attention towards a finer level of zooming generates additional costs for a local target, which manifest in terms of both an overall global precedence effect and a change in search efficiency (i.e., search RT slopes).

In this view, preattentive search guidance is biased towards a global level, whereas a second major influence may derive from post-selective stages of processing, that is, a top-down implicit short-term memory that encodes target-related information over successive trials. We assume that, across trials, the identification of a target at a particular hierarchical level is influenced by some kind of target template that instantiates a level-specific, global bias by default (see also Nie, Müller, & Conci, 2017). Accordingly, intertrial priming effects would generally be stronger for global than for local targets, because a repetition of the global target is supported by a default 'template' that is normally lacking for local-target repetitions. Only when local-target prevalence is high and attentional zooming is consistently biased towards the local level will the corresponding (i.e., local-target) template be established and maintained, resulting in facilitated local-target identification across successive trials.

Of note, according to our interpretation, the asymmetric pattern of global/local priming revealed in our experiments arises after a particular location has been selected by focal attention for local-target identification, that is, subsequent to some global visual scanning of the display. This change in hierarchy from global scanning to local identification may effectively disrupt the establishment of a local-level template. This could also explain why in tasks that do not require active, global scanning of the display prior to target identification – that is, tasks permitting attention to be maintained consistently at the same location (e.g., the screen center) – local-level priming effects may be comparable in magnitude to the priming of global object level (e.g., Hübner, 2000; Lamb & Yund, 1996, 2000; Lamb et al., 1998; Robertson, 1996; Ward, 1982).

The fact that target-template matching during postselective processing is governed by a hierarchical bias may simply be a sequential effect of the predominant global structure of preattentive visual scanning as suggested by RHT (see above). In this view, target-template matching follows the hierarchical structure with which a given target (configuration) is compared to the memory representations of possible target objects. There is evidence that short-term (and long-term) memory representations are hierarchically structured (Brady, Konkle, & Alvarez, 2011; Nie et al., 2017), which in the current context would correspond to global and local levels of object representation. While global object representations are readily available to be matched with current visual input, matching the local detail requires zooming-in to a local level within the template representation, and this is associated with a cost that is hard to overcome.

Conclusion

The present study reveals a functional connection between the representation of objects at varying hierarchical levels and the zooming of attention. Selective attention in visual search is set by default at a global level, with this setting being adjustable by target prevalence. Intertrial priming is also dominated by global object structure, that is: global precedence also manifests in terms of an asymmetric pattern of object-level priming at a post-selective stage of attentional processing, where level-specific target templates are buffered (and carried over across trial episodes) in some stable form of short-term memory. Our findings may be taken to indicate that, by default, attentional selection operates at a flexible global level of zooming alongside a stable global bias in the target templates of post-selective stimulus analysis, and the extraction of local detail is concomitant with a resource-demanding zooming-in of attention, revealing attention and (short-term) memory sources of global/local processing are linked yet show dissociable underlying dynamics.

References

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and towards structured representations. *Journal of Vision, 11*(5), 4, 1-34. doi: 10.1167/11.5.4

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision, 10*(4), 433-436. doi: 10.1163/156856897X00357

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Conci, M., Böbel, E., Matthias, E., Keller, I., Müller, H. J., & Finke, K. (2009). Preattentive surface and contour grouping in Kanizsa figures: Evidence from parietal extinction. *Neuropsychologia, 47*(3), 726-732. doi: 10.1016/j.neuropsychologia.2008.11.029

Conci, M., Müller, H. J., & Elliott, M. A. (2007a). Closure of salient regions determines search for a collinear target configuration. *Perception & Psychophysics, 69*(1), 32-47. doi: 10.3758/BF03194451

Conci, M., Müller, H. J., & Elliott, M. A. (2007b). The contrasting impact of global and local object attributes on Kanizsa figure detection. *Perception & Psychophysics, 69*(8), 1278-1294. doi: 10.3758/BF03192945

Conci, M., Töllner, T., Leszczynski, M., & Müller, H. J. (2011). The time-course of global and local attentional guidance in Kanizsa-figure detection. *Neuropsychologia, 49*(9), 2456-2464. doi: 10.1016/j.neuropsychologia.2011.04.023

Dale, G., & Arnell, K. M. (2013). Investigating the stability of and relationships among global/local processing measures. *Attention, Perception, & Psychophysics, 75*, 394-406. doi: 10.3758/s13414-012-0416-7

Deco, G., & Heinke, D. (2007). Attention and spatial resolution: A theoretical and experimental study of visual search in hierarchical patterns. *Perception, 36*, 335-354. doi: 10.1068/p5633

Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 561-570. doi: 10.1037/0096-1523.17.2.561

Enns, J. T., & Kingstone, A. (1995). Access to global and local properties in visual search for compound stimuli. *Psychological Science, 6*(5), 283-291. doi: 10.1111/j.1467-9280.1995.tb00512.x

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. doi: 10.3758/BRM.41.4.1149

Filoteo, J. V., Friedrich, F. J., & Stricker, J. L. (2001). Shifting attention to different levels within global-local stimuli: A study of normal participants and a patient with temporal-parietal lobe damage. *Cognitive Neuropsychology, 18*(3), 227-261. doi: 10.1080/02643290125839

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimension-weighting" account. *Perception & Psychophysics, 58*(1), 88-101. doi: 10.3758/BF03205479

Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition, 18*(10), 1439-1463. doi: 10.1080/13506285.2010.500605

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791-804. doi: 10.1016/S0896-6273(02)01091-7

Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition, 32*(1), 12-20. doi: 10.3758/BF03195816

Hübner, R. (2000). Attention shifting between global and local target levels: The persistence of level-repetition effects. *Visual Cognition, 7*(4), 465-484. doi: 10.1080/135062800394612

Kimchi, R. (1992). Primacy of wholistic processing and global local paradigm: A critical review. *Psychological Bulletin, 112*(1), 24-38. doi: 10.1037/0033-2909.112.1.24

Kimchi, R. (2015). The perception of hierarchical structure. In J. Wagemans (Ed.), *Oxford Handbook of Perceptual Organization*. Oxford, UK: Oxford University Press.

Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics, 72*(1), 5-18. doi: 10.3758/app.72.1.5

Kristjánsson, Á., Ingvarsdöttir, Á., & Teitsdöttir, U. D. (2008). Object- and feature-based priming in visual search. *Psychonomic Bulletin & Review, 15*(2), 378-384. doi: 10.3758/pbr.15.2.378

Krummenacher, J., & Müller, H. J. (2012). Dynamic weighting of feature dimensions in visual search: Behavioral and psychophysiological evidence. *Frontiers in Psychology, 3*, 221. doi: 10.3389/fpsyg.2012.00221

Lamb, M. R., London, B., Pond, H. M., & Whitt, K. A. (1998). Automatic and controlled processes in the analysis of hierarchical structure. *Psychological Science, 9*(1), 14-19. doi: 10.1111/1467-9280.00003

Lamb, M. R., & Yund, E. W. (1996). Spatial frequency and attention: Effects of level-, target-, and location-repetition on the processing of global and local forms. *Perception & Psychophysics, 58*(3), 363-373. doi: 10.3758/BF03206812

Lamb, M. R., & Yund, E. W. (2000). The role of spatial frequency in cued shifts of attention between global and local forms. *Perception & Psychophysics, 62*(4), 753-761. doi: 10.3758/BF03206921

Lamy, D. F., Carmel, T., Egeth, H. E., & Leber, A. B. (2006). Effects of search mode and intertrial priming on singleton search. *Perception & Psychophysics, 68*(6), 919-932. doi: 10.3758/BF03193355

Lamy, D. F., & Kristjánsson, Á. (2013). Is goal-directed attentional guidance just intertrial priming? A review. *Journal of Vision, 13*(3), 14. doi: 10.1167/13.3.14

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition, 22*(6), 657-672. doi: 10.3758/BF03209251

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics, 58*(7), 977-991. doi: 10.3758/BF03206826

Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science, 275*(5300), 671-674. doi: 10.1126/science.275.5300.671

Menneer, T., Donnelly, N., Godwin, H. J., & Cave, K. R. (2010). High or low target prevalence increases the dual-target cost in visual search. *Journal of Experimental Psychology: Applied, 16*(2), 133-144. doi: 10.1037/a0019569

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics, 57*(1), 1-17. doi: 10.3758/BF03211845

Müller, H. J., Krummenacher, J., & Heller, D. (2004). Dimension-specific intertrial facilitation in visual search for pop-out targets: Evidence for a top-down modulable visual short-term memory effect. *Visual Cognition, 11*(5), 577-602. doi: 10.1080/13506280344000419

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353-383. doi: 10.1016/0010-0285(77)90012-3

Navon, D. (1981). The forest revisited: More on global precedence. *Psychological Research, 43*(1), 1-32. doi: 10.1007/BF00309635

Navon, D. (2003). What does a compound letter tell the psychologist's mind? *Acta Psychologica, 114*, 273-309. doi: 10.1016/j.actpsy.2003.06.002

Nie, Q.-Y., Maurer, M., Müller, H. J., & Conci, M. (2016). Inhibition drives configural superiority of illusory Gestalt: Combined behavioral and drift-diffusion model evidence. *Cognition, 150*, 150-162. doi: 10.1016/j.cognition.2016.02.007

Nie, Q.-Y., Müller, H. J., & Conci, M. (2017). Hierarchical organization in visual working memory: From global ensemble to individual object structure. *Cognition, 159*, 85-96. doi: 10.1016/j.cognition.2016.11.009

Paquet, L., & Merikle, P. M. (1988). Global precedence in attended and nonattended objects. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 89-100. doi: 10.1037/0096-1523.14.1.89

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

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3*(3), 422-435. doi: 10.1037/0096-1523.3.3.422

Robertson, L. C. (1996). Attentional persistence for features of hierarchical patterns. *Journal of Experimental Psychology: General, 125*(3), 227-249. doi: 10.1037/0096-3445.125.3.227

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29*, 203-227. doi: 10.1146/annurev.neuro.29.051605.112939

Stoffer, T. H. (1993). The time course of attentional zooming: A comparison of voluntary and involuntary allocation of attention to the levels of compound stimuli. *Psychological Research, 56*(1), 14-25. doi: 10.1007/BF00572129

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

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218-1252. doi: 10.1037/a0029334

Ward, L. M. (1982). Determinants of attention to local and global features of visual forms. *Journal of Experimental Psychology: Human Perception and Performance, 8*(4), 562-581. doi: 10.1037/0096-1523.8.4.562

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology, 20*(2), 121-124. doi: 10.1016/j.cub.2009.11.066

Chapter IV

Hierarchical organization in visual working memory:

From global ensemble to individual object structure

Abstract

When remembering a natural scene, both detailed information about specific objects and summary representations such as the gist of a scene are encoded. However, formal models of change detection that are used to estimate working memory capacity, typically assume observers simply encode and maintain memory representations that are treated independently from one another without considering the (hierarchical) object or scene structure. To overcome this limitation, we present a hierarchical variant of the change detection task that attempts to formalize the role of object structure, thus, allowing for richer, more graded memory representations. We demonstrate that detection of a global-object change precedes local-object changes of hierarchical shapes to a large extent. Moreover, when systematically varying object repetitions between individual items at a global or a local level, memory performance declines mainly for repeated global objects, but not for repeated local objects, which suggests that ensemble (i.e., summary) representations are likewise biased towards a global level. In addition, this global memory precedence effect is shown to be independent from encoding durations, and mostly can not be attributed to differences in saliency or shape discriminability at global/local object levels. This pattern of results is suggestive of a global/local difference occurring primarily during memory maintenance. Altogether, these findings challenge visual-working-memory (vWM) models that propose that a fixed number of objects can be remembered regardless of the individual object structure. Instead, our results support a hierarchical model that emphasizes the role for structured representations among objects in vWM.

Introduction

Visual working memory (vWM) enables cognitive functions to operate independently of direct retinal stimulation, with current contents in vWM supporting goal-directed behavior. However, in order to maintain a stable representation of the world, only a limited amount of sensory information of an individual's total visual input can be represented in vWM (Luck & Vogel, 2013, for a review). Hence, a major focus of studies on vWM is to describe the organizational principles by which this limited cognitive space can be used efficiently for the internal representation of visual input.

Much of the work in this regard has followed from Luck and Vogel's (1997) seminal study. They devised a change detection task in which a memory array of colored squares (varying across trials from a single square to up to 12 squares) was presented for a few hundred milliseconds (ms). Subsequent to a brief blank delay of about 1 second, a probe array was presented that contained the same items as the memory array – except (on half of the trials) for one object that was displayed in a different color. Observers were required to detect the change by giving a yes/no (two-alternative) forced-choice response. The results from these experiments indicated that participants had a vWM capacity around three to four items (Luck & Vogel, 1997). Moreover, they found that an individual's capacity did not change with the number of features that combined to form a given object. For instance, detecting a feature change was equivalent when comparing objects determined by conjunctions of four features (and where all of these features could potentially change) with objects defined by a single feature only. This observation led Luck and colleagues to propose that the capacity of vWM is (relatively) fixed: there are only a limited number of available slots, each one capable of storing a single object representation regardless of its complexity (Vogel, Woodman, & Luck, 2001).

The slot model has been challenged from at least two perspectives, and the nature of vWM capacity limits remains a topic of vigorous debate. One open question concerns the influence of object complexity. For example, contrary to the findings of Vogel et al. (2001), others have shown that vWM capacity declines with increasing object complexity (e.g., Alvarez & Cavanagh, 2004). A second challenge arises from the notion that, rather than there being a limited number of available slots, vWM capacity may depend on a single information-limited cognitive resource. Evidence for this alternative view was originally provided by Bays and Husain (2008) using a variation of the change detection paradigm. On each trial, a sample array of colored squares was presented, followed by a brief delay and a subsequent test probe. The task was to report whether the test probe was displaced to the left or the right of the corresponding item in the memory array. The results showed that performance remained near-perfect for sufficiently large displacements even when presenting rather large set sizes (e.g., a set size of 8 objects, Bays & Husain, 2008). Moreover, mnemonic precision declined monotonically as a function of memory load – an outcome expected if vWM were supported by a limited-capacity resource that requires to be distributed across more objects as the memory load increases. In this view, retaining a small number of items can be accomplished with relatively high precision; but an increase in the number of to-be-remembered items leads to a decline in the precision with which items can be remembered. This trade-off between quality and quantity of mnemonic representations implies that memory resources can be allocated flexibly among several items stored in vWM to maximize mnemonic precision, given the available resources.

Overall, these findings imply that information limits in vWM are determined both by the number and the precision of mnemonic representations (e.g., Luck & Vogel, 1997; Wilken & Ma, 2004). To accommodate this, slot models have been modified to allow for variable representational precision within a slot (Luck & Vogel, 2013; Zhang & Luck, 2008). However, one contentious question that remains is how best to explain capacity limitations: Is the amount of visual information an individual can retain in vWM limited because of a limited number of slots (i.e., caused by an absolute ceiling in performance) or because, at some point, resources have been distributed so widely that the mnemonic fidelity for any given item becomes too poor for the item to be retrievable? Moreover, vWM models as described above tend to focus on how observers encode independent features or objects from rather simple arrays of segmented geometric shapes without considering the rather complex relational structure that is usually present in the natural ambient environment.

Contrary to the simple stimulus arrays used in most vWM studies, memory for real-world scenes has been shown to depend largely on organizational principles, that is, mechanisms that impose structure on visual input. For example, when trying to remember natural scenes, the gist of that scene (e.g., a statistical summary, or ensemble representation) is encoded, in addition to the detailed information about relatively few specific objects (Conci & Müller, 2014; Hollingworth & Henderson, 2003; Oliva, 2005). Moreover, the gist can be used to guide people's choice of which specific objects to be recalled (Hollingworth & Henderson, 2000). For instance, when trying to retrieve the details of the scene, the gist can lead to recall of objects that are consistent with the scene, but were actually not present at all in the memory display (Lampinen, Copeland, & Neuschatz, 2001; Miller & Gazzaniga, 1998). Conversely, gist representations seem to facilitate the encoding of (semantic) outlier objects: items are more likely to be both fixated and encoded into memory when these are semantically inconsistent with the background scene (Hollingworth & Henderson, 2000, 2003). Arguably, these findings from studies with naturalistic scenes show that observers have a strong tendency to structure and organize a given sensory input into some higher-order regularity, that is, "compression" of the available information in order to spare the limited cognitive resources. These results seem to be strongly linked to studies that investigate the relational, or, hierarchical structure in objects (e.g., a global triangle composed of local squares, see Kimchi, 1982). For instance, the identification of local-level elements in a hierarchical stimulus configuration (e.g., Navon letters) is influenced by representations at the global object level (Navon, 1977; Wagemans, Elder, et al., 2012). Moreover, global levels of a target object guide attention more efficiently during visual search than local object levels, and this global precedence in selecting a target on a given trial is transferred to subsequent trials, evidencing a persistent global bias (Conci, Müller, & Elliott, 2007a, 2007b; Conci, Töllner, Leszczynski, & Müller, 2011; Nie, Müller, & Conci, 2016; Wagemans, Feldman, et al., 2012). Thus, an observer's representation of both real-world scenes and simpler displays with geometric objects consists not only of information about the individual objects but also includes structural information and a broad, gist-like representation of the overall information presented.

