



A New Conceptualization of Human Visual Sensory-Memory

Haluk Ögmen^{1,2*} and Michael H. Herzog³

¹ Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA, ² Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX, USA, ³ Laboratory of Psychophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Memory is an essential component of cognition and disorders of memory have significant individual and societal costs. The Atkinson–Shiffrin “modal model” forms the foundation of our understanding of human memory. It consists of three stores: Sensory Memory (SM), whose visual component is called iconic memory, Short-Term Memory (STM; also called working memory, WM), and Long-Term Memory (LTM). Since its inception, shortcomings of all three components of the modal model have been identified. While the theories of STM and LTM underwent significant modifications to address these shortcomings, models of the iconic memory remained largely unchanged: A high capacity but rapidly decaying store whose contents are encoded in retinotopic coordinates, i.e., according to how the stimulus is projected on the retina. The fundamental shortcoming of iconic memory models is that, because contents are encoded in retinotopic coordinates, the iconic memory cannot hold any useful information under normal viewing conditions when objects or the subject are in motion. Hence, half-century after its formulation, it remains an unresolved problem whether and how the first stage of the modal model serves any useful function and how subsequent stages of the modal model receive inputs from the environment. Here, we propose a new conceptualization of human visual sensory memory by introducing an additional component whose reference-frame consists of motion-grouping based coordinates rather than retinotopic coordinates. We review data supporting this new model and discuss how it offers solutions to the paradoxes of the traditional model of sensory memory.

Keywords: sensory memory, iconic memory, modal model, non-retinotopic memory, non-retinotopic processes

OPEN ACCESS

Edited by:

Britt Anderson,
University of Waterloo, Canada

Reviewed by:

Ijla G. Sligte,
University of Amsterdam, Netherlands
Jan Brascamp,
Michigan State University, USA

*Correspondence:

Haluk Ögmen
ogmen@uh.edu

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: 02 February 2016

Accepted: 18 May 2016

Published: 09 June 2016

Citation:

Ögmen H and Herzog MH (2016) A
New Conceptualization of Human
Visual Sensory-Memory.
Front. Psychol. 7:830.
doi: 10.3389/fpsyg.2016.00830

INTRODUCTION

Modal Model of Human Memory

The realization that human memory is not a unitary process but consists of multiple stores with distinct characteristics led to the Atkinson–Shiffrin, or the “modal” model of human memory (Atkinson and Shiffrin, 1968). As shown in **Figure 1**, this model consists of three major stores: The input is first stored in sensory memory (SM), which exhibits a very large capacity, but can maintain information only for a few hundred milliseconds. A subset of the contents of this rapidly decaying memory is transferred to Short-Term Memory (STM; also known as Working Memory WM). STM is severely limited in capacity and can hold information for several seconds to minutes. Finally, information is stored in Long-Term Memory (LTM), a store with very large capacity, capable of holding information as long as one’s lifetime. Since its inception, the STM and LTM components of

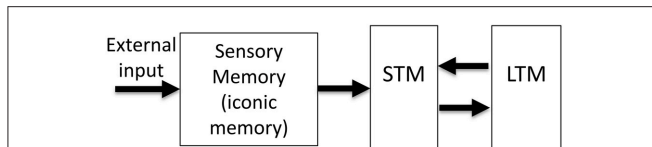


FIGURE 1 | The Atkinson-Shiffrin, or the “modal model” of human memory. The external input is first stored in a large capacity but rapidly decaying store, called Sensory Memory (SM). Alternative terms for this stage include sensory register and sensory store. The visual component of SM is also called iconic memory. A distinction has been made between visible persistence vs. informational persistence in visual SM. Some authors use the term iconic memory only for the informational persistence component of visual SM, whereas others use for both visual and informational components. A subset of the contents of SM is transferred into a more durable (few seconds) but severely capacity-limited store, called Short-Term Memory (STM). Alternative terms for this stage include Short-Term Store, or Working Memory (WM). Some authors make a distinction between the use of the terms STM and WM. Finally, the contents of STM are transferred to Long-Term Memory (LTM), also called Long-Term Store, a store with very long duration and very large capacity.

the modal model have undergone significant modifications (review: Baddeley, 2007), while SM has remained largely unchanged¹.

Sensory (Iconic) Memory

The SM component of the modal model is based on Sperling's work in 1960s (Sperling, 1960; Averbach and Sperling, 1961). By using the partial-report technique, Sperling showed that a large-capacity visual memory stores information for few hundred milliseconds (Sperling, 1960) and more recent studies indicate that this information is not implicit and unconscious but rather directly reflects the phenomenal richness of our visual experience (Vandenbroucke et al., 2014). Early information processing theories viewed SM as a real-time buffer, which briefly stores the inputs impinging on the retina to allow attentional mechanisms to select a subset of this information for transfer to the limited capacity WM. However, subsequent analyses taking into account the properties of dynamic ecological viewing conditions showed that SM cannot fulfill this function during normal viewing conditions when objects or the subject are in motion; in fact, SM appears to be a hindrance to vision (Haber, 1983). A fundamental characteristic of iconic memory is that its contents are encoded

¹More recently, a new type of STM, called fragile STM has been proposed (Slight et al., 2008; Pinto et al., 2013). It has been suggested that fragile STM takes an intermediate position between SM and STM. Whether this is a memory component genuinely distinct from STM is under debate (e.g., Matsukura and Hollingworth, 2011; Makovski, 2012). In general, fragile STM seems to share the properties of STM, rather than SM. However, because it has not been tested with moving stimuli, it is difficult to compare it with the non-retinotopic sensory memory that we discuss here. We can draw, however, two important distinctions, between non-retinotopic memory that we discuss and fragile STM: Fragile STM is very sensitive to attention and cueing; but as we discuss in Sections Sequential Metacontrast: Non-Retinotopic Information Storage and Processing and A New Conceptualization of Human Sensory Memory, the role of attention in non-retinotopic memory depends on the stimulus configuration. Second, interference (or masking) happens to the contents of fragile STM when similar stimuli are presented at the same locations. As we discuss in Section Sequential Metacontrast: Non-retinotopic Information Storage and Processing, even though stimulus can be completely masked, its informational contents are robust in non-retinotopic memory.

in retinotopic coordinates (Haber, 1983; Irwin et al., 1983, 1988; Jonides et al., 1983; Rayner and Pollatsek, 1983; van der Heijden et al., 1986; Sun and Irwin, 1987). While a retinotopically encoded memory can serve a useful function when the observer and the objects in the environment are all static, it cannot store any meaningful information when the observer's eyes, head, body and external objects are in motion. Any relative motion between the observer's retinae and the external environment will cause a shift in retinotopic coordinates for the stimulus received by SM. These shifts, in turn, will cause blurring and inappropriate integration of information over space and time: A briefly presented stimulus remains visible for about 120 ms after its offset under normal viewing conditions (Coltheart, 1980), a phenomenon known as visible persistence (the visible component of SM²). Hence, if the input shifts in retinotopic coordinates, it will create partially processed copies of the stimulus that will be superimposed upon each other at different retinotopic locations, creating a blurred version of the stimulus. For example, given a visible persistence duration of 120 ms, an object moving at 8.3°/s will generate a blur trail of 1°. This motion blur is similar to pictures of moving objects taken by a camera at relatively slow shutter speeds mimicking the duration of visible persistence (Figure 2). Similarly, when the observer moves her head, body, and eyes, the retinotopic shift of stimuli engenders multiple blurred copies superimposed upon each other in SM. Since movements of the subject and the objects are characteristics of ecological normal viewing conditions, the emerging consensus has been that a retinotopically encoded memory cannot serve any useful function under normal viewing conditions. To explain our relatively sharp and clear percepts under normal viewing conditions, there have been several attempts to identify a spatiotopic version of this memory (Davidson et al., 1973; Ritter, 1976; White, 1976; Wolfe et al., 1978a,b; Breitmeyer et al., 1982; Jonides et al., 1982; McRae et al., 1987); however, these were unsuccessful (Haber, 1983; Irwin et al., 1983, 1988; Jonides et al., 1983; Rayner and Pollatsek, 1983; van der Heijden et al., 1986; Sun and Irwin, 1987) and this area of research has been stagnant for half-century.

To move forward, three fundamental questions need to be addressed:

- Q1. How does the visual system process and store information non-retinotopically over space and time?
- Q2. How does the visual system control deleterious effects of retinotopic sensory memory?
- Q3. What purpose does a retinotopic sensory memory serve?

In the following we provide answers to these questions. First, to put retinotopy in the context of visual perception, in Section Metacontrast and Anorthoscopic Perception: A Retinotopically Extended Representation is Neither Sufficient nor Necessary for Vision we present evidence showing that retinotopic representations are neither sufficient nor necessary

²A distinction has been made between visible persistence vs. informational persistence (Coltheart, 1980). Some authors use the term iconic memory only for the informational persistence component of visual SM, whereas others use for both visual and informational components. We use the terms Sensory Memory and iconic memory to include both visible and informational persistence components.

for visual perception. In Sections Sequential Metacontrast: Non-retinotopic Information Storage and Processing and The Ternus-Pikler Paradigm to Probe Retinotopic and Non-retinotopic Processes, we review briefly two experimental paradigms that we have used to demonstrate the existence of a non-retinotopic memory. Based on these findings, in Section A New Conceptualization of Human Sensory Memory, we present a modified version of the modal model with a new sensory memory component. In Section Paradoxes of Retinotopic Sensory Memory Revisited, we revisit the paradoxes associated with the retinotopic sensory memory and discuss how the new model offers resolutions to these paradoxes.