In fact, it has been shown that perceptual organization also plays a significant role in vWM – even for rather simple memory arrays. For instance, when separate objects are grouped together into perceptual units (e.g., by means of closure or repetition), this also results in better vWM performance, as each unit in the group can be encoded into a perceptual Gestalt, thus improving memory capacity (Woodman, Vecera, & Luck, 2003; Xu, 2006; Xu & Chun, 2007). Moreover, maximizing the symmetry of an object via completion improves vWM performance (Chen, Müller, & Conci, 2016). Together, these findings point to the use of organizational principles to optimize the storage of items, so as to relieve vWM capacity (see also Jiang, Olson, & Chun, 2000).

Relatedly, there is mounting psychophysical evidence that even in simple memory displays, items are not treated independently (see Brady, Konkle, & Alvarez, 2011, for a review). For instance, if a display is changed from mostly dark squares to mostly bright squares, then observers notice this change more efficiently than a matched change that does not alter the global statistics of the scene (Alvarez & Oliva, 2009; Victor & Conte, 2004). Moreover, when computing the average visual representation in simple arrays of items from a given display, observers discount outlier objects to only represent the majority of consistent items (Haberman & Whitney, 2010). Brady and Alvarez (2011) reported further evidence suggesting that the representation of "ensemble statistics" influences the representation of individual items: Observers are biased in reporting the size of an individual item by the mean size of all (or of potentially task-relevant) items in a particular display – which they interpreted as reflecting the integration of information about the ensemble size of items in the display with information about the size of a particular item. However, existing formal models of the architecture and capacity of vWM do not take into account the possibility of such hierarchically structured representations, but only consider how many individual items are remembered when treated independently (Luck & Vogel, 2013; Ma, Husain, & Bays, 2014).

In the present study, we developed a hierarchical variant of the change detection task to investigate how different object levels (i.e., global or local representations) are represented in vWM. Within each trial, multiple hierarchical shapes were presented in a memory array, followed by a test probe that appeared after a brief delay. Observers were required to memorize all objects and hierarchical levels, and to indicate whether a change occurred in the probe item, irrespective of the level (global or local) where the change had occurred. In addition, we manipulated between-object repetition at both hierarchical levels, systematically varying the repetition in displays with several hierarchical objects presenting similar objects at global or local levels. These variations of repetition permitted investigating how repetitions of object identities within a given display potentially affect the summary representation that is generated alongside with individual item memory. Both hierarchical structure and object repetition are known to provide a structural representation and corresponding statistical information about the objects that are to be remembered (Kimchi & Palmer, 1982; Alvarez, 2011).

To anticipate, our results demonstrate a form of hierarchical storage in vWM: The remembered object representation of individual items was biased toward the global object level, with better memory performance in detecting global as compared to local changes. Moreover, object repetitions also affected vWM capacity – with a reduction in performance that was evident in particular for globally repeated objects in a given display. This suggests that, contrary to existing models of vWM, items are not retrieved as integrated and independent units, but instead, an item's (reported) representation is constructed by combining global and local structure about that specific item with information about the set of items at a global level of the available ensemble statistics.

EXPERIMENTS 1A and 1B

Experiments 1A and 1B were performed to investigate the representation of object structure in visual working memory (vWM) using a variant of the change detection task with hierarchical shapes. Observers were presented with a memory array that contained varying numbers of shapes with a global/local structure. Subsequent to a delay, a test probe was presented that could either undergo a change at a global level, at a local level, or remain the same (at both levels) compared to the target item at the same location in the preceding memory array (see Figure 1 for examples; see also Kimchi & Palmer, 1982). The key difference between the two experiments was the manipulation of set size, that is, the number of items that participants were required to remember. Set sizes were 2, 4, and 6 in Experiment 1A, and 1, 2, and 3 in Experiment 1B. On the basis of previous findings from visual search studies (e.g., Conci, Müller & Elliott, 2007a, 2007b; Nie, Maurer, Müller, & Conci, 2016), we expected greater detection sensitivity and higher memory capacity for global as compared to local object representations.

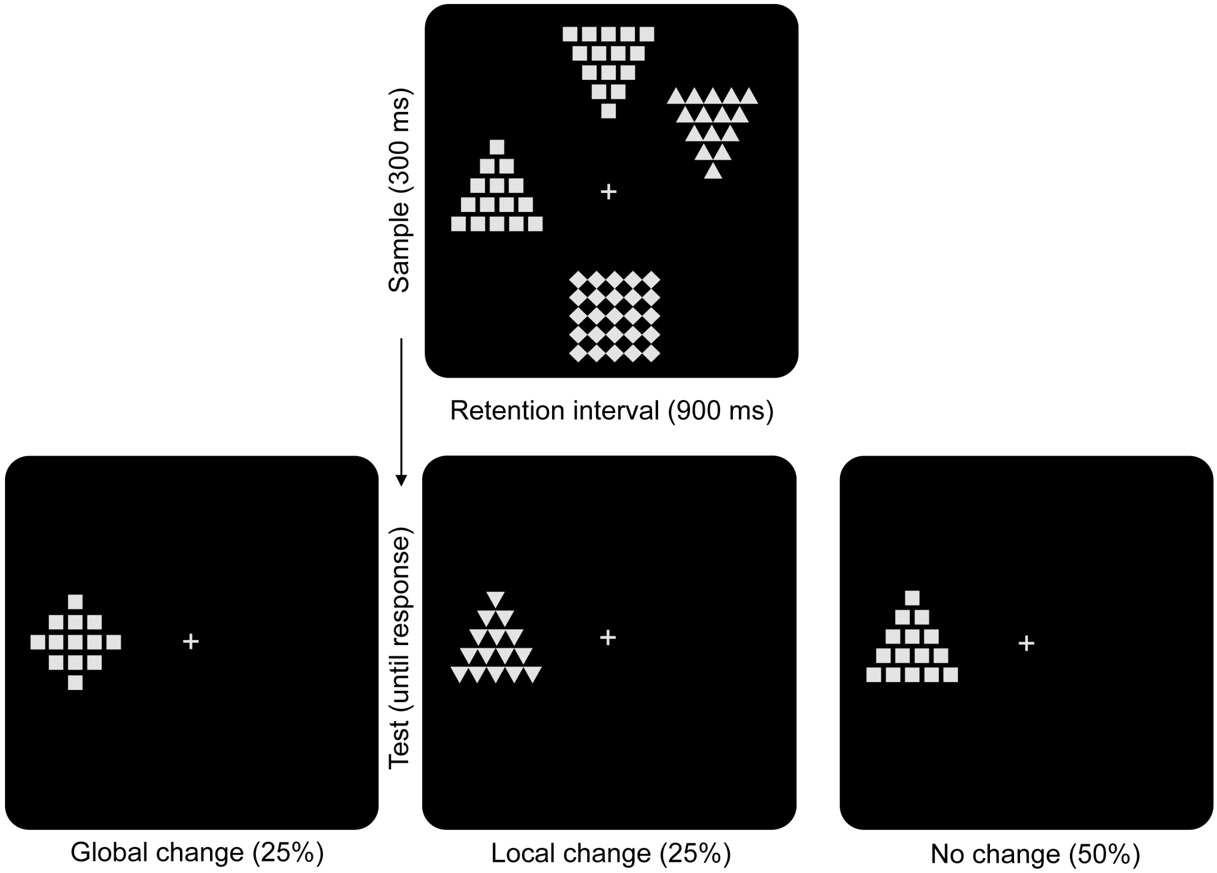

Figure 1. Example trial sequence in Experiments 1A and 1B. Observers viewed a sample display that consisted of a variable number of to-be-memorized hierarchical shapes arranged in a circle (top panel). After a brief (blank) delay, a test display was presented that either presented a probe item with a change at the global object-level (left panel), a change at the local object-level (middle panel), or an unchanged item (right panel).

Methods

Participants. Two different groups of observers participated in Experiments 1A (N = 12; 8 female; age range from 19 to 32 years; mean age = 20.5 years) and 1B (N = 10; 7 female; age range from 20 to 31 years; mean age = 22.8 years). All participants reported normal or corrected-to-normal visual acuity. Participants provided written consent to the procedure of the experiment, which was approved by the local ethics committee, in accordance with the Declaration of Helsinki. They received course credits or payment of 8 Euro per hour for their participation.

Apparatus and Stimuli. The experiments were conducted with an IBM-PC compatible computer using Matlab routines and Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were hierarchical shapes (as in Kimchi & Palmer, 1982) presented in gray (8.5 cd/m^2) against a black (0.02 cd/m^2) background on a 17-inch monitor screen (1024 × 768 pixels). Each stimulus consisted of a global shape that subtended 2.6° × 2.6° of visual angle. Global shapes were constructed from various (13-25) identical local shapes, arranged in an invisible 5 × 5 grid. Local shapes covered an area of 0.4° × 0.4°. There was a 0.15° gap between each neighboring local shape.

Memory arrays consisted of 2, 4, and 6 hierarchical shapes in Experiment 1A, and of 1, 2, and 3 shapes in Experiment 1B. All shapes were presented on an imaginary circle of 6° radius, with positions randomly selected from eight equally spaced, fixed locations on the circle. Each hierarchical shape was constructed randomly from a predefined set of 4 distinctive local shapes (squares, diamonds, and up- or downward-pointing triangles), thus forming 12 different shape configurations, with the constraint that global and local shapes were always different from each other. Target (probe) locations were randomly selected on each trial. Figure 1 shows an example display with set size 4, and three possible variants of test probes, which illustrate global and local changes, and the no-change condition (left, middle, and right panels, respectively).

Trial Sequence. Each trial started with the presentation of a central fixation cross for 500 ms. The fixation cross was followed by the memory display presented for 300 ms; observers were instructed to memorize all presented hierarchical shapes in this (memory) display at both global and local levels. Subsequent to the memory display, there was a blank screen for 900 ms, followed by a test probe that was presented at one, randomly chosen location from the preceding memory array. The task was to decide whether the test probe was the same (at both global and local levels) or different (with a change at either the global or local level) relative to the item that had been previously presented at the same location in the memory array. The probe item remained on-screen until a response was recorded. Participants were instructed to respond as accurately as possible without emphasizing response speed. In case of an erroneous response, feedback was provided by an alerting red sign ("–") presented for 1000 ms at the center of the screen. Each trial was separated from the next by a 500-ms interval.

Design and Procedure. A three-factor within-subjects design was used for both experiments. The independent variables were change, level, and set size. The first variable, change indicated the memory-probe transition and could be either present (50% of the trials) or absent (50%). The second variable, level, refers to the hierarchical level at which a potential change occurred (global or local). For a global change, the test probe differed from the memorized target item (at the same location) only at the global level (Figure 1, lower left), whereas for a local change, the test probe differed from the target item only at the local level (Figure 1, lower middle). The third variable, set size, had three levels and determined the number of items presented in the memory array: 2, 4, and 6 hierarchical shapes in Experiment 1A and 1, 2, and 3 in Experiment 1B, respectively.

At the beginning of both experiments, participants completed 1 block of 48 practice trials generated randomly to familiarize them with the task. The subsequent, actual experiment then presented 576 trials, divided into 12 blocks of 48 trials each.

Data analysis. In the present experiments (as well as the subsequent ones), vWM performance was determined by a signal-detection-theoretic sensitivity measure: d-prime (see MacMillan & Creelman, 2004). For Experiments 1A and 1B, we also estimated the number of individual objects remembered (at global or local levels) using Cowan's K (Cowan, 2001): $K = (H - FA) \times N$, where K is the number of items stored, H is the hit rate, FA is the false alarm rate, and N is the number of items presented.

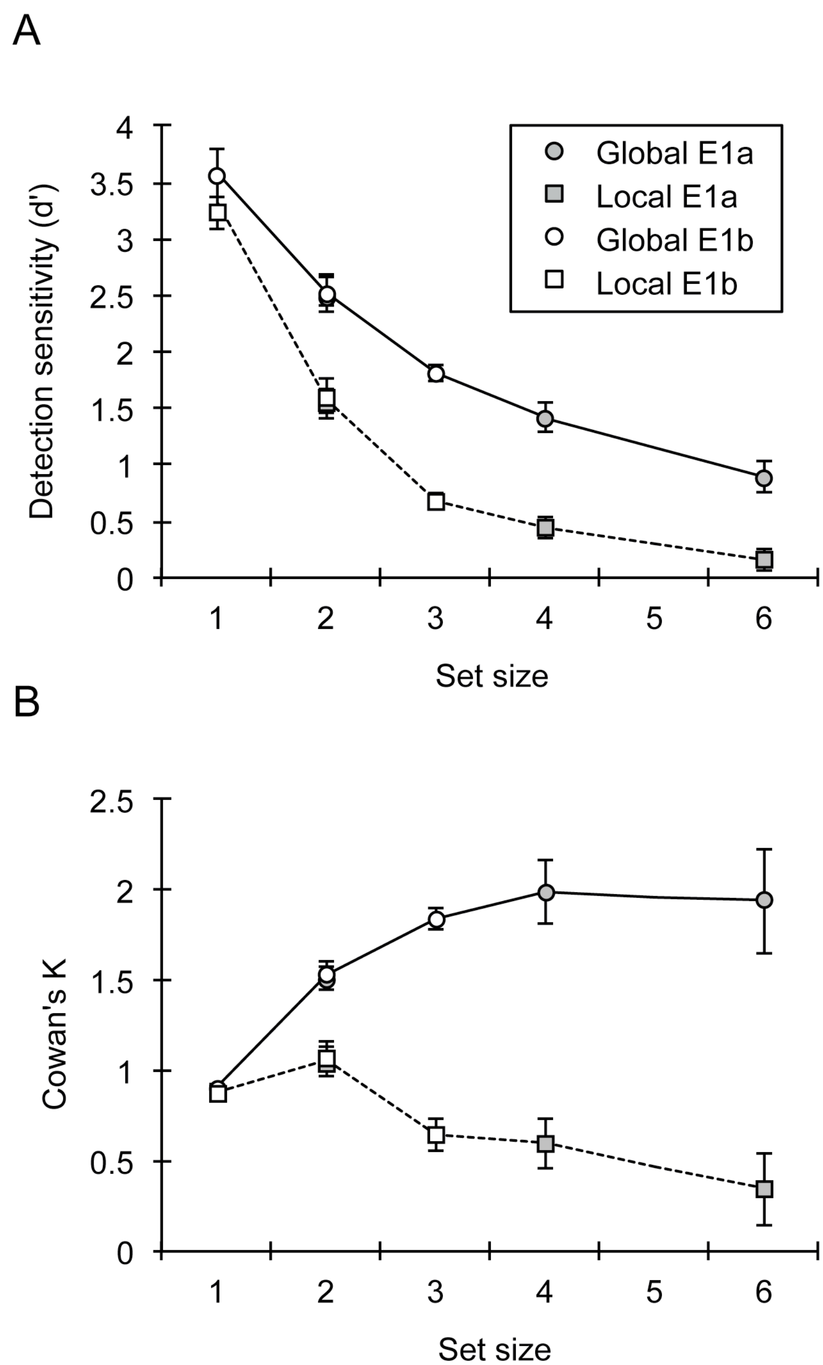

Figure 2. Mean change detection sensitivity, d-prime (A) and corresponding capacity estimates, K (B) in Experiment 1, presented as a function of set size for global and local changes (solid and dashed lines, respectively). Error bars represent ±1 SEM. Gray (set size: 2, 4, and 6) and white (set size: 1, 2, and 3) symbol colors represent values from Experiments 1A and 1B, respectively.

Results

Figure 2 shows mean d-primes and corresponding Cowan's K estimates as a function of set size, separately for global and local level changes. Results are combined across Experiments 1A (gray symbols) and 1B (white symbols).

Experiment 1A

D-prime. Figure 2A (gray symbols) presents the mean d-primes as a function of set size, separately for global and local level changes. Individual d-primes were subjected to a 2×3 repeated-measures analysis of variance (ANOVA) with the factors level (global, local) and set size (2, 4, and 6). This analysis revealed both main effects to be significant: level, $F(1,11) = 62.1$, $p < .001$, $\eta_p^2 = .85$, and set size, $F(2,22) = 97.03$, $p < .001$, $\eta_p^2 = .9$. Global changes were detected more efficiently than local changes (mean precedence effect in d-prime: 0.89), and d-primes decreased (by 1.5, $p < .001$) as set size increased (from 2 to 6). The two-way interaction was not significant, $F(2,22) = 2.17$, $p = .14$, $\eta_p^2 = .17$: global changes yielded a comparable decrease in d-primes across set sizes as local changes. To summarize, change detection sensitivity decreased with increasing set size, and sensitivity was much greater overall for global than for local changes. In addition, the decrease in d-prime across set size was comparable for both object levels, with the global precedence in change detection being rather stable across the number of to-be-remembered items[4].

[4] Even though response speed was not emphasized in the current task instructions, we nevertheless performed a comparable analysis on the median response times. The results from this analysis mirrored the pattern of the d-prime analysis, revealing significant main effects of global precedence and set size but no interaction. A similar pattern of results was also obtained in subsequent experiments. Because both accuracy-related measures and analyses of response speed yield similar effects, we can safely exclude the possibility of a speed-accuracy trade-off in our data.

Cowan's K. A subsequent analysis examined working memory capacity for global and local changes, that is, whether memory capacity for the global object level would be larger (i.e., indicative of "precedence") compared to that for the local level. To this end, a further 2×3 repeated-measures ANOVA was performed on individual Cowan's K estimates with the factors level (global, local) and set size (2, 4, and 6). This analysis revealed a significant main effect of level, $F(1,11) = 52.7$, $p < .001$, $\eta_p^2 = .83$, but no effect of set size, $F(2,22) = 0.54$, $p = .6$, $\eta_p^2 = .05$. The global object level was associated with a higher K estimate than the local level (1.97 vs. 0.47), with overall comparable K values across set size. The two-way interaction, was also significant, $F(2,22) = 17.6$, $p < .001$, $\eta_p^2 = .62$. Post-hoc comparisons showed that memory precedence for the global object level increased with set sizes larger than 2 (mean precedence in K: 0.46, 1.4, and 1.59, respectively for set size 2, 4, and 6, all $ts(11) > 5$, $ps < .001$). This result shows that, while the representation of the global object level has a greater memory capacity compared to the local level, the difference between levels becomes more pronounced with larger set sizes.

Experiment 1B

D-prime. In Experiment 1B, observers were presented with smaller set sizes (1, 2, and 3 items) of (global/local) hierarchical shapes. Despite this reduced set size, the results replicated those of Experiment 1A (see Figure 2A). Figure 2A (white symbols) presents the mean d-primes as a function of set size, separately for global and local changes. A 2×3 repeated-measures ANOVA of the individual d-primes, with the factors level (global, local) and set size (1, 2, and 3), revealed both main effects to be significant: level, $F(1,9) = 47.5$, $p < .001$, $\eta_p^2 = .84$, and set size, $F(2,18) = 94.5$, $p < .001$, $\eta_p^2 = .91$.

Detection sensitivity was again superior for global as compared to local changes (mean precedence effect in d-prime: 0.78), and sensitivity decreased (by 2.2) as set size increased from 1 to 3 (all ps $< .001$). Moreover, the two-way interaction was significant, $F(2,18) = 8.2$, $p = .003$, $\eta_p^2 = .48$: the benefit for global (relative to local) changes increased with larger memory arrays (mean precedence effects in d-prime were 0.27, 0.94, and 1.13 for set sizes 1, 2, and 3, respectively). To summarize, global change sensitivity was much bigger overall than local change sensitivity, and this difference increased with set size, again indicating that change detection operates more efficiently at the global object level.

Cowan's K. Examination of working memory capacity for global and local changes, by means of a level (global, local) × set size (1, 2, 3) ANOVA on the individual Cowan's K estimates, revealed both main effects to be significant: level, $F(1,9) = 65.04$, $p < .001$, $\eta_p^2 = .88$, and set size, $F(2,18) = 43.6$, $p < .001$, $\eta_p^2 = .83$. Capacity was higher for global than for local object levels (Ks of 1.84 and 0.64, respectively), and it increased (by 0.35) along larger set size. Moreover, the two-way interaction was significant, $F(2,18) = 61.7$, $p < .008$, $\eta_p^2 = .87$: the global precedence effect was reliable only for set sizes 2 and 3 (0.49 and 1.2, respectively; $t(9)\text{s} > 4.6$, ps $\leq .001$), but not for set size 1 (0.04, $t(9) < 1$, $p > .2$; see Figure 2B). In other words, with single objects, observers were capable of remembering both levels to a comparable extent; but, as set size increased from 2 to 3, global object representations were more reliable than local representations (with the local level in fact exhibiting a reduction of memory capacity from 2 to 3 to-be-remembered objects).

Discussion

Experiment 1 examined how the hierarchical (global/local) representation of objects affects vWM performance. Across both Experiments 1A and 1B, change detection performance was found to be superior (in terms of d-prime and K scores) at the global object level compared to the local level. Also, despite a general decrease in performance with increasing memory load, differential performance between object levels only emerged with larger set sizes: while change detection of a single object could be performed with comparable efficiency at both global and local levels, from set size 2 onwards, the global level showed a reliable benefit compared to the local level.

Analysis of the corresponding vWM capacity estimates revealed that observers could represent up to about two objects at the global level (see Figure 2B, which shows an asymptote at K=2 for the global level). At the local level, by contrast, observers could only represent up to one object; and, in fact, with larger set sizes (i.e., 3 or more objects), the K estimate for the local level is numerically reduced (see Figure 2B). Thus, when combining K values for the global and local levels, our results indicate that the memory capacity for hierarchical shapes taken together is at about 2.5 objects.

EXPERIMENT 2

Experiment 1 yielded a robust global precedence effect in vWM, which was evident both in measures of d-prime and in terms of memory capacity K – suggesting that the structure of to-be-remembered objects biases memory storage towards global object levels. Such a global bias might be related to the initial analysis of scenes in terms of their overall "gist" (Hollingworth & Henderson, 2003; Oliva, 2005; see also Conci & Müller, 2014). This representation of scene gist might in turn be related to summary statistics that represent sets of objects as a group or ensemble (e.g., of the average size of items; Alvarez, 2011; Brady & Alvarez, 2011). In other words, the global bias in vWM might, to some extent, be related to the analysis of global scene properties, which provides a summary representation of a given memory array.

To examine how such processes of scene analysis are related to global and local levels, in Experiment 2, we performed a change detection experiment with hierarchical shapes (essentially as in Experiment 1) but now varying the repetitions between to-be-remembered objects. Note that feature similarity (e.g., different degrees of redness) has been shown to affect object representations in vWM (e.g., Lin & Luck, 2008). Experiment 2 always presented a set size of four objects, and repetitions among items were manipulated systematically at the global and local levels: In the global repetition condition, always two (of the four) global-level objects were identical, and were presented alongside with four distinctive local shapes (Figure 4, left panel). By contrast, in the local repetition condition, two local-level objects were always identical and were presented with four distinctive global shapes (Figure 4, middle panel). Finally, a baseline condition presented four different hierarchical objects at both global and local levels (Figure 4, right panel). Accordingly, the manipulation of global versus local repetition allowed us to examine the relative influence of processing and resolving global and local levels of memorized objects.

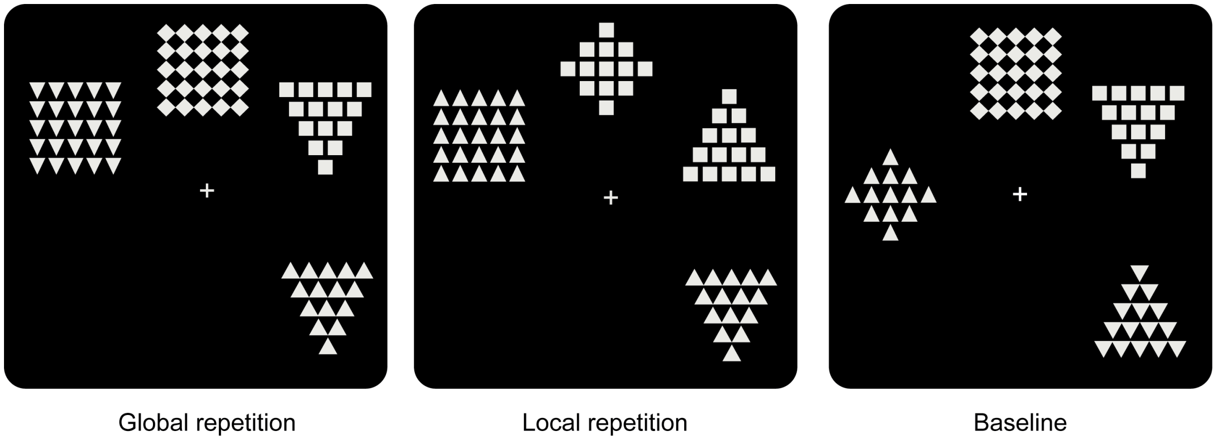

Figure 3. Examples of memory arrays with four hierarchical objects as presented in Experiment 2. In the global repetition condition (left), two pairs of globally repeated objects were presented, while the local shapes were all distinctive. The local repetition condition presented two pairs of locally repeated objects, while in turn the global shapes were all different from each other (middle). In the baseline condition (right), all objects were different from each other at both the global and local levels.

Methods

Experiment 2 was essentially identical to the previous experiments, except for a fixed set size of (always) 4 hierarchical shapes. Moreover, similarities between items were manipulated at three different levels: (i) global repetition, (ii) local repetition, and (iii) baseline. In the global and local repetition conditions, always two pairs of repeated global or local objects were presented together with four distinctive objects at the respectively other (local or global) level. In the baseline condition, there were four distinctive objects at both levels (see Figure 3 for examples). All three types of display (global repetition, local repetition, and baseline) were presented with equal probability in both change and no-change conditions. Participants initially completed 1 block of 48 practice trials to become familiar with the task. The formal experiment was divided into

12 blocks of 48 trials each. Twelve new observers (9 female; age from 21 to 31 years; mean age = 24.9 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour.

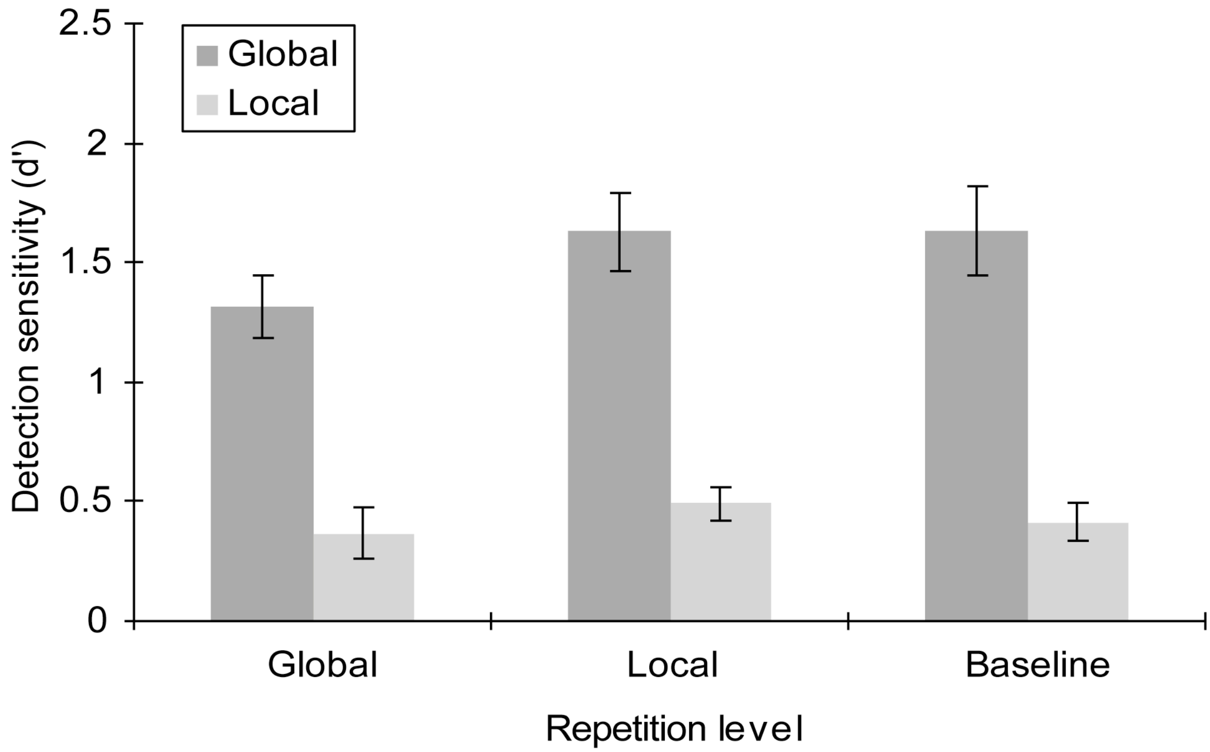

Figure 4. Mean change detection sensitivity (d-prime) in Experiment 2, presented as a function of the repetition level for both global and local changes. Error bars represent ±1 SEM.

Results

Figure 4 presents the mean d-primes as a function of the repetition level, separately for global and local object changes. Individual d-primes were subjected to a 2×3 repeated-measures ANOVA with the factors level (global, local) and repetition (global, local, and baseline). This analysis revealed both main effects to be significant:

level, $F(1,11) = 48.2$, $p < .001$, $\eta_p^2 = .81$, and repetition, $F(2,22) = 4.02$, $p < .001$, $\eta_p^2 = .27$. Global changes were detected more efficiently than local changes (mean precedence effect in d-prime: 1.1), comparable to the findings in Experiment 1. Moreover, post-hoc comparisons to decompose the main effect of repetition revealed that the mean d-prime (averaged over global- and local-change conditions) was reduced when comparing object repetitions for global relative to baseline and global relative to local conditions (mean d-prime differences were 0.18 and 0.22, respectively, $ps < .04$). By contrast, there was no difference in memory performance when comparing local repetition with baseline ($p = .99$). The two-way interaction was not significant, $F(2,22) = 2.5$, $p = .11$, $\eta_p^2 = .19$, indicating that global object repetition modulated detection of changes to a comparable extent at both global and local levels.

Discussion

Experiment 2 replicated Experiment 1 in showing an overall global bias in vWM. In addition, the results revealed an increase in object repetition at the global level to reduce mnemonic performance (affecting changes at both object levels in a similar manner). With multiple different to-be-memorized hierarchical shapes, observers' reports of a change of a given hierarchical shape displayed a cost deriving from the globally repeated objects in the memory array.

Our finding is compatible with the idea that vWM capacity reduces with competition between similar representations (Luck & Vogel, 2013). This notion would predict capacity (or, representational precision) to be lower when the to-be-remembered items are similar to each other. However, some of the available evidence appears to suggest that similarity-based perceptual grouping results in more accurate mnemonic representations, thus facilitating memory performance (Lin & Luck, 2008). It is possible that ensembles are represented in two different formats: either as ensemble averages, representing the average feature of to-be-remembered objects (as in simple feature estimation, see Bronfman, Brezis, Jacobson, & Usher, 2014); or, alternatively, as ensemble repetitions, representing the common features of presented objects, which is evident with high-order object representations in natural scenes. Accordingly, an ensemble may have distinct influences on the representation in vWM. For similar colored squares (Lin & Luck, 2008), ensemble *averages* of homogenous colors will lead to more precise mnemonic representation, whereas for globally repeated objects in the current experiment, ensemble *repetitions* likely provide an overall coarse representation of the entire display (Cohen, Dennett, & Kanwisher, 2016). Thus, our results suggest that ensemble repetitions of hierarchical memory representations interfere with mnemonic performance for individual objects. A potential alternative (not mutually exclusive) account for the reduction in performance with global object repetitions might be that observers were more likely to confuse, or "misbind" the memorized items when these were globally repeated (e.g., Bays, Wu, & Husain, 2011, for illusory bindings in vWM), illustrating once again the special role of the global scene layout in vWM, which would still be consistent with the account of global ensemble coding.

EXPERIMENT 3

Experiments 1 and 2 suggest that object structure influences vWM, such that more accurate information is retained from the global (relative to a local) object representation.

One potential reason for this difference in vWM performance, which results in an advantage for global relative to local object levels, could simply be related to stimulus encoding, that is, processes reflecting basic perceptual processing. Therefore, to ensure that observers had sufficient time to encode the stimuli presented, in Experiment 3, we varied the duration of the sample array, comparing a 300-ms presentation time of the sample display (as used in Experiments 1 and 2) with a longer presentation duration of 600 ms. Note that both presentation durations are consistent with typical encoding durations in standard vWM experiments (see Luck & Vogel, 2013, for a review), whereas a much shorter or longer presentation duration might additionally involve the recruitment of iconic memory or internal rehearsal of the memorized content (Baddeley, 1986), respectively. A previous estimate based on performance in a change detection task suggested that the rate of encoding objects into memory occurs at approximately 50 ms per item (Vogel, Woodman, & Luck, 2006). Thus, the longer stimulus duration provided substantially more time to perceptually encode the stimulus configurations, potentially leading to improved local-level detection performance if time were indeed a limiting factor.

Methods

Experiment 3 presented the baseline condition of Experiment 2 (with a fixed set size of four items, presenting distinctive shapes at both global and local levels), but with two different encoding durations of the memory array (presenting either a 300-ms or a 600-ms display duration, randomly intermixed within blocks). Participants initially completed 1 block of 48 practice trials, followed by 384 experimental trials. The experiment was divided into 8 blocks of 48 trials each. Eleven new observers (8 female;

age from 21 to 31 years; mean age = 22.5 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour.