NON-RETINOTOPIC INFORMATION PROCESSING AND STORAGE

Metacontrast and Anorthoscopic Perception: A Retinotopically Extended Representation is Neither Sufficient nor Necessary for Vision

Assume that one briefly flashes a supra-threshold stimulus; the observer will clearly perceive the shape of this stimulus. Assume now that a second stimulus is flashed *after* this stimulus in a way that it surrounds but does not spatially overlap with the first stimulus. This second stimulus can render the first one completely invisible. This phenomenon is known as visual masking, which refers to the reduced visibility of one stimulus (target), due to the presence of another stimulus (mask) (Bachmann, 1994; Breitmeyer and Öğmen, 2000, 2006). Metacontrast is a specific type of visual masking, in which the target and mask do not overlap spatially. Hence in metacontrast, the target maintains its retinotopic representation, i.e., the mask does not reduce the visibility of the target by directly occluding the retinotopic representation of the target. Hence a retinotopic representation of a stimulus is *not sufficient* for its perception



FIGURE 2 | A picture illustrating deleterious effects of a retinotopic Sensory Memory (rSM). As an object moves in the retinotopic space, its trace in rSM will create blur, as is the case for moving objects in the scene. Retinotopically shifted copies of the stimulus are superimposed and integrated, creating formless, ghost-like, appearances for moving objects. Similarly, the movements of observer's eyes, head, and body will induce retinotopic shifts in the image, blurring the whole scene and making the scene unrecognizable with a ghost-like appearance. Original photo from FreeFoto.com by permission.

or storage. The mask may be interfering *indirectly* with the retinotopic representation of the mask; but why would the visual system suppress a perfectly visible stimulus when it is embedded in a dynamic context? In the following sections, we will re-visit visual masking and its role in controlling sensory memory in Sections Sequential Metacontrast: Non-retinotopic Information Storage and Processing and How the Visual System Controls Deleterious Effects of rSM: Motion Deblurring to answer the questions Q1 and Q2 above.

Anorthoscopic perception, or slit viewing, is an experimental paradigm that derives its name from the anorthoscope, a device invented by Plateau in the nineteenth century (Plateau, 1836). Since its invention, variants of this device have been used to study human perception (e.g., Zöllner, 1862; Parks, 1965; Rock, 1981; Morgan et al., 1982; Sohmiya and Sohmiya, 1994; Öğmen, 2007; Aydin et al., 2008, 2009; Agaoglu et al., 2012). As depicted in **Figure 3**, a moving stimulus is viewed behind a narrow slit. All spatial information about the moving stimulus is restricted into a very narrow retinotopic strip. At any given time, only a very narrow spatial structure of the stimulus is visible. In other words, there is no spatially extended retinotopic representation for the moving stimulus. Moreover, as the stimulus moves, *different* parts of the object's shape fall onto the *same* retinotopic area. Hence, the contents of a retinotopic SM will consist of all stimulus parts mixed and blended into each other within the slit area. Yet, observers are able to spatiotemporally integrate the slit views to construct the spatially extended form of the moving stimulus in the absence of a retinotopically extended representation of the stimulus. Hence, anorthoscopic perception shows that a spatially extended retinotopic representation is *not*

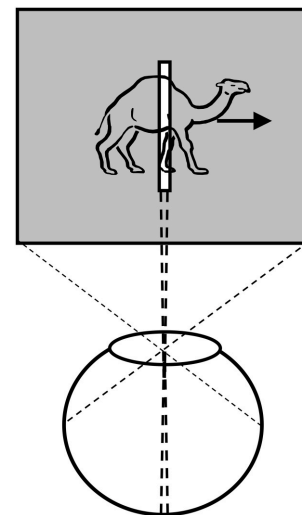
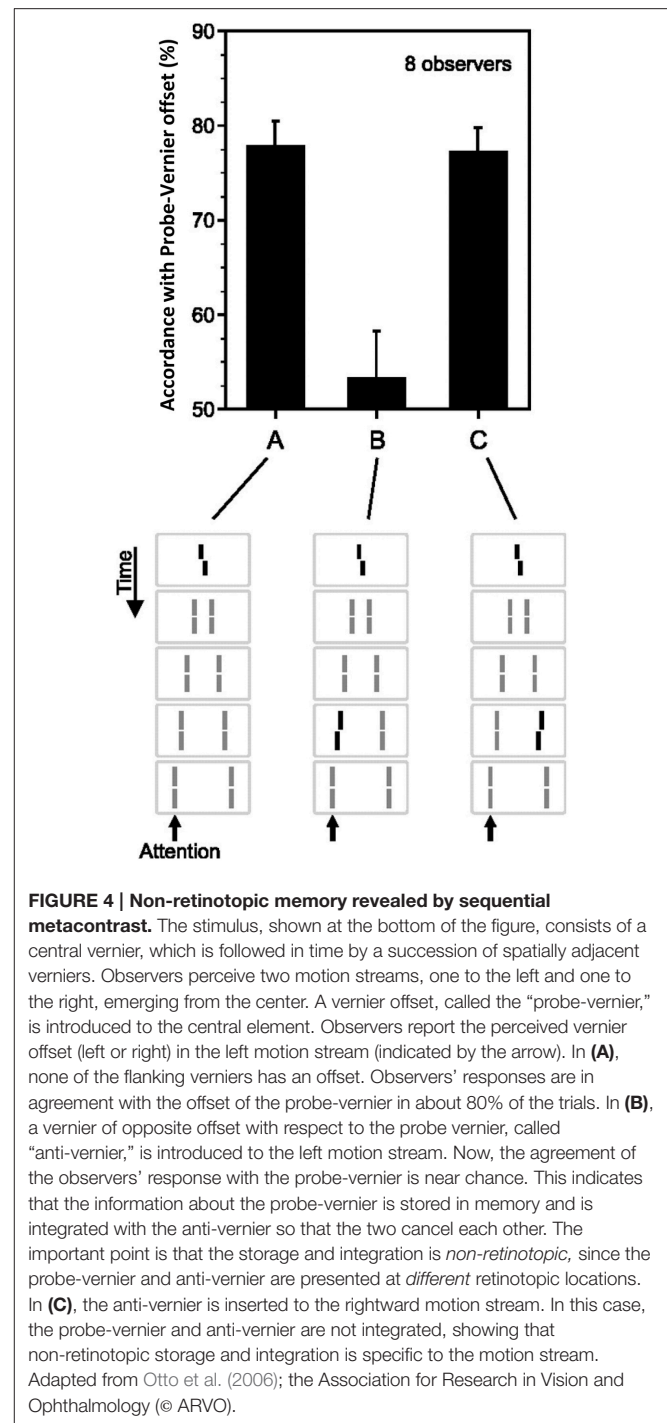


FIGURE 3 | Anorthoscopic perception. A stimulus is moved behind an opaque region with only a very narrow opening in the form of a slit. As the object moves behind the slit, samples of its form taken along the slit fall onto the same narrow retinotopic region. Hence, these different narrow two-dimensional samples will be superimposed and integrated in the retinotopic sensory memory, making the perception of spatially extended form impossible. From Öğmen (2007).

necessary for the perception of spatial form. Moreover, it also shows that, since information about different parts of the shape are shown at different time instants, the visual system is able to store this information and integrate it *non-retinotopically* in order to build the complete spatial layout of the stimulus. This indicates some type of non-retinotopic memory (Haber, 1983) but until recently it was not clear how it may be operating and its relation to the more traditional retinotopic SM. We will revisit anorthoscopic perception in Section Anorthoscopic Perception.

Sequential Metacontrast: Non-retinotopic Information Storage and Processing

Sequential metacontrast (Piéron, 1935; Otto et al., 2006) is a special case of metacontrast, consisting of *multiple* target and mask pairs as shown at the bottom of **Figure 4**. A central target is presented first, followed by two spatially flanking masks, which in turn are followed by lateral masks on one side, etc. Observers perceive two motion streams originating from the center, one to the left, and one to the right. With the appropriate choice of stimulus parameters, the central element can be completely masked making observers unable to tell whether it is presented or not (Otto et al., 2006). In order to test non-retinotopic storage and integration of information, we introduced a feature into this invisible central element in the form of a vernier offset (called “probe-vernier”), with a random spatial shift between its two segments, left or right, from trial to trial. Observers were asked to report the perceived vernier-offset in the left motion stream. Observers did not know if and where vernier offset(s) were presented in the display. In **Figure 4A**, all flanking lines are non- offset and the responses of the observers are above chance level with the offset of the probe-vernier. This indicates that the central *invisible* probe-vernier’s offset direction is *stored in a non-retinotopic memory and attributed to the left motion stream*, a process that we call feature attribution. In **Figure 4B**, a vernier of opposite offset-direction is introduced into the left stream (in reference to the probe-vernier, this is called “anti-vernier” hereafter, because its offset direction is always opposite to that of the probe-vernier). Now, the agreement of observer’s response with the offset direction of the probe-vernier is near 50%, i.e., the point of equal strength. This indicates that the two *verniers are integrated* in the non-retinotopic memory and as a result they cancel each other. Finally, in **Figure 4C** we show that this *integration is specific to the motion stream, i.e., the two verniers are integrated only if they belong to the same motion stream*. The probe-vernier is symmetric with respect to leftward and rightward motion streams; hence it will be attributed to both streams. The anti-vernier will be integrated with the probe vernier only in the specific motion stream where it appears. In **Figures 4B,C**, it will be integrated exclusively within the leftward and rightward motion streams, respectively. Since the observer is reporting the leftward motion stream, the integration is revealed in observer’s response in **Figure 4B** but not in **Figure 4C**. Taken together, these results show that information presented at the central retinotopic location is stored in memory and is integrated with the information presented at other retinotopic locations according to the motion of stimuli, hence, providing evidence



for a *non-retinotopic memory that depends on stimulus motion*. Additional results supporting this finding (with multiple vernier’s inserted at multiple locations) can be found in Otto et al. (2006, 2009, 2010a,b).

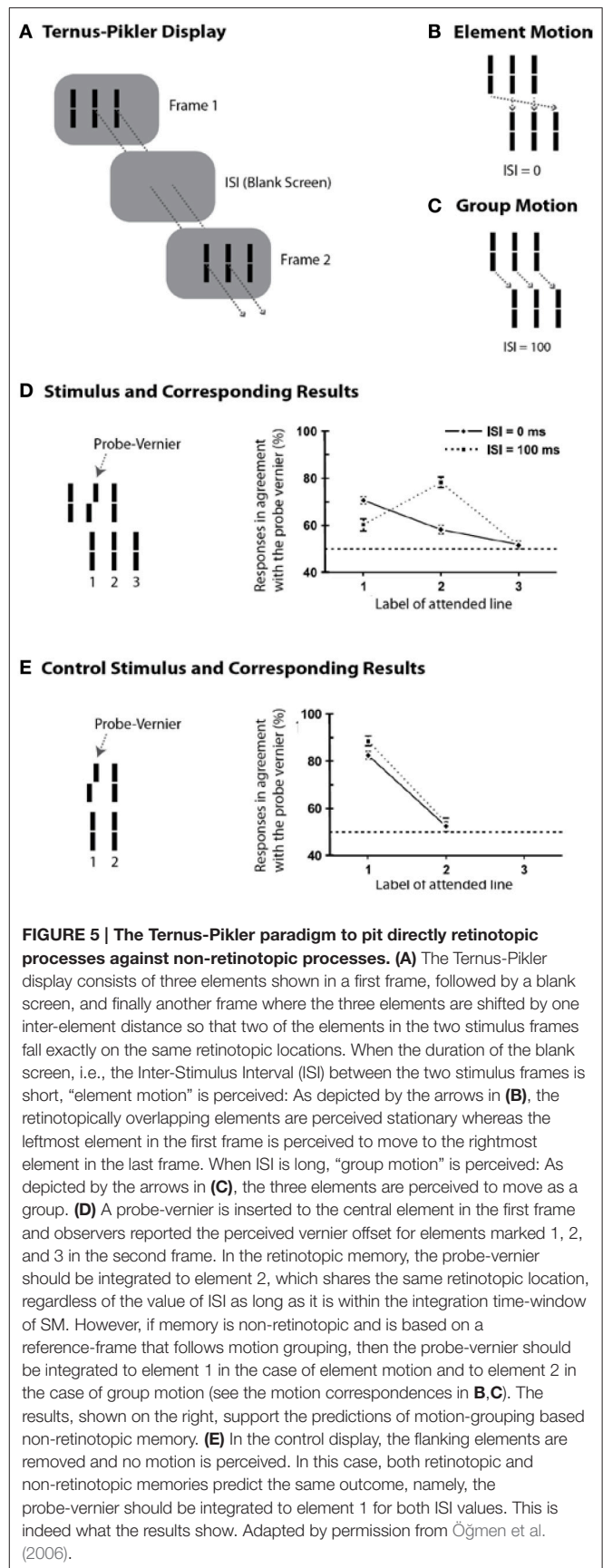
A methodological difference between traditional studies of memory and the sequential metacontrast paradigm outlined above is the ways cues are used. In the partial-report technique, after the offset of the stimulus, a cue is delivered (with a delay) to

indicate which item(s) to report (Sperling, 1960). As soon as the cue is delivered, the observer can initiate the reporting process. More recent studies combined change-detection paradigms with *retro-cueing* to investigate memory processes, in particular STM (e.g., Griffin and Nobre, 2003; Sligte et al., 2008; Hollingworth and Maxcey-Richard, 2013; Rerko et al., 2014; van Moorselaar et al., 2015). In these studies, an array of items is presented, followed by a retro-cue, and finally by a comparison item or display. The task of the observer is to report whether the cued item has changed. Hence, in this paradigm, the cue itself is not sufficient to initiate the response. Sperling's original purpose for introducing the cue was to design a *partial* report task so as to avoid the decay of information during the time it takes to report the contents of the memory. In addition to indexing the items to be reported, cues also guide attention and hence allow the examination of the role attention may play in the storage, maintenance, or transfer of information in memory. For example, a retro-cue indicates to the observer which particular memory item to attend in order to complete the impending comparison.

We have combined cueing with sequential metacontrast to examine the role of attention in non-retinotopic memory (Otto et al., 2010a). In a first experiment, we used a stimulus as shown in **Figure 4**. The stimulus could contain only a central vernier (as in **Figure 4A**), a central and a flanking vernier (as in **Figures 4B,C**), or only a flanking vernier. In the experiment described in **Figure 4**, the observers were instructed at the beginning of the block of trials which motion stream to attend (Otto et al., 2006). In Otto et al. (2010a), we used an auditory cue that indicated which motion stream (leftward or rightward) to attend. We varied the timing of the auditory cue with respect to the stimulus. When the cue was delivered before stimulus onset, observers focused their attention exclusively on the cued stream. By delaying the cue, we could control when unifocal attention could be devoted to a particular motion stream. Accordingly, the cue could focus attention preferentially on the central or the flanking vernier depending on its timing with respect to the onset of the central and flanking vernier. Our results showed that neither the timing nor the distribution of attention (focused on one stream vs. distributed to both streams) had any specific effect on non-retinotopic feature integration. These findings indicate that attention cannot directly access single lines and mandatory feature integration occurs within the attended motion stream.

The Ternus-Pikler Paradigm to Probe Retinotopic and Non-retinotopic Processes

Whereas the sequential metacontrast paradigm provides evidence for non-retinotopic memory, it does not pit directly retinotopic and non-retinotopic processes against each other. In order to achieve this goal, we developed an experimental paradigm that can pit directly retinotopic and non-retinotopic memories against each other, while revealing the direct role of motion in the process. To this end, we modified a stimulus paradigm developed by Gestalt psychologists Ternus and Pikler (Pikler, 1917; Ternus, 1926). **Figure 5A** shows a basic Ternus-Pikler display. The first frame contains three elements.



After an inter-stimulus interval (ISI), these three elements are shifted to the right by one inter-element distance so that two of the elements overlap retinotopically across the two frames (these retinotopically overlapping elements allow the testing of retinotopic information storage and integration). For small values of ISI, observers report seeing the leftmost element of the first frame move to the rightmost element of the second frame, while the other two elements appear stationary (**Figure 5B**). This percept is called “element motion” (Pantle and Picciano, 1976). For longer ISIs, observers report seeing all three elements moving in tandem to the right as a group (**Figure 5C**). This percept is called “group motion.” These motion-based non-retinotopic correspondences between the elements in the two frames allow the testing of motion-based, non-retinotopic information storage and integration. The probe-vernier was inserted to the central element of the first frame as shown in **Figure 5D-left** (Öğmen et al., 2006). We asked observers to report the perceived offset-direction for elements in the second frame, numbered 1, 2, and 3 in **Figure 5D-left**. None of these elements contained a vernier offset and naïve observers did not know where the probe-vernier was located. Consider first the control condition in **Figure 5E**, obtained by removing the flanking elements from the two frames. In this case no motion is perceived. Based on the traditional retinotopic iconic memory account, the information about the probe vernier should be stored at its retinotopic location and it should be integrated with element 1 in the second frame, which occupies the same retinotopic location. Thus, the agreement of observers’ responses with the direction of probe-vernier offset should be high for element 1 and low for element 2. In agreement with the large body of findings on iconic memory, this is indeed what we found (data in **Figure 5E-right**). A retinotopic iconic memory predicts the same outcome for the Ternus-Pikler display regardless of whether element or group motion is perceived, as long as the ISI is within the time-scale of iconic memory. On the other hand, if there is a memory mechanism that stores and integrates information non-retinotopically according to motion grouping relations (**Figures 5B,C**), one would expect the probe vernier to integrate with element 1 in the case of element motion (**Figure 5B**) and with element 2 in the case of group motion (**Figure 5C**). Our results supported the predictions of the non-retinotopic motion-based memory hypothesis (**5D-right**). Additional results supporting this finding (with multiple vernier’s inserted at multiple locations) can be found in Öğmen et al. (2006), Scharnowski et al. (2007), Otto et al. (2008), Boi et al. (2009, 2011), and Noory et al. (2015a,b).

A NEW CONCEPTUALIZATION OF HUMAN SENSORY MEMORY

Extensive research supports the existence of a retinotopic sensory memory (review: Coltheart, 1980). The research reviewed in the previous section supports the existence of a *non-retinotopic, motion-based*, sensory memory. Taken into account these recent findings, we modified the sensory memory component of the modal model by introducing a non-retinotopic store (**Figure 6**). To be consistent with the terminology used in the literature, we

keep the term iconic memory for the retinotopic component of sensory memory and also refer to this component as the “*retinotopic Sensory Memory*” (*rSM*). We call the newly introduced non-retinotopic component, the “*non-retinotopic Sensory Memory*” (*nrSM*).

Below, we discuss the fundamental properties of the new model and the key differences between its *rSM* and *nrSM* components. Specifically, we will discuss the differences in terms of reference-frames used in each, how masking affects the contents of each memory component, the distinct but complementary roles masking and motion play with respect to these two components, and finally the influence of attention mechanisms:

- (i) *Retinotopic vs. motion-based reference-frames*: Whereas the reference-frame, or the coordinate system, of *rSM* is anchored in retinotopic coordinates, *nrSM* uses motion-grouping-based non-retinotopic reference-frames or coordinate systems. **Figure 7** depicts the operations underlying *nrSM*. At a first stage, motion is analyzed within retinotopic representations and motion vectors are grouped according to Gestalt grouping principles (e.g., common fate). For example, in **Figure 7**, the moving dots are grouped into two distinct groups based on their direction of motion. For each group, a common motion vector is computed and this common motion vector serves as the reference-frame according to which the contents of memory are encoded. As illustrated in the example, multiple motion groupings can be extracted simultaneously across the visual field and hence *nrSM* can contain multiple reference-frames, unlike *rSM* which has a single reference-frame anchored in retinotopic coordinates.
- (ii) *Immunity to masking*: In the experiments discussed in Section Sequential Metacontrast: Non-retinotopic Information Storage and Processing, the probe-vernier can be completely masked; however, the information about its vernier-offset is not masked since it is integrated to other verniers in the motion stream and observers can reliably report the direction of vernier offset in behavioral experiments. By using the Ternus–Pikler display, we tested directly whether masking operates in retinotopic coordinates and whether *nrSM* is susceptible to masking (Noory et al., 2015b). Our results showed that masking operates in retinotopic coordinates and *nrSM* is immune to masking (Noory et al., 2015b). Hence, unlike *rSM* whose contents are suppressed by masking (Averbach and Coriell, 1961; Averbach and Sperling, 1961; Coltheart, 1980), the contents of *nrSM* are immune to masking (Noory et al., 2015b).
- (iii) *Distinct and complementary roles of masking and motion in sensory memory*: Masking “turns off” *rSM* whereas motion “activates” *nrSM*³. To test the proposed distinct

³Note that for static stimuli and static observer, the velocity is 0 and hence “motion-based” reference-frame with a null velocity vector becomes identical to a retinotopic reference-frame. However, from a mechanistic point of view, if *nrSM* uses the activities of motion detectors to synthesize its reference-frame, in the case of static stimuli, there will be no motion-detector activity to generate the reference-frame. Hence to store information about static stimuli, a memory component