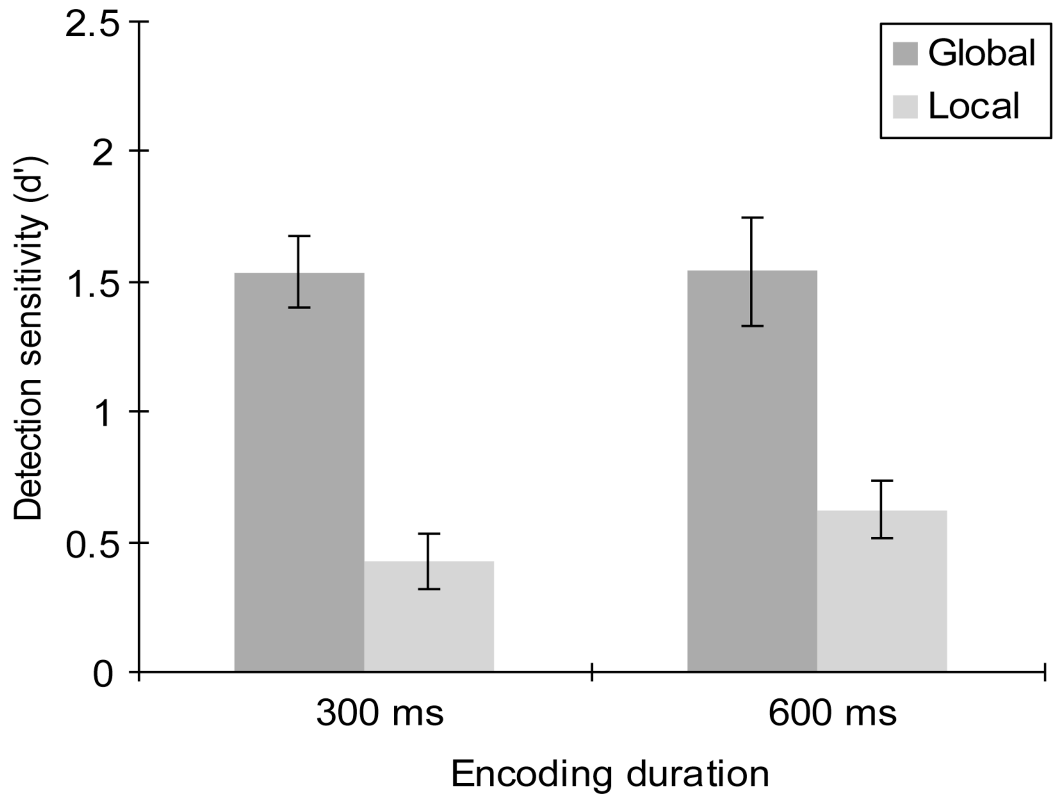

Figure 5. Mean change detection sensitivity (d-prime) in Experiment 3, presenting global and local changes for short and long encoding durations. Error bars represent ±1 SEM.

Results

Figure 5 presents the mean d-primes as a function of the encoding duration, separately for global- and local-level changes. Individual d-primes were subjected to a 2×2 repeated-measures ANOVA with the factors level (global, local) and encoding duration (300, 600 ms). This analysis revealed only a significant main effect of level: $F(1,10) = 37.5$, $p < .001$, $\eta_p^2 = .79$. Global changes were detected more efficiently than local changes (mean precedence effect in d-prime: 1.01). Importantly, however, there were no effects involving the factor encoding duration (ps > .14), suggesting that global precedence in working memory is not due to encoding limitations that might arise because of a too short duration of the presented memory array.

Discussion

We again replicated a robust global precedence effect in vWM as already seen in the previous experiments. Our results also demonstrate that performance was not significantly influenced by variations of the encoding duration: Reliable global precedence effects in vWM were found with both shorter (300 ms) and longer (600 ms) durations of the memory display. Following prolonged exposure of a stimulus array, mnemonic representations should be constrained by limits of storage only (Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011). Thus, our findings likely indicate that the global/local hierarchy primarily reflects limitations in storage capacity, rather than limitations in perceiving, that is, encoding of the presented objects.

EXPERIMENT 4

The present results demonstrate that structural relations in objects are represented in vWM, revealing a global precedence effect that might originate from limited mnemonic resources. However, a potential alternative explanation to account for these results might be that global object representations are to some extent more salient than corresponding local object representations. For instance, a global change comprises a change to a single, large (global) configuration while all local elements remain constant.

Conversely, in the case of a local change, many, small (local) shapes undergo a change (and the global configuration remains the same). Thus, differences in the number of the depicted changes, differences in relative size and/or the amount of crowding between change levels, may provide potential confounds that could alternatively (at least to some extent) account for our findings (see also Navon, 1981; Kimchi, 1992).

Experiment 4 was performed to test whether the detection of changes at global and local object levels differs when the changes (at varying levels) occur independently of each other. To this end, Experiment 4 essentially repeated the 300-ms presentation duration condition of Experiment 3, except that global- and local-change detections were now presented in separate sessions, such that participants only needed to memorize one task-relevant object level while ignoring the other level in the respective session. If the global bias is predominantly related to vWM maintenance of multiple hierarchical levels, we would predict that global memory precedence is substantially reduced when only one specific object level is relevant, while any residual differences might be taken to reflect additional influences that relate to stimulus saliency (e.g., relative size, or crowding).

Methods

Experiment 4 presented the 300-ms display duration condition of Experiment 3, but with global and local changes presented in separate halves of the experiment (counterbalanced across participants). A new group of eleven observers (5 female; age from 18 to 31 years; mean age = 20.4 years) performed the global and local change-detection tasks in two separate, consecutive sessions. Each session started with a practice block of 24 trials, followed by 96 experimental trials that were divided into 4 blocks of

24 trials each. All participants had normal or corrected-to-normal visual acuity, and received course credits or payment of 8 Euro per hour.

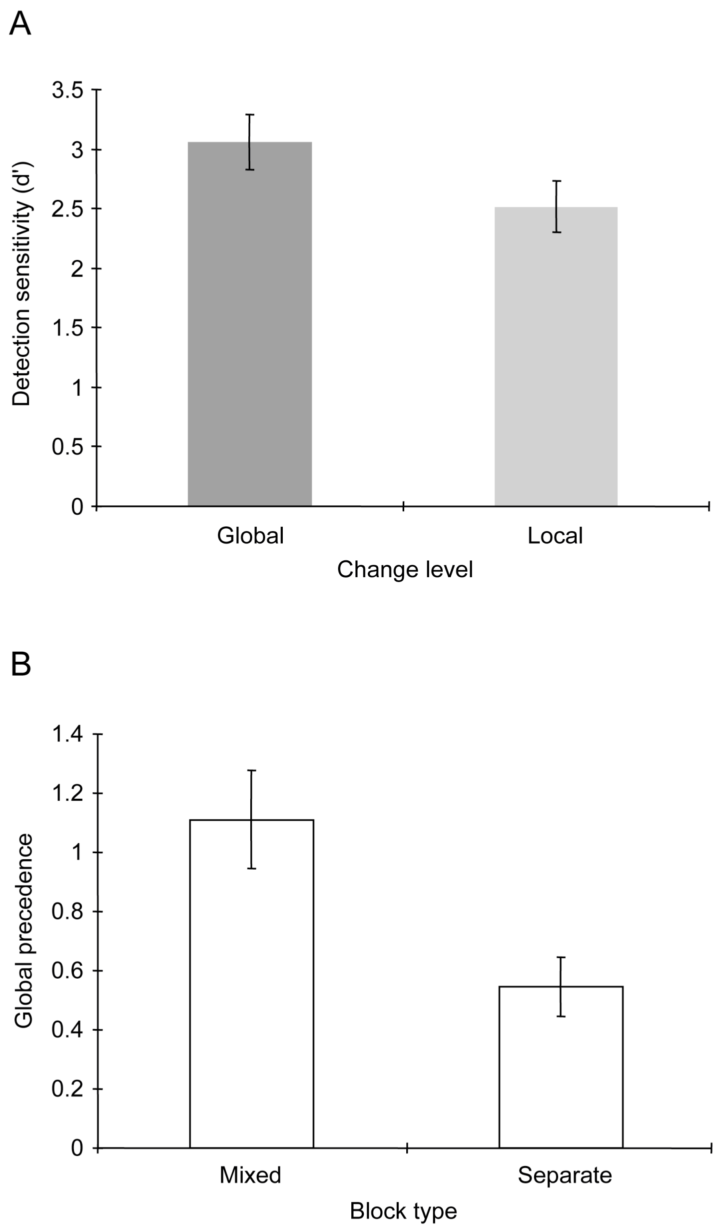

Figure 6. Mean change detection sensitivity, d-prime (A) in Experiment 4, with global and local changes presented in separate halves of the experiment. (B) Global precedence effect (global minus local d-primes) in Experiment 3 (300-ms encoding duration, in which global and local change trial were *mixed* within blocks) and in Experiment 4 (with global and local change trials presented in *separate* blocks). Error bars represent ±1 SEM.

Results

Figure 6A presents the d-prime values separately for global and local change detections. Paired t test revealed that d-primes for detecting a global change were significantly higher than local change detections: 3.06 vs. 2.52, $t(10) = 5.48$, $p < .001$, indicating that the global object level is still processed with priority even when only one hierarchical level is task-relevant during an entire half of the experiment.

In a next step, individual d-primes were subjected to a 2×2 mixed-design ANOVA with the within-subject factor level (global, local) and the between-subject factor block type (mixed [Experiment 3], separate [Experiment 4]). In the mixed block type, global and local change trials were presented in *mixed* order within a given block of trials (Experiment 3, 300-ms encoding duration), while for the *separate* block type global and local changes were presented in separate blocks (Experiment 4). The results from this analysis revealed significant main effects of level: $F(1,20) = 73.1$, $p < .001$, $\eta_p^2 = .78$ and of block type: $F(1,20) = 57.1$, $p < .001$, $\eta_p^2 = .74$. The main effect of level simply depicts the above mentioned global precedence effect (i.e., the difference in d-prime between global and local change detections), which was present with both mixed and separate presentations. Moreover, mixed, relative to separate, block types led to substantially reduced change detection sensitivities (d-primes: 0.98 vs. 2.8, see Figures 5 and 6A, respectively). Importantly though, our analysis also revealed a significant level by block type interaction, $F(1,20) = 8.6$, $p < .009$, $\eta_p^2 = .30$, showing that the global precedence effect in the current experiment (*separate* block type) is significantly reduced as compared to Experiment 3 (*mixed* block type; global precedence in d-primes: 0.54 vs.

1.11, respectively; see Figure 6B). This finding suggests that global precedence is substantially reduced with the number of task-relevant object levels.

As described above, monitoring only one task-relevant object level in Experiment 4 (as opposed to both object levels in previous experiments) led to a substantial increase in the overall detection sensitivity. However, this overall difference in performance might additionally have influenced the size of the global precedence effect. We therefore calculated a *relative* global precedence score that normalizes the difference between global and local change detections relative to the default, global level of performance (i.e., [global – local]/global d-primes). This relative difference score revealed that global precedence in Experiment 4 modulated detection performance by 17.6%, as compared to a much larger difference of 72.1% in Experiment 3.

Discussion

Without having to remember both global and local object levels simultaneously, the performance for both global- and local-change detections was found to be significantly enhanced. Nevertheless, we still obtained a reliable (but reduced) global precedence effect that showed a bias in processing global-level object information. This suggests that a small, but yet reliable part of the global precedence effect might be attributed to differences in saliency between global and local levels of representation (relative global precedence effect of 17.6%). Importantly, however, the larger part of the global precedence effect (i.e., 72.1%) appears to be related to hierarchical-level differences as they are maintained in vWM.

General discussion

Studies that investigate vWM typically present relatively simple features or objects that are assumed to be represented independently of each other. However, recent evidence suggests that, rather than representing only individual items, vWM also provides a structural representation, that is: ensemble statistics relating to aspects of the "gist" of the presented scene (Alvarez, 2011; Brady & Alvarez, 2011; Brady et al., 2011). In an attempt to elaborate this notion, the aim of the present study was to investigate how hierarchical relations within and across objects are represented in vWM. Our experiments yielded four main results, namely: (i) vWM representations are organized in a hierarchically structured fashion reflecting the global/local structure of the perceptual input; (ii) repetition between items particularly at a global object level gives rise to vWM capacity detriments; (iii) global precedence effects in the current change detection paradigm primarily reflect the way items are stored during the retention phase; and (iv) this global benefit mainly reflects the globality of memory itself, and can only be partially attributed to differences in saliency across object levels.

Beyond objects and features: hierarchical representations in vWM

To explore the hierarchical structure within given object representations in vWM, we introduced a change detection task with hierarchical, global/local shapes, in which a change occurred either at the global or the local level of a given target object. We found a robust pattern of global precedence: Measures of memory performance (d-prime, Cowan's K) for the global object level revealed higher sensitivity and larger capacity compared to measures for the local level. Importantly, this global memory bias increased significantly with larger set sizes, reaching an asymptote at a K value of approximately 2; the corresponding local memory capacity was overall smaller, with a K value of about 1 that decreased to 0.5 with larger memory arrays. Thus, limited memory resources are distributed asymmetrically across object levels: with an increased memory load, the global benefit becomes larger. Moreover, the total amount of visual information maintained in vWM turned out to be rather stable, with a maximum capacity of around K=2.5 items when global and local levels are pooled together (see Hardman & Cowan, 2015, for a comparable procedure). This is consistent with the view that, overall, there is a limited amount of visual information that can be held in vWM (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Ma, Husain, & Bays, 2014). Our study extends this idea by showing that the distribution of limited memory resources is hierarchically organized within a given object, reflecting the inherent, to-be-remembered visual structure.

Any measure of memory capacity is meant to estimate its underlying “units”, where what exactly counts as the proper unit depends on the structure of the represented information. For instance, it has been suggested that objects are represented as separate visual features that are stored in independent “channels”, each with their own capacity limitation (e.g., color or orientation; Magnussen, Greenlee, & Thomas, 1996), or in terms of integrated object representations (Luck & Vogel, 1997; Vogel et al., 2001). Recent evidence suggests that there are significant benefits to remembering multiple features of a single object compared to the same set of features distributed across multiple objects (Fougnie, Asplund, & Marois, 2010; Olson & Jiang, 2002). For instance, it is easier to store five different colors and five orientations that define the same five objects than to store the same 10 features for separate objects (Fougnie, Cormiea, & Alvarez, 2013). The finding that vWM improves with fewer discrete objects (Olson & Jiang, 2002; Xu, 2002) has been taken to suggest that the representations that underlie vWM are object-based (Luck & Vogel, 1997; Vogel et al., 2001). According to this notion, vWM can store a small, fixed number of objects and integrate multiple features into a single object representation. However, there is also evidence showing that vWM representations are, in fact, not purely object-based. For instance, having to remember more features engenders significant costs in terms of the fidelity of each feature representation (Fougnie et al., 2010). Moreover, attending to one relevant feature reduces the mnemonic precision of a second task-irrelevant feature of the same object, such that memory precision for multiple features of an object may vary as a function of the attentional engagement devoted to each feature (Shin & Ma, 2016; Swan, Collins, & Wyble, 2016). Such results indicate that what counts as the right "unit" in vWM is neither a fully integrated object representation nor a collection of independent features. Instead, vWM units appear to encompass rather flexible organizational principles. In light of the current results, the actual unit in vWM would appear to reflect a hierarchical structure, with the global-level representations of this "unit" being prioritized for vWM storage, while less resources are assigned to local-level object representations.

Beyond slots vs. resources: hierarchical ensembles in vWM

To further investigate the relational structure between object representations, we tested the role of inter-item repetition in Experiment 2, in which pairs of hierarchical objects were similar either at the global or at a local level, compared to a baseline condition in which none of the objects were similar to other items at the global/local object levels. While replicating the basic pattern established in Experiment 1, we further observed that repetition at the global level interfered with the detection of both global and local changes. Local repetition, by contrast, appeared to have no influence (with performance comparable to the baseline level).

Prominent theories of vWM have proposed that memory limitations arise entirely from the availability of some limited resource that is either quantized into slots (Zhang & Luck, 2008) or continuously divisible (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Wilken & Ma, 2004). These notions, and the supporting studies, leave open the question of how contextual or ensemble representations interact with representations of individual items in vWM. A potential account in this regard assumes that the representation of ensemble statistics could take up space in memory that could otherwise be used to represent information about individual items (Cohen, Dennett, & Kanwisher, 2016). In agreement with this view, our findings show that vWM representations are biased towards the global level and interference arises in particular among similar global-object representations. Restated, a global ensemble representation of the entire display impairs (via repetition) both global and local memory representations of individual items, illustrating that the different hierarchical object levels of vWM representations are linked and dependent on each other, rather than being maintained independently (Fougnie et al., 2013).