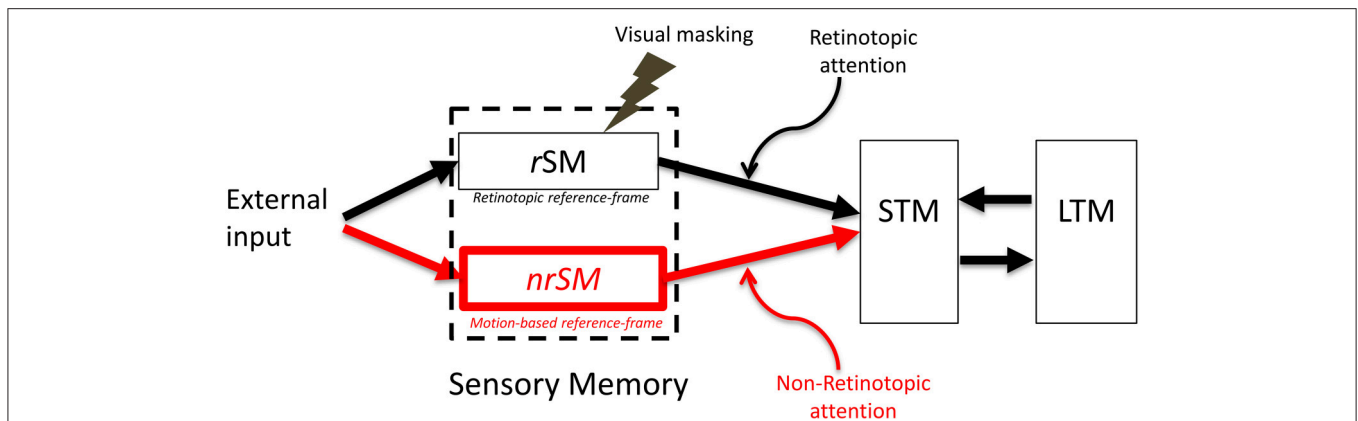


FIGURE 6 | The proposed new model. The SM stage of the modal model is modified to include two stores, a retinotopic SM (*rSM*), corresponding to the SM in the modal model and an additional non-retinotopic SM (*nrSM*). The reference-frame for *rSM* is anchored in retinotopic coordinates, whereas the reference-frame of *nrSM* is non-retinotopic, flexible and based on motion groupings in the stimulus. Hence, *nrSM* can have simultaneously multiple reference-frames depending on how the stimulus generates motion groupings (see **Figure 7**, for an example). The contents of *rSM* can be suppressed by visual masking whereas *nrSM* is immune to masking. Information from *rSM* and *nrSM* is transferred to STM through retinotopic and non-retinotopic attention mechanisms.

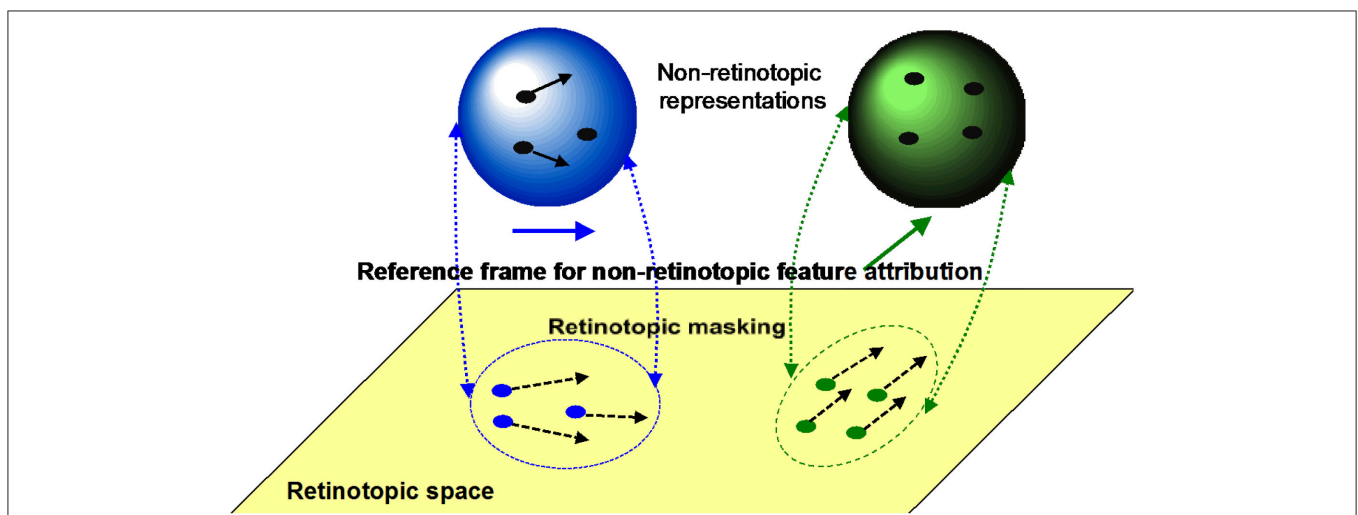
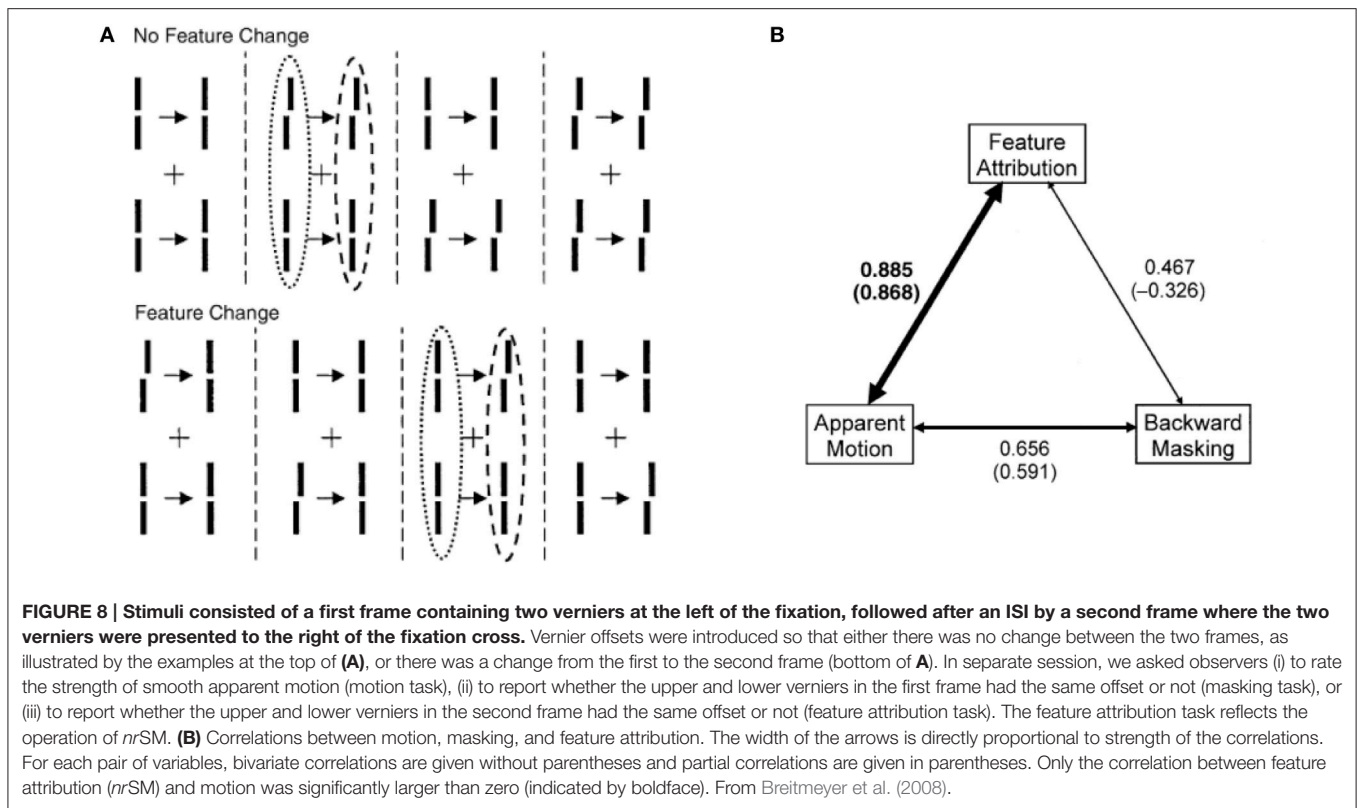


FIGURE 7 | The operation of *nrSM*. First, a motion analysis is carried out in the retinotopic areas and stimuli are grouped according to Gestalt grouping principles (e.g., common fate). In this example, two groups are formed, the rightward moving blue dots, and upward moving green dots. For each group, a common motion is extracted as the reference-frame and stimuli are transferred into non-retinotopic representations according to this reference-frame. The non-retinotopic representations are depicted at the top of the figure. *nrSM* stores information within this non-retinotopic representation.

but complementary roles of masking and motion, we determined the correlations between the non-retinotopic storage and integration in *nrSM* (we call this effect “feature attribution” since features are not perceived according to their retinotopic coordinates but are attributed non-retinotopically to motion streams), masking, and motion (Breitmeyer et al., 2008). The first frame contained a vernier pair presented to the left of the fixation cross. The second frame contained a vernier pair presented to

the right of the fixation cross. Offsets were introduced so that features (vernier offset) either changed or remained the same across the two frames (see **Figure 8**). In the feature-attribution task, subjects judged the vernier pair presented in the second frame and reported whether the upper and lower verniers in the second frame were the same or different (examples highlighted by dashed ovals). For example, the correct response is “same” for the rightmost stimulus sequence in the top panel of **Figure 8** and “different” for the rightmost stimulus sequence in the bottom panel. On trials in which feature attribution occurred, a larger number of misidentifications of the vernier pair presented in the second frame would be

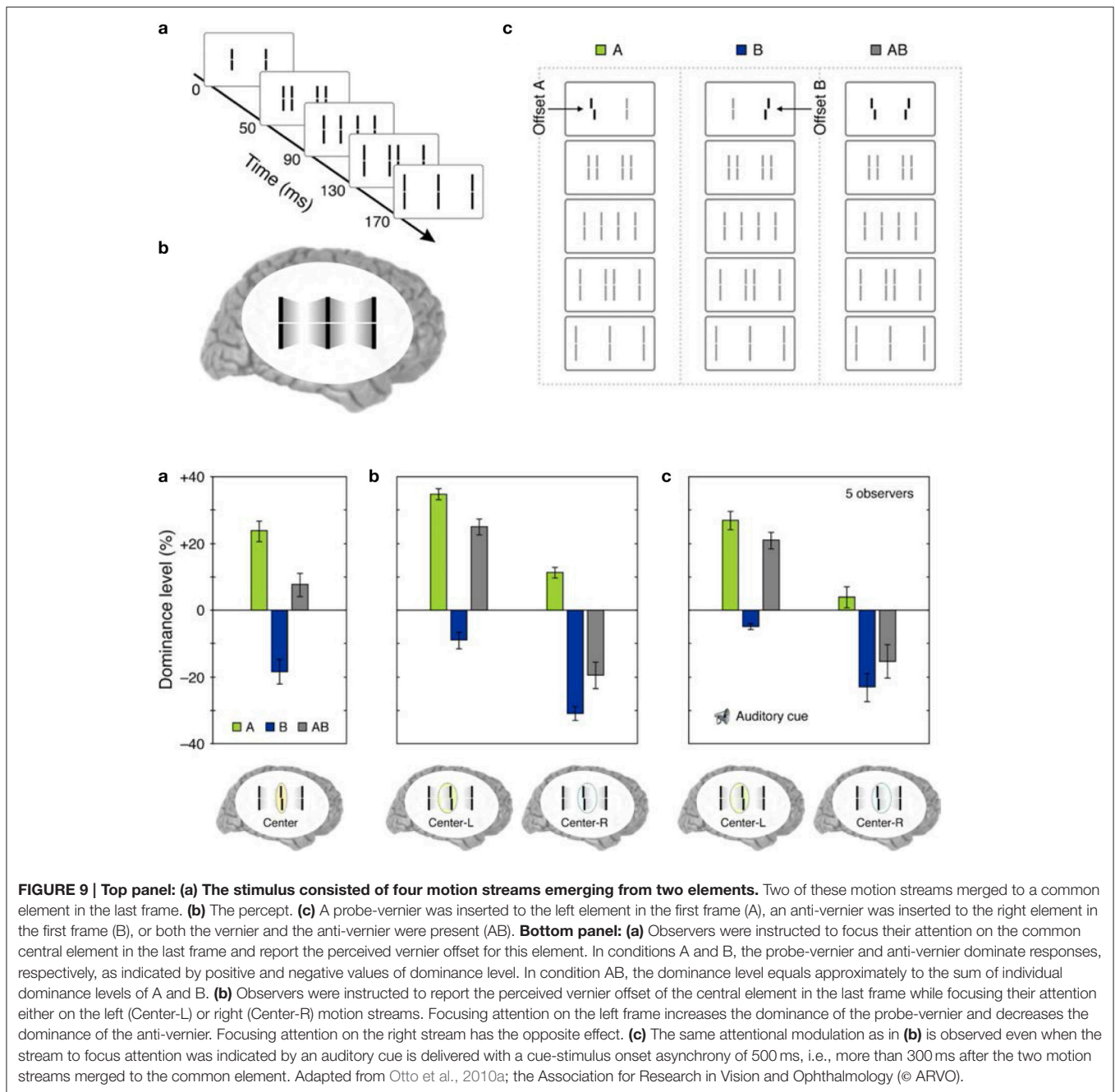
which is directly anchored in retinotopic coordinates (*rSM*) is needed. This is analogous to on and off channels in the visual system. Although these channels can be viewed as part of a single continuum of contrast, computationally, they involve different operations and are represented by separate distinct channels.



expected when the stimulus sequences had feature changes than when they did not. Hence, the difference between the numbers of misidentifications in the no-change and feature-change conditions provide an index of feature attribution, with larger differences corresponding to stronger feature attribution. Since feature attribution requires temporal storage and non-retinotopic integration, it measures *rSM*. In the backward-masking task, subjects judged the vernier pair in the first frame and reported whether the upper and lower verniers were the same or different (examples highlighted by dotted ovals). In the apparent motion task, observers rated the strength of smooth apparent motion, using a scale ranging from 0 (no motion) to 5 (optimal smooth motion). **Figure 8B** shows the correlations between these three variables computed across several values of stimulus-onset asynchronies between the two frames. As one can see from the figure, feature attribution correlated strongly with motion (significance: $p < 0.01$ for bivariate correlation and $p < 0.03$ for partial correlation) while the correlation of feature attribution with masking was weaker and not significant ($p > 0.175$ for bivariate correlation and $p < 0.295$ for partial correlation). Thus, these results support that the operation of *nrSM* has strong correlation with motion, which according to our theory constitutes its reference-frame, whereas the effect of masking is related to the operation of *rSM*.

(iv) *Retinotopic vs. non-retinotopic attention mechanisms:* Attention is a key process that controls the transfer of information from SM to STM and various lines of evidence

suggest that temporal dynamics of SM plays a fundamental role in determining how attention can select information from SM for transferring into STM (Wutz and Melcher, 2014). Attentional processes can be classified into two broad types, endogenous and exogenous (e.g., Posner, 1980; Jonides, 1981; Weichselgartner and Sperling, 1987; Müller and Rabbitt, 1989; Nakayama and MacKeben, 1989; Cheal and Lyon, 1991; Egeth and Yantis, 1997). Endogenous attention is a relatively slow process under voluntary control and its allocation to stimuli is flexible. It can be allocated to a static stimulus (fixed retinotopic location) as well as dynamic stimuli when we track for example a moving stimulus (changing retinotopic location; Pylyshyn and Storm, 1988). Exogenous attention is a relatively fast reflexive component. It has been shown that exogenous attention can also be deployed non-retinotopically according to the motion and motion-based perceptual of grouping of stimuli (Boi et al., 2011; Theeuwes et al., 2013; Gonen et al., 2014). Hence both endogenous and exogenous attention can operate in terms of retinotopic and non-retinotopic motion-grouping based coordinate systems and can control information flow from SM to STM. In Section Sequential Metacontrast: Non-retinotopic Information Storage And Processing, we discussed findings from sequential metacontrast with cueing, indicating that feature integration within a motion stream does not depend on the spatial allocation or the timing of attention. In the same study, we have also investigated a more complex stimulus where two motion streams merge to form a more complex



Gestalt (Figure 9). The stimulus consisted of four motion streams, two moving rightward and two moving leftward. Two of these streams merged at a common point. When observers were asked to report the vernier offset of this common point, the outcome did depend on the allocation of attention (Figure 9). The vernier offset in the attended stream dominated the outcome (Otto et al., 2010a). To summarize, nrSM has both pre-attentive and attentive components. Storage and integration of information within motion streams are pre-attentive whereas storage and integration of information across motion streams that merge (i.e., grouped into a more complex Gestalt) are

flexible and depend on attention. This modulatory effect of attention on non-retinotopic integration of information may be related to the findings of Cavanagh et al. (2008) who showed that attributes of a moving stimulus which is spatio-temporally embedded in a distractor stimulus can be integrated non-retinotopically when the moving stimulus is tracked by attention. A difference between Cavanagh et al. study and ours is that in their study color and motion attributes integrated non-retinotopically whereas letter and digit shapes did not. In our study, we showed integration for vernier offsets, which would imply integration for shapes. Future studies are needed to clarify this difference.

PARADOXES OF RETINOTOPIC SENSORY MEMORY REVISITED

Having introduced the new model, we can now compare it to the standard model containing only *rSM* and discuss what it predicts for the data that have been problematic for *rSM*.

Anorthoscopic Perception

One possible way *rSM* can account for anorthoscopic perception is to assume that the observers eyes move and hence different parts of the figure fall in different parts of the retina, building up over time a retinotopic image of the stimulus which can be stored by *rSM*. In fact, this is the “retinal painting” hypothesis which was put forth by von Helmholtz (1867). While it is possible to store an anorthoscopic stimulus in *rSM* via eye movements through gradually built-up retinotopic representations, numerous studies showed that anorthoscopic perception does occur in the absence of eye movements, i.e., without retinal painting, for example by moving stimuli in opposite directions (since the eyes cannot pursue simultaneously both stimuli) or by carefully monitoring eye movements during anorthoscopic perception (McCloskey and Watkins, 1978; Rock, 1981; Morgan et al., 1982; Fujita, 1990; Sohmiya and Sohmiya, 1992, 1994; Nishida, 2004; Fendrich et al., 2005; Rieger et al., 2007). In the absence of eye movements, since the stimulus moving behind the slit activates the *same* retinotopic area successively in time, these successive stimulations should be integrated together and stored in *rSM* as a meaningless blend of different parts. To explain anorthoscopic percepts, Parks (1965) proposed a non-retinotopic memory using the “time-of-arrival coding.” The storage in this memory is based, not on retinotopic coordinates, but on *temporal coordinates* with each stimulus part assuming as its coordinate its time-of-arrival. However, time-of-arrival theory was rejected by experimental studies that used a stimulus moving to the right and its mirror-image version moving to the left (McCloskey and Watkins, 1978; Sohmiya

and Sohmiya, 1992, 1994). **Figure 10** shows the stimulus. Two mirror-image symmetric triangular shapes composed of dots travel in opposite directions through the slit. Their timing and speed is arranged so that equivalent parts of the upper and lower triangles pass through the slit simultaneously. If time-of-arrival were the encoding principle in non-retinotopic memory, the upper and lower stimuli should appear identical since the arrival-times of their parts are identical⁴. However, observers report, not two identical triangles, but two mirror-image symmetric triangles.

We have proposed an alternative non-retinotopic process to explain anorthoscopic percepts (Ögmen, 2007; Aydın et al., 2008, 2009). The aforementioned experiment suggests that the critical variable is not the time-of-arrival of the stimulus but it is its direction of motion. As illustrated in **Figure 7**, we suggested that at a first stage motion vectors are extracted within the retinotopic slit region and these motion vectors (and not the time-of-arrival) serve as the reference-frame for *nrSM*. Accordingly, for the upper triangle, the leftward motion will be the reference-frame; whereas for the lower triangle rightward motion will be the reference frame. This allows the recovery and storage of the shape information into *nrSM*. Moreover, we made a novel prediction from our theory that shape distortions observed in anorthoscopic stimuli should be the result of differences in the perceived speeds of different parts of the stimuli. Our data provided support for this prediction (Aydın et al., 2008).