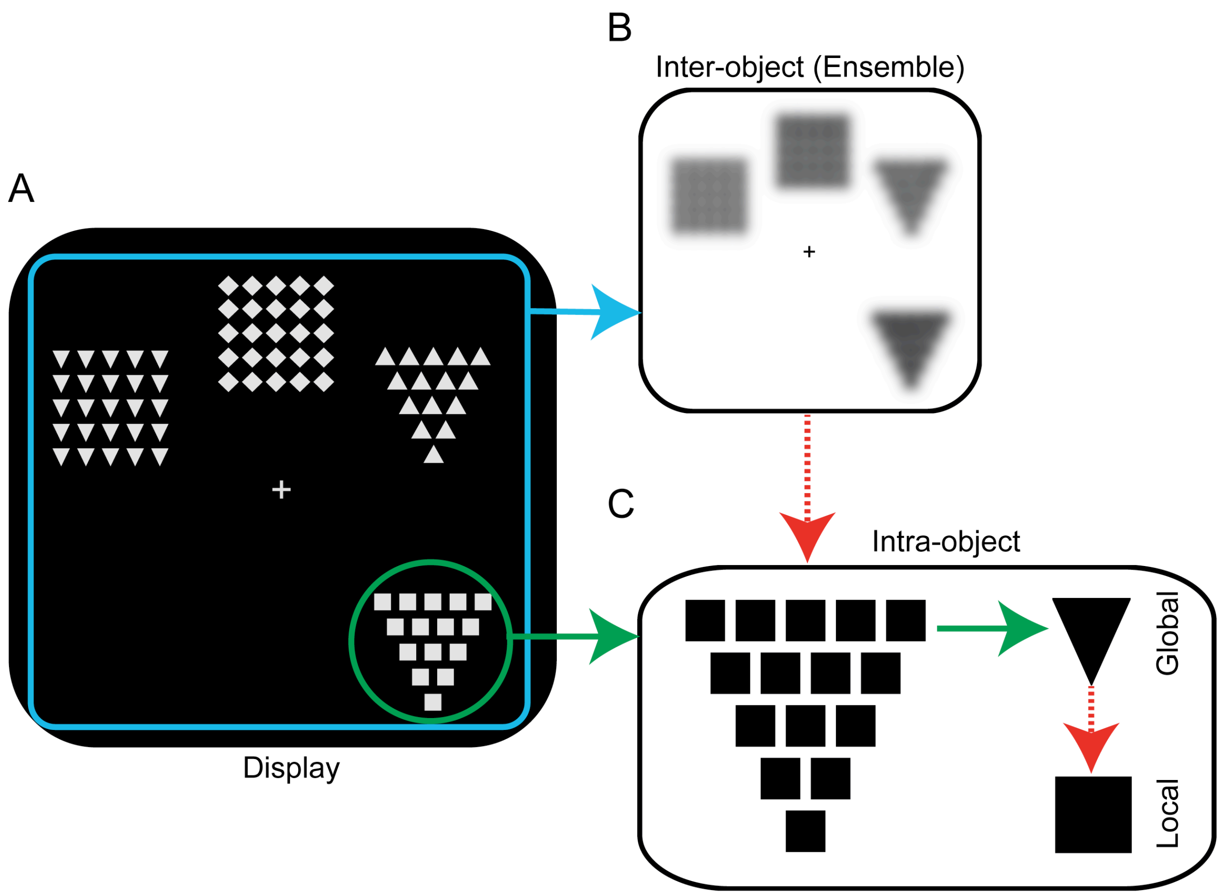

Figure 7. Example memory array (A) and a schematic model of hierarchically structured representations in vWM, designed to illustrate the interaction between inter-object ensemble representations (B) and storage of individual items and their intra-object (global/local) relations (C). At an inter-object level (blue arrows), the global display characteristics are encoded into an ensemble representation (B). In addition, individual items are stored at an intra-object level (C), with separate representations of the global and local levels of a given to-be-remembered object (green arrows). The model also incorporates inhibitory links between different levels (red, dashed arrows), reflecting the reduction of mnemonic precision from the inter- to the intra-object level (e.g., for multiple, repeated items), and from the global to local mnemonic representations, to account for our finding of a robust global precedence in vWM.

A theoretical model of hierarchical working memory

Recently, a hierarchical feature-bundle model has been proposed with the aim to integrate both object- and feature-based effects in vWM (Brady et al., 2011). According to this model, each unit of vWM is a hierarchically structured feature bundle, consisting of an integrated object representation at the top level and individual features represented at a lower level. The main idea of the model is that a unit in vWM is determined by the top level representation of an integrated object, while, at the same time, the lower level elemental feature of an object can be accessed by means of top-to-bottom decomposition within a given bundle.

In light of the current findings and in general agreement with the feature-bundle model, we propose that various levels of representation of a set of items in vWM are likewise stored in a hierarchically organized fashion. In a process paralleling how people memorize real scenes (Oliva, 2005), observers might represent summary statistics (i.e., the "gist") of the entire memory display in addition to information about each specific item. Each item in turn has its own global and local representation, which are also maintained in terms of a hierarchically structured representation with global and local object levels. This hierarchical storage format within and across individual items would permit observers to represent not only the individual identity of the to-be-remembered items, but also the structural relations (global/local) across the display layout. However, storing the various object levels plus the global ensemble might come at an overall cost, which is reflected in the overall low vWM capacity of about 2.5 items (as compared to capacity estimates of about 4 with displays that present relatively simple colored squares; e.g., Luck & Vogel, 1997).

A schematic model to accommodate the various levels of representation is depicted in Figure 7 (on the basis of Brady et al., 2011). At an intra-object level, the vWM representation of a given individual item consists of two hierarchically structured layers, with a global object representation being stored at the top level and the corresponding local representation at the bottom level (Figure 7C). In this view, representational units in vWM are conceived as being hierarchically structured across global/local levels, thus reflecting the global and local object properties of the stimulus input. The two layers of the representation are not equal; rather, the global level receives a representational bias. Consequently, the global object representation would be available at the top-level unit, whereas information concerning the local object would be stored at the lower hierarchical level. Critically, the top-level unit receives priority such that the majority of the available mnemonic resources are used to maintain the global-level representation. By contrast, the subordinate, lower-level representation of local object information would receive only a smaller amount, that is, the remaining resources (as indicated by the dashed arrow) – in line with the current observation of global precedence in change detection. Of note, in the current experiments, an asymmetry in performance primarily results from global and local levels reflecting the inherent hierarchical structure of a given object. However, comparable differences in processing can also be observed in non-hierarchical objects with multiple features where an asymmetry in mnemonic performance results from varying levels of attentional engagement (Shin & Ma, 2016; Swan et al., 2016).

Moving beyond individual vWM representations, at an inter-object level, the entire display layout is retained in particular with reference to the global object characteristics; as a result, globally similar items are merged into a single ensemble representation (Figure 7B). This ensemble representation essentially reflects the observed global bias overall, that is, global precedence and repetition effects might both arise from the global-level representation – enhancing the global objects but also causing interference as revealed by the impaired mnemonic representation of the entire individual object (dashed arrow).

One counterintuitive prediction of the model is that if the objects all had the exact same shape at the global level, performance should actually be worse than if there were several distinct global shapes presented. This is obviously not a likely experimental outcome as the task should be much easier when shapes at a given level are the same. A potential explanation might be that for homogeneous global display representations, redundant information presented at all item locations obviates the need to encode separate units of information from each location, but simply requires memorizing the repeated structure overall, thus reducing vWM load. A potential constraint of our schematic model therefore is that it can only account for competition between several repeated global object representations in heterogeneous displays (as in the example of Figure 7B), which leads to an overall impairment of individual (item) memory representation by means of ensemble coding.

In brief, this schematic model extends previous vWM models by taking into account the hierarchical relations both within and across objects, thus (to some extent) reflecting the typical structures in our natural environment with both representations of the overall scene layout and the more detailed object information. In this view, items are stored across three layers of representation, from the overall scene layout to the fine, detailed object information at the local level. Importantly, the model encompasses interference from top to bottom layers to illustrate the hierarchical organization of visual information, which in general assigns priority to the global level.

Conclusion

The present study reveals a functional connection between the representation of objects at varying hierarchical levels and the organization of vWM. Object representations in vWM are, by default, biased towards the global level, with the global bias existing across varying encoding durations and mainly reflecting the globality of memory itself. This suggests that global precedence in change detection primarily originates from hierarchically structured representations that are held in vWM. Memory performance is also influenced by the ensemble structure of the displays, that is: the interference of repetition among objects at the global level manifests in terms of impaired mnemonic representations for both global and local object levels. Together, our findings challenge models that propose that a fixed number of independent objects can be remembered regardless of the presented object structure. Instead, our results support a more flexible account that emphasizes the role for hierarchically structured representations in vWM.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106-111. doi: 10.1111/j.0963-7214.2004.01502006.x

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, 106*(18), 7345-7350. doi: 10.1073/pnas.0808981106

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision, 11*(10), 6-6. doi: 10.1167/11.10.6

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851-854. doi: 10.1126/science.1158023

Bays P. M. Wu E. Y. Husain M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia, 49*(6), 1622-1631. doi: 10.1016/j.neuropsychologia.2010.12.023

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384-392. doi: 10.1177/0956797610397956

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 4-4. doi: 10.1167/11.5.4

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

Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We See More Than We Can Report: "Cost Free" Color Phenomenality Outside Focal Attention. *Psychological Science, 25*(7), 1394-1403. doi: 10.1177/0956797614532656

Chen, S., Müller, H. J., & Conci, M. (2016). Amodal completion in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1344-1353. doi: 10.1037/xhp0000231

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the Bandwidth of Perceptual Experience? *Trends in Cognitive Sciences, 20*(5), 324-335. doi: 10.1016/j.tics.2016.03.006

Conci, M., & Müller, H. J. (2014). Global scene layout modulates contextual learning in change detection. *Frontiers in Psychology, 5*, 89. doi: 10.3389/fpsyg.2014.00089

Conci, M., Müller, H. J., & Elliott, M. A. (2007a). The contrasting impact of global and local object attributes on Kanizsa figure detection. *Perception & Psychophysics, 69*(8), 1278-1294. doi: 10.3758/BF03192945

Conci, M., Müller, H. J., & Elliott, M. A. (2007b). Closure of salient regions determines search for a collinear target configuration. *Perception & Psychophysics, 69*(1), 32-47. doi: 10.3758/BF03194451

Conci, M., Töllner, T., Leszczynski, M., & Müller, H. J. (2011). The time-course of global and local attentional guidance in Kanizsa-figure detection. *Neuropsychologia, 49*(9), 2456-2464. doi: 10.1016/j.neuropsychologia.2011.04.023

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114. doi: 10.1017/S0140525X01003922

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision, 10*(12), 27-27. doi: 10.1167/10.12.27

Fougnie, D., Cormiea, S., & Alvarez, G. A. (2013). Object-based benefits without object-based representations. *Journal of Experimental Psychology: General, 142*(3), 621- 626. doi:10.1037/a0030300

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72*(7), 1825-1838. doi: 10.3758/APP.72.7.1825

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition, 7*(1-3), 213-235. doi: 10.1080/135062800394775

Hollingworth, A., & Henderson, J. M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. *Memory & Cognition, 31*(6), 930-940. doi: 10.3758/BF03196446

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 683-702. doi: 10.1037/0278-7393.26.3.683

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance, 8*(4), 521-535. doi: 10.1037/0096-1523.8.4.521

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112*(1), 24-38. doi: 10.1037/0033-2909.112.1.24

Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1211-1222. doi: 10.1037/0278-7393.27.5.1211

Lin, P.-H., & Luck, S. J. (2008). The influence of repetition on visual working memory representations. *Visual Cognition, 17*, 356-372. doi: 10.1080/13506280701766313

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281. doi: 10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391-400. doi: 10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17*(3), 347-356. doi: 10.1038/nn.3655

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide (2nd Edition)*. Cambridge, UK: Cambridge University Press.

Magnussen, S., Greenlee, M. W., Thomas, J. P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 202-212. doi: 10.1037/0096-1523.22.1.202

Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia, 36*(6), 513-520. doi: 10.1016/S0028-3932(97)00148-6

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353-383. doi:10.1016/0010-0285(77)90012-3

Navon, D. (1981). The forest revisited: More on global precedence. *Psychological Research, 43*(1), 1-32. doi: 10.1007/BF00309635

Nie, Q.-Y., Maurer, M., Müller, H. J., & Conci, M. (2016). Inhibition drives configural superiority of illusory Gestalt: Combined behavioral and drift–diffusion model evidence. *Cognition, 150*, 150-162. doi: 10.1016/j.cognition.2016.02.007

Nie, Q.-Y., Müller, H. J., & Conci, M. (2016). Searching for forest or trees: Attentional zooming and level-specific memory in hierarchical objects. *(submitted for publication)*

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics, 64*(7), 1055-1067. doi: 10.3758/BF03194756

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

Shin, H., & Ma,W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision, 16*(5):10, 1-8, doi:10.1167/16.5.10

Swan, G., Collins, J., & Wyble, B. (2016). Memory for a single object has differently variable precisions for relevant and irrelevant features. *Journal of Vision, 16*(3):32, 1-12, doi:10.1167/16.3.32

Victor, J. D., & Conte, M. M. (2004). Visual working memory for image statistics. *Vision Research, 44*(6), 541-556. doi: 10.1016/j.visres.2003.11.001

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92-114. doi: 10.1037/0096-1523.27.1.92

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1436-1451. doi: 10.1037/0096-1523.32.6.1436

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

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218-1252. doi: 10.1037/a0029334

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12), 11-11. doi: 10.1167/4.12.11

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10*(1), 80-87. doi: 10.3758/BF03196470

Xu, Y. (2006). Encoding objects in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics, 68*, 815-828. doi: 10.3758/BF03193704

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences, 104*(47), 18766-18771. doi: 10.1073/pnas.0705618104

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235. doi: 10.1038/nature06860

Chapter V

The structure of visual working memory representations adapts to task demands

Abstract

Natural scenes consist of multiple hierarchical levels, though typically more global levels are prioritized over more local levels. A global-object benefit has also been revealed in visual working memory (vWM), but it is unknown whether such structured representations can be adapted flexibly according to task demands. To test the flexibility of object structure in vWM, we performed two experiments that presented to-be-remembered hierarchical configurations with global and local orientation information in change-detection (Experiment 1) and continuous-report (Experiment 2) tasks while systematically varying the amount of required memory precision. Our results revealed a consistent influence of precision demands on the structure of memory: the typical global-object benefit was reduced and eventually reversed into a local-object benefit with the degree to which object details were to be remembered. These findings indicate that structured memory representations are flexibly adjusted according to task demands, thus challenging accounts that assume fixed representations in vWM.

Introduction

Natural environments consist of cluttered arrays of objects with hierarchical structures. For example, a forest has trees, and the trees in turn have branches and leaves, illustrating that natural scenes tend to be organized in a hierarchically structured fashion. Accordingly, scenes might be represented at varying levels of the visual hierarchy, with global representations (e.g., the forest) at the top and more local representations (e.g., trees) towards the bottom (Kimchi, 1992). In order to meaningfully interact with such hierarchical visual environments, remembering objects in scenes requires registration not only of the individual elements, but also of the structural relations among them, that is: their part-whole organization. To achieve this, our visual system interprets the visual input by integrating (often fragmentary) local elements into global, 'holistic' percepts.

Likewise, there is evidence that visual-working-memory (vWM) representations are hierarchically structured, maintaining part-whole relations (see Brady, Konkle, & Alvarez, 2011, for a review). For instance, Nie, Müller, and Conci (2017) developed a hierarchical variant of the change detection task (Luck & Vogel, 1997) to investigate how global/local object levels are represented in vWM. On each trial, a variable number of hierarchical shapes (e.g., a global triangle composed of several local squares) were presented in a memory array, followed by a test probe that appeared after a brief delay. Observers were required to memorize all objects with their respective hierarchical levels, and to indicate whether or not the test probe differed (at either the global or the local level) from the object at the respective location in the memory array (observers had to issue a simple change/no-change [two-alternative forced-choice] response). The results revealed that global changes were more efficiently detected than local changes, suggesting that global object levels are prioritized over local levels in vWM, with considerably fewer mnemonic resources being devoted to the storage of local as compared to global object levels.