Hence, during anorthoscopic perception information is conveyed through *nrSM*, providing a solution to the paradox of anorthoscopic perception.

⁴Note that there is a slight difference in the way the individual disks arrive within the slit; according to time-of-arrival coding, the disks of the top triangle will be constructed from left to right whereas the disks of the bottom triangle will be constructed from right to left. However, in both cases, the same global triangular shape will emerge from these disks regardless whether the individual disks are constructed from left to right or right to left.

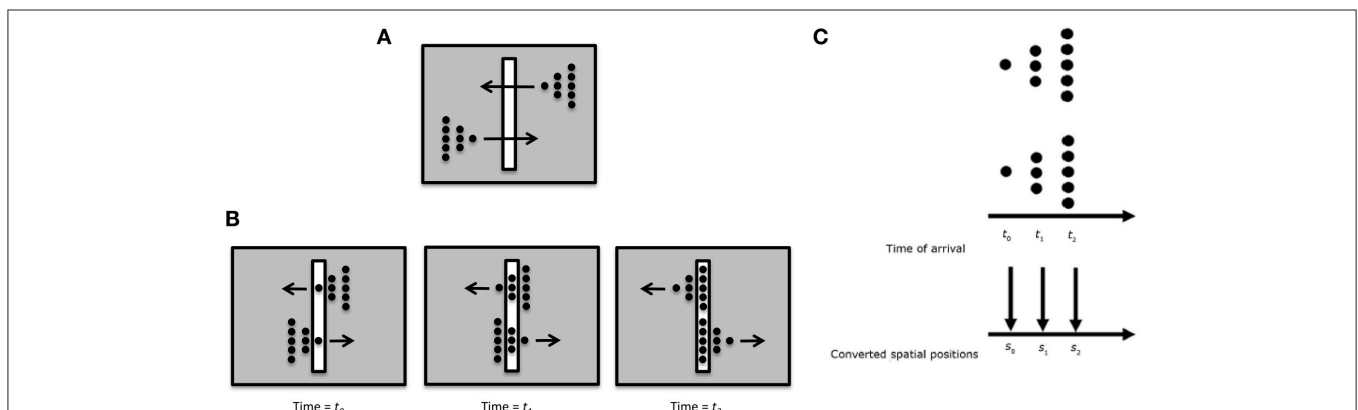
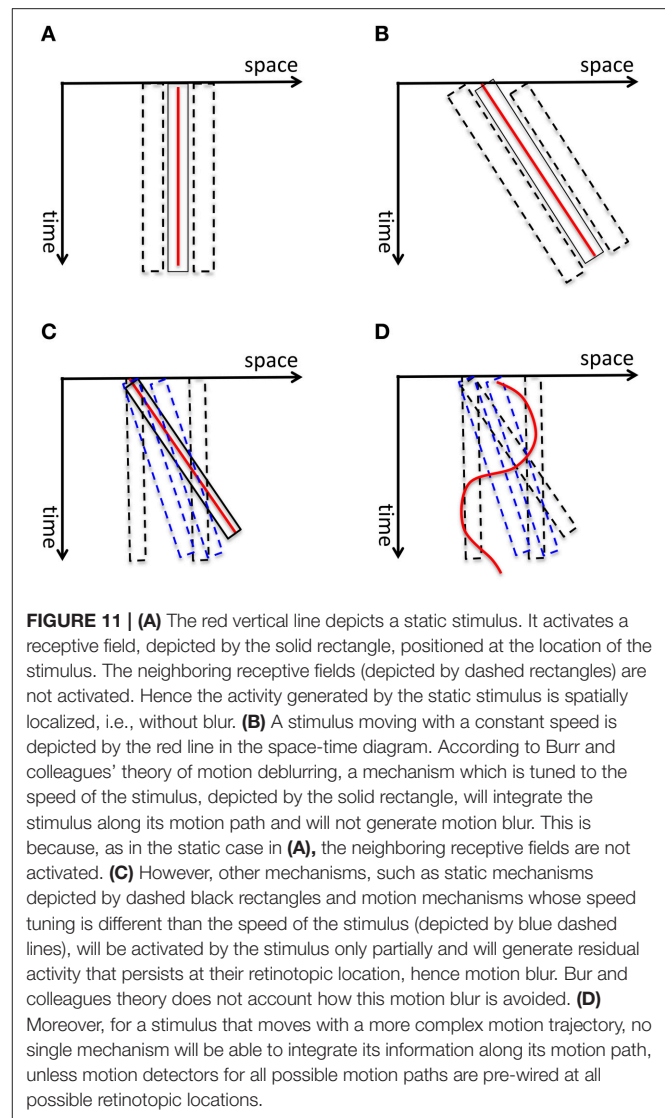


FIGURE 10 | (A) Stimulus configuration used to test the “retinal painting” account based on *rSM* and a non-retinotopic explanation based on “time-of-arrival” coding. **(B)** Equivalent parts of the upper and lower stimuli traverse through the slit synchronously. Since an eye movement will generate the same retinal painting for the upper and lower stimuli, the retinal painting account predicts that observers will perceive the same shape for the upper and lower stimuli. **(C)** The time-of-arrival account converts the arrival times to spatial positions. Since the equivalent parts of the upper and lower stimuli are synchronized, the time-of-arrival account also predicts that observers will perceive the same shape for the upper and lower stimuli. Contrary to these predictions, observers perceive two mirror-image triangles.

How the Visual System Controls Deleterious Effects of *rSM*: Motion Deblurring

As mentioned before, the visible component (i.e., visible persistence) of the retinotopic sensory memory retains information for about 120 ms under normal viewing conditions (Coltheart, 1980). Based on this estimate, one would expect moving objects to appear highly smeared; however, our typical perception is relatively sharp and clear (e.g., Ramachandran et al., 1974; Burr, 1980; Hogben and Di Lollo, 1985; Castet, 1994; Bex et al., 1995; Westerink and Teunissen, 1995; Burr and Morgan, 1997; Hammett, 1997). This leads to two fundamental questions: (i) how does the visual system generate and store clear percepts if no meaningful information is conveyed by the SM stage of the modal model? and (ii) how does it avoid motion blur; i.e., how does the visual system control deleterious effects of *rSM* [Q2 in Section Sensory (Iconic) Memory]?

Burr and colleagues measured the perceived extent of motion blur produced by a field of moving dots and showed that it increases as a function of exposure duration up to 40 ms after which it decreases, a phenomenon called *motion deblurring* (Burr, 1980; Burr and Morgan, 1997). They proposed that spatiotemporally-oriented receptive-fields of motion mechanisms can account for motion deblurring since these receptive fields can collect information along the motion path of the object (Burr and Morgan, 1997). Hence, according to this theory, the computation of form for moving objects is carried out by motion mechanisms. To clarify this concept, consider first the space-time diagram shown in **Figure 11A**. The red line represents a static stimulus (since its position with respect to the horizontal space-axis is fixed). It will activate a receptive field collecting information from this position over time (depicted by the solid rectangle). Neighboring receptive fields (depicted by dashed rectangles) will not be activated since the stimulus does not fall within their “space-time window.” Hence, the activity generated by the static stimulus will be spatially localized without any blur. A stimulus moving with a constant speed can be represented by an oriented line in the space-time diagram (the red line in **Figure 11B**). Motion-sensitive neurons can be described by spatio-temporally oriented receptive fields (Adelson and Bergen, 1985). In terms of motion mechanisms that are tuned to the velocity of the stimulus, only one motion mechanism will be activated. Since the case in **Figure 11B** is equivalent to the static case in **Figure 11A** with a rotation, Burr and colleagues argued that the stimulus will not generate motion blur provided that it remains with the receptive field of the matching motion detector (the solid rectangle in **Figure 11B**) to sufficiently activate it. However, this theory fails to explain the following: As depicted in **Figure 11C**, mechanisms whose receptive fields are not aligned with the motion path of the object (e.g., motion detectors tuned to different speeds than the speed of the moving object; mechanisms that are not tuned to motion) will be partially activated by the moving object and will generate extensive blur. This theory cannot explain how this blur is avoided by the visual system. Furthermore, since object trajectories can be arbitrarily complex, a fixed set of oriented receptive fields cannot guarantee that a



match between object motion and receptive field profile would occur in general (**Figure 11D**).

Kahneman et al. (1992) proposed the object-file theory to explain how the attributes of moving objects can be computed. According to this theory, an “object-file” is opened and the attributes of the moving object are inserted into this file. Since this insertion can take place over multiple retinotopic locations over the motion path, the theory could in principle provide an answer to question Q1. However, the theory gives no details, nor mechanisms to explain how object files are opened and information is inserted over the motion pathway. The theory does not answer questions Q2 and Q3 either.

Contrary to the predictions of these two theories, it has been long known that isolated targets in motion do exhibit extensive blur (Bidwell, 1899; McDougall, 1904; Dixon and Hammond, 1972; Farrell, 1984; Di Lollo and Hogben, 1985; Farrell et al., 1990). In order to reconcile the apparently contradictory observations of motion deblurring for a field of moving dots

and extensive blur for isolated moving targets, we conducted experiments where we showed that (data, modeling, and review: Chen et al., 1995; Purushothaman et al., 1998; Öğmen, 2007): (1) isolated targets moving on a uniform background are perceived with extensive motion blur; (2) the presence of spatio-temporally proximal stimuli can reduce the spatial extent of perceived motion blur (motion deblurring); (3) motion mechanisms cannot account for motion deblurring; (4) metacontrast mechanisms can account for motion deblurring. Hence, to put these results in the context of our model in **Figure 6**, when isolated targets are in motion *rSM* becomes active and its side-effect, motion blur, is perceived. On the other hand, in the presence of spatiotemporally proximal stimuli, masking “turns off” *rSM* and no motion blur is perceived. Thus, the answer to the question Q2 is: visual masking. While earlier analyses also acknowledged that visual masking can turn *rSM* off under most ecologically valid viewing conditions, this observation led to a paradox: If the contents of *rSM* are suppressed during natural viewing, no information can be conveyed to WM and LTM, making the whole memory system inoperational under normal viewing conditions! Our theory offers a solution to this paradox by suggesting that information is conveyed to STM/WM and LTM through *nrSM*.

What Purpose Does *rSM* Serve?