This pattern of global precedence in vWM is in line with a number of studies that investigated global/local processing in classic attention tasks (e.g., Navon, 1977; Kimchi, 1994; Conci, Müller, & Elliott, 2007; Wagemans, Elder, et al., 2012). However, prioritization of the global level of to-be-memorized objects may not reflect a rigid bias of encoding in vWM; rather, the structure of a given vWM representation might depend on the current task goals. For instance, Machizawa, Goh, and Driver (2012) examined whether the precision of vWM representations can vary with the expected magnitude of a given to-be-detected change. In their study, observers performed a change detection task with orientation stimuli, where the color of the to-be-memorized items was informative about the magnitude of a to-be-expected orientation change. The results revealed memory precision to be higher when observers expected a fine orientation change, as compared to a coarse change (at least when the memory load was low), indicating that both the number and precision of memorized items can be constrained by top-down task goals. However, to the best of our knowledge, there is no evidence of how hierarchically structured representations in vWM adapt to changing task demands. One – reasonable – hypothesis would be that the more a task requires fine (local) detail to be retained, the lower the benefit for the global object level. On this view, there is a trade-off such that maintaining finer object details requires a higher proportion of mnemonic resources to be devoted to representing the local level, which comes at the expense of resources available for representing the global level. The present study was designed to examine for such a task-dependent trade-off in the structure of vWM representations, by using change-detection (Luck & Vogel, 1997) and continuous-report (Zhang & Luck, 2008) tasks applied to hierarchical orientation configurations, where changes in the task would systematically alter the degree of the required precision at both global and local levels.

Of note in this context, the type of compound letters (Navon, 1977) and composite shapes (Kimchi & Palmer, 1982) often used to examine the global/local structure of visual perception cannot be used to probe the fidelity of vWM representations because they only allow for discrete changes (e.g., from a triangle to a square at either global or local levels). To overcome this limitation, we developed a novel, textured stimulus that permits continuous changes to be implemented at both global and local levels, specifically, a global ellipse composed of local oriented lines (Figure 1A; see also Kimchi, 1994, for comparable hierarchical stimuli). For these configurations, the global orientation is entirely defined by the (boundary contour connecting the terminations of the) local oriented lines; and the global and local orientations can vary independently of each other in continuous feature space (i.e., both can vary independently from -90° to +90°). These orientation stimuli thus possess a hierarchical structure that permits the fidelity of their vWM representations to be probed at both object levels.

EXPERIMENT 1

Experiment 1 investigated the flexibility of structured representations in vWM using a variant of the change-detection task with hierarchical configurations that depict both global and local orientation information (see Figure 1A). In two separate phases of the experiment, observers were presented with different magnitudes of change (large/60°, or small/20°). On the basis of previous findings (Nie et al., 2017), we expected greater detection accuracy for global than for local orientation changes. In addition, change magnitude was predicted to modulate the global benefit, with observers displaying stronger global precedence when having to detect large- as compared to small-magnitude changes (see above).

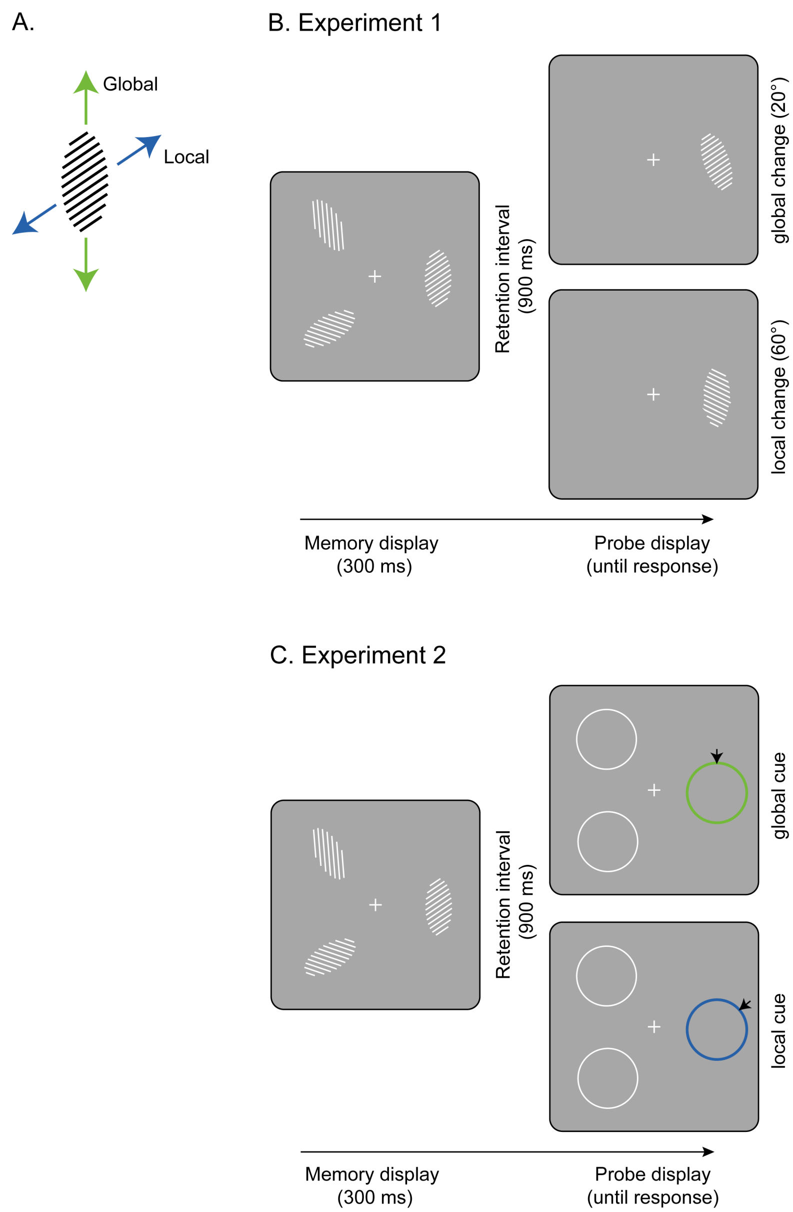

Figure 1. Stimuli and example trial sequence in Experiments 1 and 2. (A) Example of a hierarchically oriented ellipse with global and local orientations (as indicated by the

green and blue arrows, respectively). (B) In Experiment 1, a trial started with a memory display containing one or three hierarchical objects of varying global/local orientations. After a blank retention interval, a probe display was presented which, on change trials, depicted either a small (20°, top panel) or large (60°, bottom panel) change in orientation at either a global (top) or a local (bottom) object level. (C) In Experiment 2, the trial sequence was essentially the same, except that the probe display contained circular placeholders (C, bottom panel), and participants were asked to reproduce either the global or the local orientation of the target item in the memory array by clicking on the corresponding position of the colored wheel (black arrows). Different colors (blue, green) were used to cue the global and local orientation, respectively.

Method

Participants. Eighteen observers (6 male; age range 19 to 32 years, mean age = 25.3 years; all reporting normal or corrected-to-normal visual acuity) participated in Experiment 1. Participants received course credits or payment of 8 Euro per hour, and they provided (prior) written informed consent to the study procedure, which was approved by the local ethics committee, in accordance with the Declaration of Helsinki.

The choice of sample size in the present study was based on previous studies on global/local structure in selective attention and visual memory (Nie, Maurer, Müller, & Conci, 2016; Nie et al., 2017); we aimed for 75% power to detect an effect size of approximately 0.66 with an alpha level of .05, for both experiments.

Apparatus and Stimuli. The experiment was controlled by a Matlab program, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were oriented ellipses (0.75° × 2.0° of visual angle in size) composed of oriented lines (thickness: 1 pixel; 5 lines/°; see Figure 1A), which were presented in white (27.1 cd/m^2) against a gray (8.5 cd/m^2) background on an LCD monitor screen placed at a viewing distance of approximately 57 cm.

Memory arrays consisted of 1 or 3 hierarchical stimuli presented on an imaginary circle of 4° radius around the central fixation. Their positions were randomly selected, the only restriction being that neighboring stimuli were separated by at least 2.8° (center-to-center distance). The test display consisted of a single probed object (which, on trials with three to-be-remembered stimuli, was randomly selected from the memory array). Figure 1 shows an example display with a set size of 3, and possible variants of test probes, which illustrate global and local changes (Figure 1B, top and bottom panels, respectively) at two different change magnitudes.

Trial Sequence. Each trial started with the presentation of a central fixation cross (800-1600 ms; randomly jittered), followed by the memory display (300 ms), a blank retention interval (900 ms), and then a test probe that presented one item at a randomly chosen location from the preceding memory array. The fixation cross was visible throughout the entire trial sequence. The task was to decide whether the test probe was the same (at both the global and local levels) or different (with a change at either the global or the local level) relative to the item that had been previously presented at the same location in the memory array. The probe item remained on-screen until a response was recorded, or until time-out after 5,000 ms. Participants were instructed to respond as accurately as possible (without time pressure). In case of an erroneous or late response, feedback was provided by changing the color of the fixation cross to red or blue, respectively, for 1000 ms, before the next trial started.

Design and Procedure. The experiment was divided into two consecutive phases that either presented a variant with a large (60° of memory-probe transition) or one with a small (20°) change magnitude, with the order of presentation counterbalanced across observers. A three-factors within-subjects design was used for both phases. The independent variables were change (present vs. absent), level (global vs. local), and set size (1 vs. 3 items). All conditions of this design were equally probable and were presented in a random order.

Participants were comfortably seated in a dimly lit room. Each phase (large vs. small changes) started with 24 practice trials for participants to become familiar with the task and with the current change magnitude (large, or small), followed by 320 experimental trials presented in 8 blocks of 40 trials each, resulting in 40 trials for each factorial combination.

Results

Figure 2 shows the mean percentage of change responses (hits and false alarms) for large (A) and small (B) change magnitudes as a function of (memory array) set size, separately for global- and local-level changes. Individual hit rates were subjected to a 2 (change magnitude: large, small) × 2 (level: global, local) × 2 (set size: 1, 3) repeated-measures analysis of variance (ANOVA), which revealed all three main effects to be significant: change magnitude, $F(1,17) = 86.05$, $p < .001$, $\eta_p^2 = .84$; level, $F(1,17) = 238.8$, $p < .001$, $\eta_p^2 = .94$; and set size, $F(1,17) = 56.23$, $p < .001$, $\eta_p^2 = .77$. Large changes were detected more accurately than small changes (mean effect of change magnitude on accuracy: 0.16); global changes were detected more accurately than local changes (mean precedence effect on accuracy: 0.22); and accuracies decreased overall (by 0.2) from set size 1 to set size 3. In addition, the three-way interaction was significant, $F(1,17) = 42.96$, $p < .001$, $\eta_p^2 = .72$. Post-hoc comparisons showed that memory precedence for global orientation increased with set size for large changes (global-precedence effect at set sizes 1 vs. 3: 0.15 vs. 0.3, $t(17) = -3.67$, $p = .002$, $d_z = -0.87$, see Figure 2A), whereas it decrased for small changes (global-precedence effect at set sizes 1 vs. 3: 0.27 vs. 0.15, $t(17) = 3.96$, $p = .001$, $d_z = 0.93$, see Figure 2B). No other significant effects were obtained, Fs < 1, ps > .36.

The global-precedence pattern for the large changes effectively mirrors our previous findings, demonstrating a large and reliable global benefit that increases with set size (Nie et al., 2017). For small changes, by contrast, global precedence decreases with increasing memory load. At first glance, the opposite effects of (large vs. small) change magnitude on global precedence with increasing set size may appear surprising. However, there might be a simple, purely statistical explanation for this: Hit rates were at ceiling (96% hits) for global, large changes at set size 1[5]. As a result, the global-precedence effect for large changes at set size 1 might be curtailed, that is, underestimated in this condition. Consequently, had performance not been at ceiling, the global-precedence effect for large changes at set size 1 might have been considerably larger than that for small changes at set size 1 (where performance was not at ceiling). However, when considering only the 3-item displays (thus avoiding the ceiling effect), the pattern is very clear: the global-precedence effect is approximately twice as large for

[5] Note that performance in change-detection tasks hardly ever reaches 100%, owing to trial-by-trial fluctuations in attentional engagement (Adam, Mance, Fukuda, & Vogel, 2015).

large as compared to small changes, $t(17) = 3.51$, $p < .004$ (Figure 2C) – indicating that the bias across hierarchical levels is reduced, shifting towards the local object level when finer (orientation) details, at both object levels, need to be remembered.

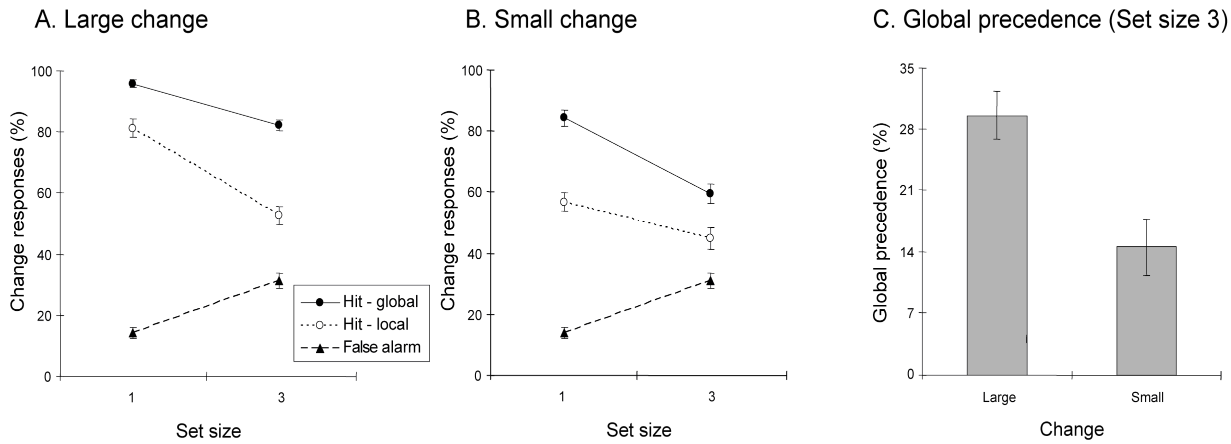

Figure 2. Mean percentage of change responses for large (A) and small (B) changes, and corresponding global precedence effects for set size 3 (C) in Experiment 1. Mean hit and false alarm rates are presented as a function of set size and change level (panels A and B). The global precedence effect on change-detection accuracies (C) is presented separately for large and small changes. The error bars represent ±1 SEM.

Discussion

Experiment 1 examined how change magnitude would affect hierarchical representations in vWM. The results replicated our earlier findings (Nie et al., 2017): change-detection performance was revealed to be superior for changes at the global relative to changes at the local object level. While the ability to detect a given change primarily depends on whether the item that is subject to the change is actually stored in memory (Nie et al., 2017; Awh, Barton, & Vogel, 2007), detection of small-magnitude changes may additionally depend on whether items are maintained with sufficient fidelity for the difference to be discernible. In line with this prediction, detection accuracy was overall lower for small-magnitude changes than for large-magnitude changes. Moreover, global precedence was reduced for detecting small- versus large-magnitude changes, suggesting that information maintained in vWM can be biased towards the local level when object details become task-relevant. Based on this finding, Experiment 2 was designed to explore how a further increase in demands for precision would affect global precedence in vWM.

EXPERIMENT 2

The standard change-detection paradigm as used above only allows us to quantitatively assess whether or not items were remembered, but it provides little information about how well each individual object was actually remembered (Fougnie, Asplund, & Marois, 2010). To gain a better idea of the quality of the maintained item representations, in Experiment 2, a continuous-report task was employed. Stimuli and trial sequence were essentially the same as in Experiment 1, except that the test probe was now replaced by a response wheel (see Figure 1C), requiring participants to reproduce, as precisely as possible, an orientation of the object at the indicated location in the preceding memory array. Different color cues were used to probe the global or local orientation of the hierarchical configurations. The observed deviation of the reported orientation from the correct orientation constitutes a much finer measure compared to the rather coarse change-detection accuracy. Importantly, as maximum precision (i.e., a 0°-deviation from the correct orientation) is never achieved in continuous-report tasks, ceiling performance is not an issue – ensuring a valid measure of the global-precedence effect, even for set size 1. Based on Experiment 1, we expected that the generally high demand for precision in the continuous-report paradigm would lead to an even stronger vWM bias towards the local object level.

Method

Experiment 2 was essentially identical to Experiment 1, except that participants had to reproduce the exact value of either the global or local orientation on a response wheel. The memory display and the subsequent delay were the same as in Experiment 1, but the final probe display was replaced by a recall display (Figure 1C). In this display, a colored circle (i.e., the response wheel) appeared at the location of the probed vWM stimulus (with white, placeholder circles appearing at non-probed locations). For half of the participants, a blue probe required the reproduction of the item's global orientation, and a green probe the reproduction of the item's local orientation. For the other half of participants, the color-to-object-level assignments were reversed. Participants responded by clicking the appropriate position on the response wheel (e.g., for a perfect response, the positions marked by the black arrows in Figure 1C); note that participants were free to respond to, on the wheel, either the upper or the lower end of the (oriented) lines/ellipses. After providing the response, feedback was given by displaying the correct orientation on the wheel for 1 s. Participants initially completed 8 trials with the memory items remaining on the screen until a response was issued, in order to understand the task. Next, one block of 40 practice trials was presented to familiarize observers with the actual memory task. The formal experiment was divided into 15 blocks of 40 trials each. Eighteen new observers (9 male; age range 22 to 40 years, mean age = 27.6 years; all with normal or corrected-to-normal visual acuity) participated in the experiment, receiving course credits or payment of 8 Euro per hour.