As mentioned in the previous section, data showing that isolated moving-targets do generate motion blur indicate that *rSM* cannot be completely removed from SM, but its deleterious effects for dynamic viewing conditions are in general controlled by visual masking mechanisms. This leads to a more general question: If *rSM* cannot be eliminated from SM, is it a simple side-effect or does it serve a purpose? Ecological viewing consists of periods of fixations, saccades, pursuit, and vergence eye movements. The head and the body of the observer can be also in movement and vestibulo-ocular movements can compensate for some but not all retinotopic motions generated by these movements. For example, when the observer moves, the eyes may reflexively compensate for these movements to keep their positions on the fixation point of interest, thereby stabilizing the fixation point. However, observer’s movements can also generate motion parallax for the rest of the scene and the amount of motion for different parts of the stimulus depends on the depth of objects relative to the observer. Hence, under normal viewing conditions, the retinotopic stimulus typically contains both static and dynamic components. As we have noted earlier (Footnote 2), static stimuli have a null velocity vector and their reference-frame is equivalent to a retinotopic reference frame. From a mechanistic point of view, if *nrSM* uses the activities of motion detectors to synthesize its reference-frame, in the case of static stimuli, there will be no motion-detector activity to generate the reference-frame. Hence to store information about static stimuli, a memory component that is directly anchored in retinotopic coordinates is needed and this memory component is *rSM*. Within *nrSM*, there can be multiple reference-frames deployed at different parts of the visual field depending on the motion patterns across the visual field. Hence, our theory suggests that sensory memory operates according to retinotopic motion patterns and *rSM* is a special case

with a reference-frame corresponding to the null velocity. From this perspective, information can flow simultaneously from *rSM* and *nrSM*, the former carrying out the information about fixated stimuli and the rest of the scene which is static with respect to the fixated stimuli, whereas *nrSM* conveys information about objects that are in relative motion with respect to the fixation target.

DISCUSSION

Sensory memory was discovered in 1960s and, by the end of the decade, it became an important and integral part of the modal model of human memory. However, about two decades after its discovery, Haber placed it on a death-bed and suggested that the concept should be removed from textbooks (Haber, 1983). The inability of the sensory memory to operate under normal viewing conditions not only challenged any role it may have in information processing, but also positioned it as a “road block” to information flow from external inputs to the rest of the modal model. However, during the last decade, evidence has been accumulating on non-retinotopic processing for various stimulus attributes such as form (Nishida, 2004; Öğmen et al., 2006; Otto et al., 2006; Öğmen and Herzog, 2010), luminance (Shimozaki et al., 1999), color (Nishida et al., 2007), size (Kawabe, 2008), and motion (Boi et al., 2009; Noory et al., 2015a). We suggest that this non-retinotopic processing extends to sensory memory in the form of non-retinotopic sensory memory (*nrSM*). Furthermore, we have also shown that attention, a key process in the transfer of information from SM to STM, also operates on motion-based non-retinotopic coordinates (Boi et al., 2009, 2011). Based on these findings, we proposed here a new model for SM and discussed how it can resolve the paradoxes that stem from the Achilles’ heel of the traditional SM, namely its retinotopic basis.

The traditional SM has been conceptualized as a low-level, image-like representation. However, our results and model suggest that grouping operations already take place in SM. One can also trace the roots of processing stages, such as object permanence and invariance, hitherto thought to take place at higher levels, already in SM. Having a flexible motion-based reference-frame makes this memory position-invariant. Moreover, the ability to carry information across occlusions plays a key role in achieving object permanence. Having these properties already at the SM level does make sense if one considers the ecology of vision. Gestalt psychologists have long argued that atomistic approaches, which build complex percepts by gradually combining simpler ones, cannot handle the complexity of our visual environment and grouping operations need to take place early on. Gibsonian ecological optics (Gibson, 1979) emphasizes the importance of motion in a natural environment. Duncker’s (1929) and Johansson’s (1975) work provided several examples of relativity of motion and the underlying motion-based reference frames (reviews: Mack, 1986; Öğmen and Herzog, 2015). Our new model for sensory memory combines these concepts and suggests how memory systems can be interfaced to our natural environment.

AUTHOR CONTRIBUTIONS

HO, Developed the theory; wrote the original manuscript draft.
MH, Developed the theory; read and commented on the original manuscript draft.

REFERENCES

- Adelson, E. H., and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2, 284–299. doi: 10.1364/JOSAA.2.000284
- Agaoglu, M. N., Herzog, M. H., and Ögmen, H. (2012). Non-retinotopic feature processing in the absence of retinotopic spatial layout and the construction of perceptual space from motion. *Vision Res.* 71, 10–17. doi: 10.1016/j.visres.2012.08.009
- Atkinson, R. C., and Shiffrin, R. M. (1968). “Human memory: a proposed system and its control processes,” in *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed K. W. Spence (New York, NY: Academic Press) 89–195.
- Averbach, E., and Coriell, A. (1961). Short-term memory in vision. *Bell Syst. Tech. J.* 40, 309–328. doi: 10.1002/j.1538-7305.1961.tb03987.x
- Averbach, E., and Sperling, G. (1961). “Short-term storage of information in vision,” in *Information Theory*, ed C. Cherry (London: Butterworth), 196–211.
- Aydin, M., Herzog, M. H., and Ögmen, H. (2008). Perceived speed differences explain apparent compression in slit viewing. *Vision Res.* 48, 1603–1612. doi: 10.1016/j.visres.2008.04.020
- Aydin, M., Herzog, M. H., and Ögmen, H. (2009). Shape distortions and Gestalt grouping in anorthoscopic perception. *J. Vis.* 9, 8.1–8.8. doi: 10.1167/9.3.8
- Bachmann, T. (1994). *Psychophysiology of Visual Masking: The Fine Structure of Conscious Experience*. New York, NY: Nova Science Publishers.
- Baddeley, A. (2007). *Working Memory, Thought, and Action*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198528012.001.0001
- Bex, P. J., Edgar, G. K., and Smith, A. T. (1995). Sharpening of blurred drifting images. *Vision Res.* 35, 2539–2546. doi: 10.1016/0042-6989(95)00060-D
- Bidwell, S. (1899). *Curiosities of Light and Sight*. London: Swan Sonnenschein.
- Boi, M., Ögmen, H., and Herzog, M. H. (2011). Motion and tilt after effects occur largely in retinal, not in object coordinates, in the Ternus-Pikler display. *J. Vis.* 11, 1–11. doi: 10.1167/11.3.7
- Boi, M., Ögmen, H., Krummenacher, J., Otto, T. U., and Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *J. Vis.* 9.5. doi: 10.1167/9.13.5
- Breitmeyer, B. G., and Ögmen, H. (2000). Recent models and findings in backward visual masking: a comparison, review, and update. *Percept. Psychophys.* 62, 1572–1595. doi: 10.3758/BF03212157
- Breitmeyer, B. G., and Ögmen, H. (2006). *Visual Masking: Time Slices Through Conscious and Unconscious Vision, 2nd Edn*. Oxford, UK: Oxford University Press.
- Breitmeyer, B. G., Herzog, M. H., and Ögmen, H. (2008). Motion, not masking, provides the medium for feature attribution. *Psychol. Sci.* 19, 823–829. doi: 10.1111/j.1467-9280.2008.02163.x
- Breitmeyer, B. G., Kropfl, W., and Julesz, B. (1982). Tex existence and role of retinotopic and spatiotopic forms of visual persistence. *Acta Psychol.* 52, 175–196. doi: 10.1016/0001-6918(82)90007-5
- Burr, D. (1980). Motion smear. *Nature* 284, 164–165. doi: 10.1038/284164a0
- Burr, D. C., and Morgan, M. J. (1997). Motion deblurring in human vision. *Proc. Soc. Lond. B* 264, 431–436. doi: 10.1098/rspb.1997.0061
- Castet, E. (1994). Effect of the ISI on the visible persistence of a stimulus in apparent motion. *Vision Res.* 34, 2103–2114. doi: 10.1016/0042-6989(94)90320-4
- Cavanagh, P., Holcombe, A. O., and Chou, W. (2008). Mobile computation: spatiotemporal integration of the properties of objects in motion. *J. Vis.* 8, 1–23. doi: 10.1167/8.12.1
- Cheal, M., and Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *Q. J. Exp. Psychol.* 43A, 859–880. doi: 10.1080/14640749108400960
- Chen, S., Bedell, H. E., and Ögmen, H. (1995). A target in real motion appears blurred in the absence of other proximal moving targets. *Vision Res.* 35, 2315–2328. doi: 10.1016/0042-6989(94)00308-9
- Coltheart, M. (1980). Iconic memory and visible persistence. *Percept. Psychophys.* 27, 183–228. doi: 10.3758/BF03204258
- Davidson, M. L., Fox, M. J., and Dick, A. O. (1973). Effect of eye movements on backward masking and perceived location. *Percept. Psychophys.* 14, 110–116. doi: 10.3758/BF03198624
- Di Lollo, V., and Hogben, J. H. (1985). Suppression of visible persistence. *J. Exp. Psychol.* 11, 304–316. doi: 10.1037/0096-1523.11.3.304
- Dixon, N. F., and Hammond, E. J. (1972). The attenuation of visual persistence. *Br. J. Psychol.* 63, 243–254. doi: 10.1111/j.2044-8295.1972.tb02107.x
- Duncker, K. (1929). Über induzierte Bewegung. *Psychol. Forsch.* 22, 180–259. doi: 10.1007/BF02409210
- Egeth, H. E., and Yantis, S. (1997). Visual attention: control, representation, and time course. *Annu. Rev. Psychol.* 48, 269–297. doi: 10.1146/annurev.psych.48.1.269
- Farrell, J. E. (1984). Visible persistence of moving objects. *J. Exp. Psychol.* 10, 502–511. doi: 10.1037/0096-1523.10.4.502
- Farrell, J. E., Pavel, M., and Sperling, G. (1990). Visible persistence of stimuli in stroboscopic motion. *Vision Res.* 30, 921–936. doi: 10.1016/0042-6989(90)90058-S
- Fendrich, R., Rieger, J. W., and Heinze, H.-J. (2005). The effect of retinal stabilization on anorthoscopic percepts under free-viewing conditions. *Vision Res.* 45, 567–582. doi: 10.1016/j.visres.2004.09.025
- Fujita, N. (1990). Three-dimensional anorthoscopic perception. *Perception* 19, 767–771. doi: 10.1068/p190767
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
- Gonen, F. F., Hallal, H., and Ögmen, H. (2014). Facilitation by exogenous attention for static and dynamic gestalt groups. *Atten. Percept. Psychophys.* 76, 1709–1720. doi: 10.3758/s13414-014-0679-2
- Griffin, I. C., and Nobre, A. C. (2003). Orienting attention to locations in internal representations. *J. Cogn. Neurosci.* 15, 1176–1194. doi: 10.1162/089892903322598139
- Haber, R. N. (1983). The impending demise of the icon: a critique of the concept of iconic storage in visual information processing. *Behav. Brain Sci.* 6, 1–54.
- Hammett, S. T. (1997). Motion blur and motion sharpening in the human visual system. *Vision Res.* 37, 2505–2510. doi: 10.1016/S0042-6989(97)00059-X
- Hogben, J. H., and Di Lollo, V. (1985). Suppression of visible persistence in apparent motion. *Percept. Psychophys.* 38, 450–460. doi: 10.3758/BF03207176
- Hollingworth, A., and Maxcey-Richard, A. M. (2013). Selective maintenance in visual working memory does not require sustained visual attention. *J. Exp. Psychol.* 39, 1047–1058. doi: 10.1037/a0030238
- Irwin, D. E., Brown, J. S., and Sun, J.-S. (1988). Visual masking and visual integration across saccadic eye movements. *J. Exp. Psychol.* 117, 276–287. doi: 10.1037/0096-3445.117.3.276
- Irwin, D. E., Yantis, S., and Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Percept. Psychophys.* 34, 49–57.
- Johansson, G. (1975). Visual motion perception. *Sci. Am.* 232, 76–88. doi: 10.1038/scientificamerican0675-76
- Jonides, J. (1981). “Voluntary vs. Automatic control over the mind’s eye’s movement,” in *Attention and Performance IX*, eds J. Long and A. Baddeley (Hillsdale, MI: Erlbaum), 187–203.