Mixture modeling analysis

On any given trial, we measured the deviation of the response from the correct orientation in degrees, with variations between 0° (perfect memory) and ±90° (poor memory, large deviation). In the continuous-report paradigm, the histogram of errors produced across trials typically shows that responses are centered at around 0°, but that, across all responses, errors are distributed across the entire range provided. The error histograms obtained for the four conditions in the current experiment showed a good fit by a mixture of two distributions: (a) a Gaussian-like distribution (defined on a circular space in terms of a von-Mises distribution), assumed to reflect successful memory retrieval with some, variable degree of precision; and (b) a uniform distribution reflecting random guessing (Zhang & Luck, 2008). We used Zhang and Luck's method to separate trials in which the orientation – at either the global or the local level – was retrieved with some degree of fidelity, and trials in which the orientation of the probed item was forgotten. Figure 3 shows an example of a representative participant, where the individual responses (as plotted in the histograms) are fitted to a mixture distribution for each condition. In addition, Figure 4 provides the overall distributions across participants from the mixture-model fitting procedure.

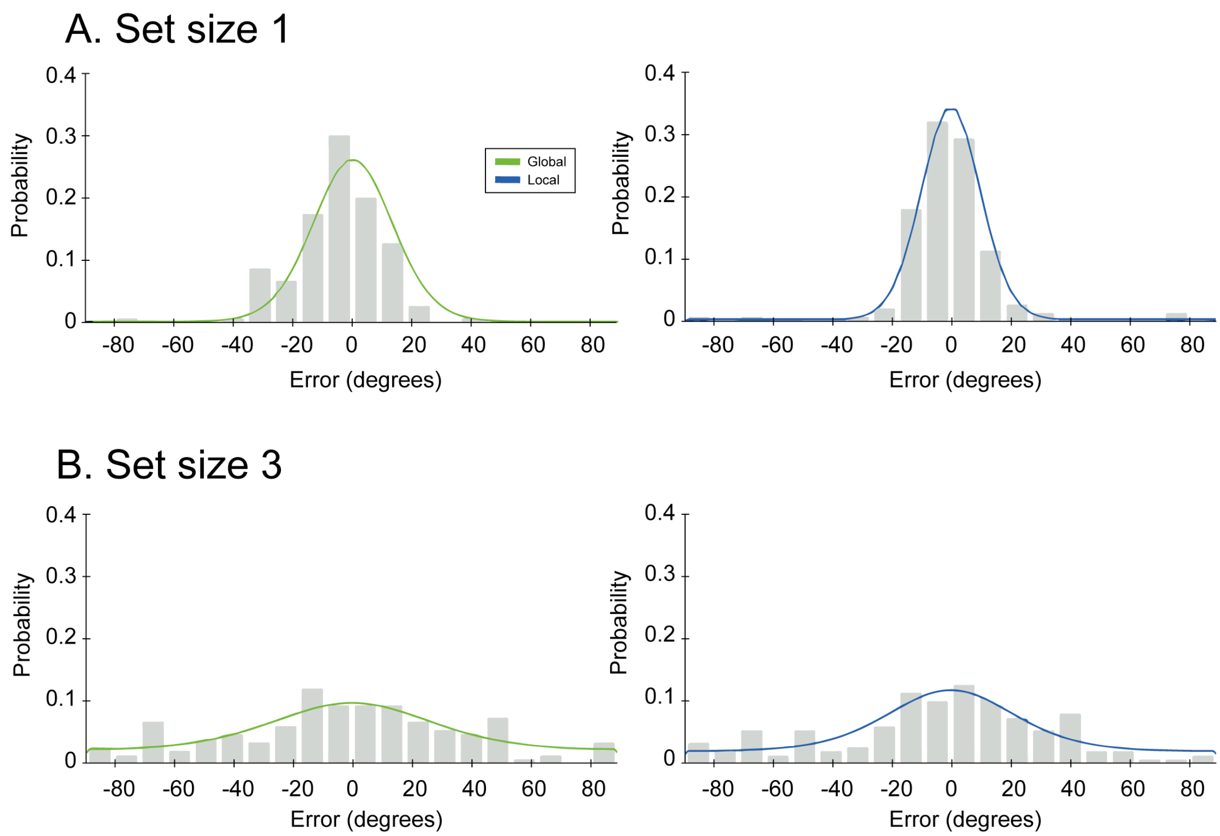

Figure 3. Results from one representative participant in Experiment 2. The histograms represent the distributions of the response errors (degree of deviation of response from target orientation) for set sizes 1 (A) and 3 (B). The green and blue curves show the model fits of a mixture model that combines a uniform guessing distribution with a Gaussian-like distribution of memory-based responses for the global (green) and local (blue) reproductions, respectively.

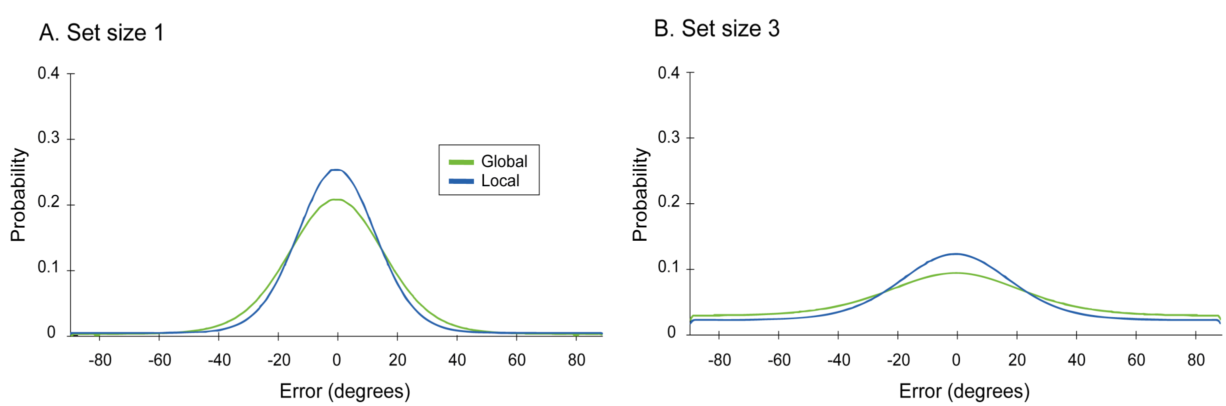

Figure 4. Modeled response-error distributions using the average of the best fitting parameter values for each participant for set size 1 (A) and 3 (B), for global (green) and local (blue) orientation reproductions.

The fidelity (precision) of a given memory representation was estimated as the standard deviation of the von-Mises distribution (σ). Accordingly, the narrower the distribution (with a relatively small standard deviation from 0°), the more precise the memory representation. The probability of guessing (g) was estimated by the height of the uniform distribution. Maximum-likelihood estimation was used to estimate these two parameters for each condition with the MemToolbox (Suchow, Brady, Fougnie, & Alvarez, 2013).

Results

Figure 4 presents the distribution of errors in all four experimental conditions. In addition, average guessing rate (g) and memory fidelity (σ) parameters for each experimental condition are depicted in Figure 5. To determine how g and σ differed across experimental conditions, a 2 (level: global, local) × 2 (set size: 1, 3) repeated-measures ANOVA was performed for each parameter.

First, the analysis of the *guessing rate* (g) revealed only the main effect of set size to be significant, with lower guess rates for set size 1 as compared to 3 (0.08 vs. 0.5), $F(1,17) = 98.07$, $p < .001$, $\eta_p^2 = .85$. Neither the main effect of level nor the interaction between level and set size were significant (ps > .24). As can be seen from Figure 5A, the guessing rate did not differ between the two object levels (global and local) at either set size (though there was a 'hint' of a *local*-precedence effect at set size 3, $t(17) = 1.13$, $p = .28$, dz = .27).

Next, the same analysis on *memory fidelity* (σ) revealed both main effects to be significant: level (global vs. local: 21.99 vs. 17.23), $F(1,17) = 7.5$, $p = .014$, $\eta_p^2 = .31$, and set size (1 vs. 3: 15.44 vs. 23.78), $F(1,17) = 30.22$, $p < .001$, $\eta_p^2 = .64$. The main effect of level indicates that mnemonic precision was substantially higher for local relative to global orientation, thus depicting a reversal of the typical global benefit, which in Experiment 2 actually manifested in terms of a *local*-precedence effect. In addition, a main effect of set size showed that memory precision was reduced for the larger set size. The interaction between level and set size was not significant ($F(1,17) = .44$, $p > .5$, $\eta_p^2 = .025$).

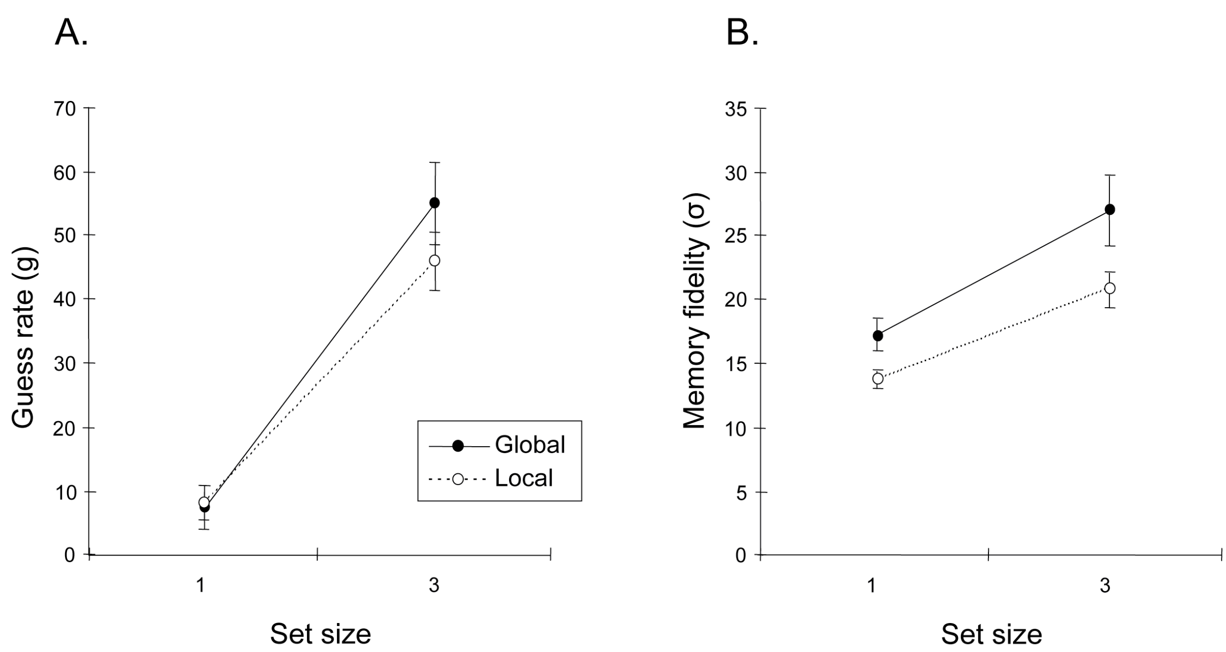

Figure 5. Estimated probability of guessing (A) and memory fidelity (standard deviation of the von Mises distribution, B) in Experiment 2. Results are depicted as a function of set size, separately for global and local object levels. Error bars represent ±1 SEM.

Discussion

Experiment 2 revealed a reliable local – as opposed to the typical global – bias in maintaining items in vWM: the general demand for precision associated with having to report the orientation (at both object levels) of a stored hierarchical object engendered more precise representation of the task-critical object attribute at the local than at the global level. In addition, the responses became more random as memory load increased.

This finding thus shows that with a change of the task from change detection (Experiment 1) to continuous report (Experiment 2), the very same stimuli are memorized differently: while a reliable global-precedence effect was obtained in Experiment 1, this pattern reversed into a local-precedence effect in Experiment 2. That is, in contrast to the pattern typically observed in paradigms using hierarchical stimuli (Kimchi, 1992), the observers in the present experiment were actually more precise in storing local, as compared to global, task-critical object information. Thus, the results of Experiments 1 and 2 taken together reveal a consistent trend: (i) a robust global-precedence effect for large to-be-detected changes (see also Nie et al., 2017), (ii) a considerably reduced global-precedence effect when the to-be-detected changes are small, and (iii) a reversal into local precedence when the task requires high precision.

General discussion

It is commonly thought that vWM is sensitive to the number of integrated objects (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). However, recent evidence suggests that vWM not only represents individual objects, but also the structural relations within and across memorized objects. For example, vWM for (global/local) hierarchical shapes displays separable influences of the objects' component parts and their respective wholes on memory capacity (Nie et al., 2017). The aim of the current study was to investigate the flexibility with which global/local structure is represented in vWM. To this end, two experiments were performed, which revealed clear modulations of mnemonic structure as a function of the required memory precision.

In Experiment 1, we used a change-detection task presenting hierarchical shapes with a global/local orientation, in which a small or large change occurred only at one of the two possible levels in a given target object. Overall, we found a robust pattern of global precedence: the hit rate (i.e., the rate of correctly detected orientation changes) was higher for the global than for the local object level. Importantly, the bias for the global object level was weaker when the orientation changes to-be-detected (at both object levels) were small rather than large. This indicates that, while the stored representations are overall biased towards the global level, this bias is flexible, that is, top-down-modulable: detecting large-magnitude changes potentially strengthens a default, 'gist'-level representation of the overall, global object structure(s) (Brady & Alvarez, 2015; Nie et al., 2017). Detecting small-magnitude changes, by contrast, requires a more detailed representation of the memorized objects, strengthening the system set for the local object level to support the representation of finer details.

To investigate whether this local bias persists when even more fine-grained detail is required to solve the task, a continuous-report paradigm was introduced in Experiment 2, which required observers to reproduce the exact memorized orientation at (randomly across trials) either the global or the local object level. Consistent with the reduction of global precedence with small-magnitude changes in Experiment 1, Experiment 2 actually revealed a pattern of *local* precedence. That is, when the demands for detailed representation were (consistently across levels) high, mnemonic precision reversed, with enhanced performance for the local relative to the global level. Taken together, these findings indicate that, with increasing demands for memorizing fine details, a default bias towards the global object level reverts gradually into an advantage for the local level. That is, within a given object to be memorized, the distribution of (limited) mnemonic resources is adjusted flexibly, in accordance with task demands, among the various levels of representation.

Interestingly, this flexible dynamics with which object hierarchies are represented in vWM does not appear to reflect global/local processing as revealed in classic attention tasks. Amongst the latter, visual-search studies, for instance, have typically revealed a persistent global bias (Kimchi, 1994; Conci, Töllner, Leszczynski, & Müller, 2011; Wagemans, Feldman, et al., 2012; Nie et al., 2016), which was little modulated by changes in task demands (Navon, 1977; Rauschenberger & Yantis, 2001).

Recent evidence suggests that change-detection and continuous-report tasks differ mainly in their demands for mnemonic precision (Brady et al., 2011; Fougnie et al., 2010), and vWM resolution may vary as a function of the required detail in change-detection tasks (Keshvari, Van den Berg, & Ma, 2013). That is, detecting changes of a small magnitude requires high-resolution object representations in vWM, whereas large changes can be detected with representations of relatively low resolution. In agreement with this view, the present findings demonstrate that the pattern of vWM precedence adjusts from global to local object levels for the very same stimuli, with demands for mnemonic precision varying both within a task (detection of changes at different magnitudes) and across tasks (change detection vs. continuous report) – thus supporting accounts that propose flexible allocation of mnemonic resources in vWM (Bays & Husain, 2008; Ma, Husain, & Bays, 2014).

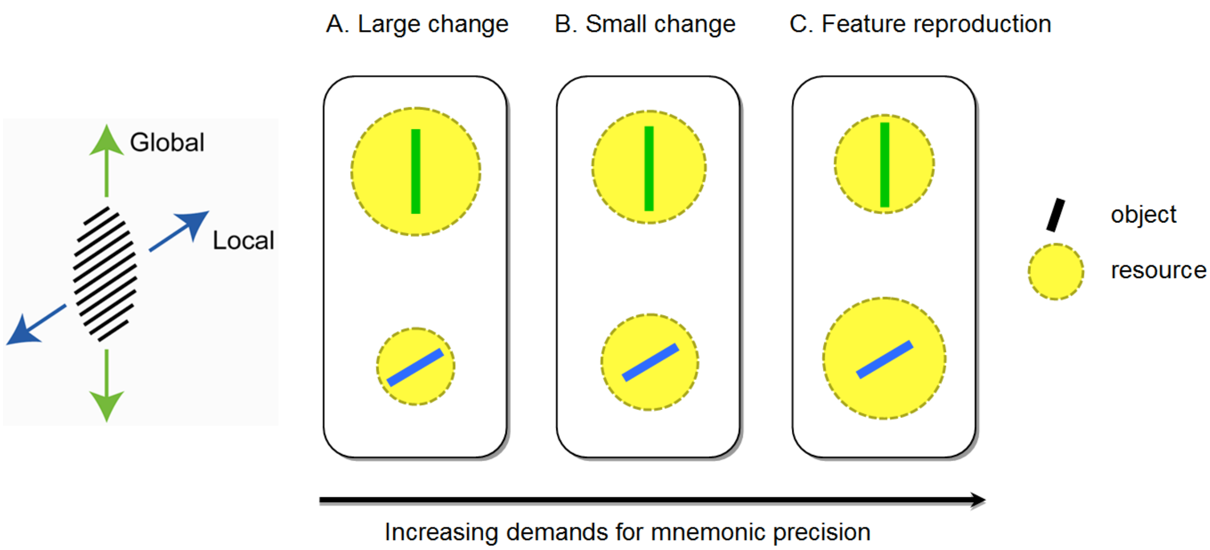

Figure 6. Flexible resource allocation under varying demands for mnemonic precision: As the demands for memory fidelity increase from (A) large-magnitude change detection through (B) small-magnitude change detection to (C) precise orientation reproduction, the allocation of limited mnemonic resources gradually changes from a default bias towards the global object level to a bias towards local object details.

Influential models of vWM assume that memory limits arise entirely from the availability of some limited resource that is either quantized into slots (Awh et al., 2007; Zhang & Luck, 2008; Luck & Vogel, 2013) or continuously divisible (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Wilken & Ma, 2004; Ma et al., 2014). These studies leave open the question of how structural relations relate to object representations in vWM. The present findings are, in principle, compatible with an account that proposes a flexible allocation of limited memory resources over hierarchical object levels (see Figure 6): Large-magnitude (standard) change-detection tasks make low demands as regards memory fidelity; as a result, a default, global memory bias becomes manifest, By contrast, higher fidelity representations are needed to detect small-magnitude changes, engendering a re-distribution of memory resources away from the global towards the local object level(s). Finally, for the precise reproduction of a given feature, the allocation of limited resources completely reverses and resources are predominantly deployed to represent the local level. In summary, this demand-based model suggests that the distribution of limited mnemonic resources among hierarchical object levels is flexible and adjusts to top-down requirements for precision.

Conclusion

Using standard vWM tasks in combination with hierarchical (orientation) stimuli, the present study provides novel evidence for structured memory representations that can be adapted in line with the current task demands. The modulation of hierarchical representations by requirements for precision points to a degree of flexibility in how structured representations are maintained in vWM. That is, hierarchical structure in vWM is not fixed; rather, observers can adjust the distribution of limited mnemonic resources over to-be-retained global and local object levels, to optimize memory performance in accordance with changing environmental demands.

References

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The Contribution of Attentional Lapses to Individual Differences in Visual Working Memory Capacity. *Journal of Cognitive Neuroscience, 27*(8), 1601-1616. doi: 10.1162/jocn_a_00811

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106-111. doi: 10.1111/j.0963-7214.2004.01502006.x

Awh, E., Barton, B., Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*(7), 622-628. doi: 10.1111/j.1467-9280.2007.01949.x

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851-854. doi: 10.1126/science.1158023

Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 921-929. doi: 10.1037/xlm0000075

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 4-4. doi: 10.1167/11.5.4

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

Conci, M., Müller, H. J., & Elliott, M. A. (2007). The contrasting impact of global and local object attributes on Kanizsa figure detection. *Perception & Psychophysics, 69*(8), 1278-1294. doi: 10.3758/BF03192945

Conci, M., Töllner, T., Leszczynski, M., & Müller, H. J. (2011). The time-course of global and local attentional guidance in Kanizsa-figure detection. *Neuropsychologia, 49*(9), 2456-2464. doi: 10.1016/j.neuropsychologia.2011.04.023

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision, 10*(12), 27-27. doi: 10.1167/10.12.27

Keshvari, S., Van den Berg, R., Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology, 9*(2): e1002927. doi: 10.1371/journal.pcbi.1002927

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance, 8*(4), 521-535. doi: 10.1037/0096-1523.8.4.521

Kimchi, R. (1992). Primacy of wholistic processing and global local paradigm: a critical review. *Psychological Bulletin, 112*(1), 24-38. doi: 10.1037/0033-2909.112.1.24

Kimchi, R. (1994). The role of wholistic/configural properties versus global properties in visual form perception. *Perception, 23*(5), 489-504. doi: 10.1068/p230489

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281. doi: 10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391-400. doi: 10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17*(3), 347-356. doi: 10.1038/nn.3655

Machizawa, M. G., Goh, C. C. W., & Driver, J. (2012). Human Visual Short-Term Memory Precision Can Be Varied at Will When the Number of Retained Items Is Low. *Psychological Science, 23*(6), 554-559. doi: 10.1177/0956797611431988

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353-383. doi:10.1016/0010-0285(77)90012-3

Nie, Q.-Y., Maurer, M., Müller, H. J., & Conci, M. (2016). Inhibition drives configural superiority of illusory Gestalt: Combined behavioral and drift–diffusion model evidence. *Cognition, 150*, 150-162. doi: 10.1016/j.cognition.2016.02.007

Nie, Q.-Y., Müller, H. J., & Conci, M. (2017). Hierarchical organization in visual working memory: From global ensemble to individual object structure. *Cognition, 159*, 85-96. doi: 10.1016/j.cognition.2016.11.009

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

Rauschenberger R., Yantis S. (2001). Attentional capture by globally defined objects. *Perception & Psychophysics, 63*(7), 1250-1261. doi: 10.3758/BF03194538

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision, 13*(10), 9-9. doi: 10.1167/13.10.9

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92-114. doi: 10.1037/0096-1523.27.1.92

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

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218-1252. doi: 10.1037/a0029334

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12), 11-11. doi: 10.1167/4.12.11

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235. doi: 10.1038/nature06860

VI. Summary and conclusions

The experimental and computational modeling work reported in this dissertation contributes to a better understanding of the structured representations in human attention and visual memory. Examining hierarchical stimuli that consist of more than one feature level, I established a new structure-based framework that provides novel perspectives for future work. Finally, I devised a novel hierarchical stimulus and examined the fidelity of structured representations in vWM.

Using a visual search variant to test the configural superiority effect (CSE) with illusory figures as either the target or distractors, we demonstrate the first evidence that distractor inhibition is the major driving force of the CSE. Our results provide a novel view to the question of how complete objects are derived from basic feature properties, and we demonstrate that inhibitory effects might emerge from a mid-level stage of visual processing, for instance, as reflected by faster evidence accumulation.

Next, we devised a hierarchical variant of a visual search task with Navon letters (Navon, 1977) as the global/local targets and nontargets. Our findings provides a novel theoretical perspective, which suggests that differences between hierarchical levels are not simply reflecting early visual processing (i.e., mechanisms of perceptual grouping) but also relate to a consistent and automatic bias in memory. We believe that our results challenge the view advocated by classical global/local paradigms (e.g., Navon, 1977), demonstrating that global precedence reflects the resolution of attention, i.e., the level at which we tend to select information. Moreover, we show that global biases might emerge at multiple stages of processing, for instance, reflecting preattentive and postselective information processing. As such global precedence represents a major property of the visual system that is evident throughout the entire object-processing stream.

By combining hierarchical shapes with a change detection task, we demonstrate the first evidence of hierarchically structured representations in visual working memory. Our results challenge models that simply propose a fixed number of units/objects, which are supposed to be retained independently of the other, to-be-remembered items. Instead, our findings support a more ecologically valid account – which we suggest in terms of a hierarchical model that emphasizes both the role of structured representations of objects and the scene layout within which they are presented.

The stimuli used in the above studies, e.g., Navon letters (Navon, 1977) or hierarchical shapes (Kimchi & Palmer, 1982), only allows for discrete change at either a global or a local level. We therefore devised a novel hierarchical stimulus that permits continuous changes in an orientation feature space. Combining such stimulus with a continuous report task, we demonstrate the first evidence that structured representations in vWM can dynamically adapt in line with task demands. Our results challenge models that propose relatively fixed representation of units/objects in vWM, but support adaptive, hierarchical accounts that emphasize both the structured nature of vWM representations and adaptive control of representational resolution to optimize mnemonic precision.

Overall, the findings in this dissertation provide a novel perspective to the prevailing feature- and object-based attention and working memory literature and support a new alternative theoretical framework, the hierarchical attention and working memory system, in which attention and visual memory are conceptualized as being hierarchically organized by local or part-level object representations.

References

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 521-535.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.

VII. Deutsche Zusammenfassung

Die in dieser Dissertation berichteten experimentellen Arbeiten und computationalen Modellierungen tragen zu einem besseren Verständnis bei, wie strukturierte Repräsentationen in visuellen Aufmerksamkeits- und Gedächtnismechanismen verankert sind. Indem hierarchische Stimulusanordnungen untersucht wurden, die aus mehreren Merkmalsebenen bestanden, konnte diesbezüglich ein strukturbasiertes Erklärungsmodel etabliert werden, welches neue Perspektiven für zukünftige Arbeiten bietet. Zudem wurden im Rahmen der präsentierten Untersuchungen innovative hierarchische Stimuli konzipiert, die es erlauben die Genauigkeit einer strukturellen Repräsentation im visuellen Arbeitsgedächtnis zu erfassen.

Im Rahmen einer visuellen Suchaufgabe wurde zunächst in Kapitel II der sogenannte „configural superiority effect“ (CSE) mit virtuellen Zielreiz- und Distraktorobjekten untersucht. Diese Arbeiten zeigen erste Evidenz dafür, dass Distraktorinhibition (und nicht etwa die Aktivierung eines Zielreizes) einen der wesentlichen Mechanismen zur Erklärung des CSE darstellt. Unsere Ergebnisse offerieren dabei eine neue Sicht auf die Frage wie sich integrierte Objekte aus basalen Merkmalseigenschaften ableiten lassen. Wir zeigen außerdem, dass inhibitorische Effekte auf einer mittleren Stufe der visuellen Verarbeitung entstehen, wie z.B. über die schnelle Akkumulation von Evidenz im Rahmen einer „drift/diffusion“ Modellierung dargelegt wurde.

In Kapitel III haben wir schließlich eine hierarchische Variante einer Suchaufgabe mit Navon-Buchstabenkonfigurationen entwickelt (Navon, 1977), bei der Zielreize und Distraktoren auf globalen/lokalen Ebenen repräsentiert sein konnten. Unsere Befunde mit diesem Paradigma zeigen eine neue theoretische Perspektive auf, die zeigt dass Unterschiede zwischen hierarchischen Ebenen nicht einfach nur frühe Stufen der visuellen Verarbeitung abbilden (z.B. Mechanismen der perzeptuellen Gruppierung) sondern auch eine konsistente und automatische Tendenz im Gedächtnis reflektieren. Entsprechend stellen diese Ergebnisse die Ideen aus klassischen globalen/lokalen Suchaufgaben (z.B. Navon, 1977) in Frage, nämlich, dass die globale Präzedenz im wesentlichen die Auflösung von Aufmerksamkeit widerspiegelt, d.h. die Ebene auf der wir Informationen auswählen. Vielmehr zeigen wir, dass ein globaler Vorteil auf multiplen Stufen der Verarbeitung entstehen kann, und somit sowohl präattentive als auch postselektive Abschnitte betrifft. Entsprechend zeigt sich in den Befunden zur globalen Präzedenz eine wesentliche Eigenschaft des visuellen Systems welche den gesamten Objektverarbeitungspfad betrifft.

In Kapitel IV wurden schließlich hierarchische Formkonfigurationen mit einer „change detection“ Aufgabe kombiniert. Wir zeigen dabei erste Evidenz dafür, dass hierarchisch strukturierte Repräsentationen im visuellen Arbeitsgedächtnis abgelegt sind. Unsere Befunde hinterfragen dabei Modelle die einfach nur annehmen, dass im Arbeitsgedächtnis eine festgelegte Anzahl von Einheiten oder Objekten repräsentiert ist, welche unabhängig von den anderen, zu erinnernden Einheiten abgespeichert sind. Statt dessen deuten unsere Ergebnisse auf einen ökologisch validen Ansatz hin, welcher besagt, dass strukturierte Repräsentationen in einem hierarchischen Modell, innerhalb der Szene in der die Objekte präsentiert wurden, abgebildet werden.

Die in den oben genannten Studien verwendeten Stimuli, z.B. Navon Buchstaben (Navon, 1977) oder hierarchische Formen (Kimchi & Palmer, 1982), erlauben es nur eine diskrete Änderung auf einer globalen oder lokalen Ebene zu implementieren. Aus diesem Grund haben wir in Kapitel V der vorliegenden Dissertation neue Varianten von hierarchischen Stimulusanordnungen entwickelt, welche uns erlauben fortlaufende

Änderungen von Objektorientierungen zu untersuchen. Die Kombination dieser Stimuli mit einer sogenannten „continuous report“ Aufgabe erlaubte es uns zu zeigen, dass strukturierte Repräsentationen im Arbeitsgedächtnis mit Bezug auf die spezifische Aufgabenstellung dynamisch adaptiert werden können. Unsere Ergebnisse unterstützen dabei adaptive und hierarchische Ansätze, die sowohl die strukturierte Repräsentation als auch die adaptive Kontrolle von repräsentationaler Auflösung zur Optimierung von mnemonischer Präzision vorschlagen.

Insgesamt zeigen die Befunde dieser Dissertation eine neue Perspektive zu vorherrschenden merkmals- und objektbasierten Ansätzen in Aufmerksamkeits- und Arbeitsgedächtnisfunktionen auf. Sie stützen dabei eine alternative Konzeption nach der die attentionale und gedächtnisbasierte Verarbeitung von Informationen hierarchisch organisiert ist.

Literatur

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 521-535.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.

Acknowledgements

This cumulative dissertation was written at the Ludwig-Maximilians University of Munich and supported by grants CO 1002/1-1from the German Research Foundation (DFG), and from the "LMUexcellent" Junior Researcher Fund.

A number of people have contributed to the successful completion of this work. I would like to thank Prof. Hermann J. Müller for his kind support and supervision. In addition, I am grateful to Dr. Markus Conci who has accompanied me throughout all steps from the first ideas to thesis completion and supervision. Thanks are also addressed to Dr. Heinrich René Liesefeld for great collaboration in Chapter V. In addition, I would like to thank Dr. Zhuanghua Shi for his invaluable support and technical help during the final stage of thesis completion, Birgitt Aßfalg for assisting me through finalizing the Ph.D process, Dr. Dragan Rangelov and Dr. Donatas Jonikaitis for many insightful discussions, and Mara Maurer and Dr. Xiaowei Ding for help with data collection.

Last but not least, I would like to thank my parents Zibing and Dexia, and my two sisters Qi-Yun and Qi-Mei for their support during the past years.

Qi-Yang Nie

Munich, November 2017

Curriculum Vitae

Qi-Yang Nie

born on October 10^{th} 1984 in Lu An/Anhui. (China)

Education

2003-2007 Bachelor in Environmental Engineering (Wuhan University, Hubei, China)

2007-2011 M.Sc. in Cognitive Neuroscience (Institute of Psychology at the Chinese Academy of Sciences, Beijing, China)

2012-2017 Ph.D. in Psychology (Ludwig-Maximilians-University, München, Germany)

Professional Experience

2012-2017 Research Fellow (Department of Psychology, Ludwig-Maximilians-University, München, Germany)

Publication list

Chen, S., **Nie, Q.-Y.**, Müller, H. J., & Conci, M. (*submitted*). Kanizsa-figure object completion attenuates the attentional blink. *Journal of Experimental Psychology*: *Human Perception and Performance*.

Nie, Q.-Y., Ding, X., Chen, J., & Conci, M. (2018). Social attention directs working memory maintenance. *Cognition*, 171, 85-94.

Sun, Y., Stein, T., Liu, W., Ding, X., & **Nie, Q.-Y.** (2017). Biphasic attentional orienting triggered by invisible social signals. *Cognition*, 168, 129-139.

Nie, Q.-Y., Müller, H. J., & Conci, M. (2017). Hierarchical organization in visual working memory: From global ensemble to individual object structure. *Cognition*, 159, 85-96. **(thesis study, Chapter 3)**

Nie, Q.-Y., Maurer, M., Müller, H. J., & Conci, M. (2016). Inhibition drives configural superiority of illusory Gestalt: Combined behavioral and drift-diffusion model evidence. *Cognition*, 150, 150-162. **(thesis study, Chapter 2)**

Tang, X.*, Pang, J.*, **Nie, Q.-Y.**, Conci, M., Luo, J., & Luo, J. (2016). Probing the cognitive mechanism of mental representational change during chunk decomposition: A parametric fMRI study. *Cerebral Cortex*, 26(7), 2991-2999.

Nie, Q.-Y., & Luo, J. (2012). “Aha!” and “haha!”: The common and distinct cognitive brain processes underlying insight and humor. *Advances in Psychological Science*, 20(2): 219-227.