ACKNOWLEDGMENTS

MH is supported by the Swiss National Science Foundation (SNF) Project “Basics of visual processing: from retinotopic encoding to non-retinotopic representations.”

- Jonides, J., Irwin, D. E., and Yantis, S. (1982). Integrating visual information from successive fixations. *Science* 215, 192–194. doi: 10.1126/science.7053571
- Jonides, J., Irwin, D. E., and Yantis, S. (1983). Failure to integrate information from successive fixations. *Science* 222, 188. doi: 10.1126/science.6623072
- Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219. doi: 10.1016/0010-0285(92)90007-o
- Kawabe, T. (2008). Spatiotemporal feature attribution for the perception of visual size. *J. Vis.* 8, 7.1–7.9. doi: 10.1167/8.7
- Mack, A. (1986). “Perceptual aspects of motion in the frontal plane,” in *Handbook of Perception and Human Performance*, eds K. R. Boff, L. Kaufman, and J. P. Thomas (New York, NY: John Wiley and Sons), 17.1–17.38.
- Makovski, T. (2012). Are multiple visual short-term memory storages necessary to explain the retro-cue effect? *Psychon. Bull. Rev.* 19, 470–476. doi: 10.3758/s13423-012-0235-9
- Matsukura, M., and Hollingworth, A. (2011). Does visual short-term memory have a high-capacity stage? *Psychon. Bull. Rev.* 18, 1098–1104. doi: 10.3758/s13423-011-0153-2
- McCloskey, M., and Watkins, M. J. (1978). The seeing- more-than-is-there phenomenon: implications for the locus of iconic storage. *J. Exp. Psychol.* 4, 553–564. doi: 10.1037/0096-1523.4.4.553
- McDougall, W. (1904). The sensations excited by a single momentary stimulation of the eye. *Br. J. Psychol.* 1, 78–113. doi: 10.1111/j.2044-8295.1904.tb00150.x
- McRae, K., Butler, B. E., and Popiel, S. J. (1987). Spatiotopic and retinotopic components of iconic memory. *Psychol. Res.* 49, 221–227. doi: 10.1007/BF00309030
- Morgan, M. J., Findlay, J. M., and Watt, R. J. (1982). Aperture viewing: a review and a synthesis. *Q. J. Exp. Psychol.* 34A, 211–233. doi: 10.1080/14640748208400837
- Müller, H. J., and Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *J. Exp. Psychol.* 15, 315–330. doi: 10.1037/0096-1523.15.2.315
- Nakayama, K., and MacKeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Res.* 11, 1631–1647. doi: 10.1016/0042-6989(89)90144-2
- Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Curr. Biol.* 14, 830–839. doi: 10.1016/j.cub.2004.04.044
- Nishida, S., Watanabe, J., Kuriki, I., and Tokimoto, T. (2007). Human brain integrates colour signals along motion trajectory. *Curr. Biol.* 17, 366–372. doi: 10.1016/j.cub.2006.12.041
- Noory, B., Herzog, M. H., and Ögmen, H. (2015a). Spatial properties of non-retinotopic reference frames in human vision. *Vision Res.* 113, 44–54. doi: 10.1016/j.visres.2015.05.010
- Noory, B., Herzog, M. H., and Ögmen, H. (2015b). Retinotopy of visual masking and non-retinotopic perception during masking. *Atten. Percept. Psychophys.* 77, 1263–1284. doi: 10.3758/s13414-015-0844-2
- Ögmen, H. (2007). A theory of moving form perception: synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Adv. Cogn. Psychol.* 3, 67–84. doi: 10.2478/v10053-008-0015-2
- Ögmen, H., and Herzog, M. H. (2010). The geometry of visual perception: retinotopic and non-retinotopic representations in the human visual system. *Proc. IEEE Inst. Electr. Electron. Eng.* 98, 479–492. doi: 10.1109/JPROC.2009.2039028
- Ögmen, H., and Herzog, M. H. (2015). “Apparent motion and reference frames,” in *Oxford Handbook of Perceptual Organization*, ed J. Wagemans (Oxford, UK: Oxford University Press), 487–503.
- Ögmen, H., Otto, T., and Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Res.* 46, 3234–3242. doi: 10.1016/j.visres.2006.04.007
- Otto, T. U., Ögmen, H., and Herzog, M. H. (2006). The flight path of the phoenix—the visible trace of invisible elements in human vision. *J. Vis.* 6, 1079–1086. doi: 10.1167/6.10.7
- Otto, T. U., Ögmen, H., and Herzog, M. H. (2008). Assessing the microstructure of motion correspondences with non-retinotopic feature attribution. *J. Vis.* 8, 16.1–16.15. doi: 10.1167/8.7.16
- Otto, T. U., Ögmen, H., and Herzog, M. H. (2009). Feature integration across space, time, and orientation. *J. Exp. Psychol.* 35, 1670–1686. doi: 10.1037/a0015798
- Otto, T. U., Ögmen, H., and Herzog, M. H. (2010a). Attention and non-retinotopic feature integration. *J. Vis.* 10:8. doi: 10.1167/10.12.8
- Otto, T. U., Ögmen, H., and Herzog, M. H. (2010b). Perceptual learning in a nonretinotopic frame of reference. *Psychol. Sci.* 21, 1058–1063. doi: 10.1177/0956797610376074
- Pantle, A., and Picciano, L. (1976). A multistable movement display: evidence for two separate motion systems in human vision. *Science* 193, 500–502. doi: 10.1126/science.941023
- Parks, T. E. (1965). Post-retinal visual storage. *Am. J. Psychol.* 78, 145–147. doi: 10.2307/1421101
- Piéron, H. (1935). Le processus du métacontraste. *J. Psychol. Normale Pathol.* 32, 1–24.
- Pikler, J. (1917). *Sinnesphysiologische Untersuchungen*. Leipzig: Barth.
- Pinto, Y., Sligte, I. G., Shapiro, K. L., and Lamme, V. A. F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychon. Bull. Rev.* 20, 732–739. doi: 10.3758/s13423-013-0393-4
- Plateau, J. (1836). Notice sur l’anarthroscope. *Bull. Acad. R. Sci. Belles Lett. Bruxelles* 3, 7–10.
- Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25. doi: 10.1080/00335558008248231
- Purushothaman, G., Ögmen, H., Chen, S., and Bedell, H. E. (1998). Motion deblurring in a neural network model of retino-cortical dynamics. *Vision Res.* 38, 1827–1842. doi: 10.1016/S0042-6989(97)00350-7
- Pylyshyn, Z. W., and Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197. doi: 10.1163/156856888X00122
- Ramachandran, V. S., Rao, V. M., and Vidyasagar, T. R. (1974). Sharpness constancy during movement perception. *Perception* 3, 97–98. doi: 10.1068/p030097
- Rayner, K., and Pollatsek, A. (1983). Is visual information integrated across saccades? *Percept. Psychophys.* 34, 39–48. doi: 10.3758/BF03205894
- Rerko, L., Souza, A. A., and Oberauer, K. (2014). Retro-cue benefits in working memory without sustained focal attention. *Mem. Cogn.* 42, 712–728. doi: 10.3758/s13421-013-0392-8
- Rieger, J. W., Grüşchow, M., Heinze, H.-J., and Fendrich, R. (2007). The appearance of figures seen through a narrow aperture under free viewing conditions: effects of spontaneous eye motions. *J. Vis.* 7:10. doi: 10.1167/7.6.10
- Ritter, M. (1976). Evidence for visual persistence during saccadic eye movements. *Psychol. Res.* 39, 67–85. doi: 10.1007/BF00308946
- Rock, I. (1981). Anorthoscopic perception. *Sci. Am.* 244, 145–153. doi: 10.1038/scientificamerican0381-145
- Scharnowski, F., Hermens, F., Kammer, T., Ögmen, H., and Herzog, M. H. (2007). Feature fusion reveals slow and fast memories. *J. Cogn. Neurosci.* 19, 632–641. doi: 10.1162/jocn.2007.19.4.632
- Shimozaki, S., Eckstein, M. P., and Thomas, J. P. (1999). The maintenance of apparent luminance of an object. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1433–1453. doi: 10.1037/0096-1523.25.5.1433
- Sligte, I. G., Scholte, H. S., and Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE* 3:e0001699. doi: 10.1371/journal.pone.0001699
- Sohmiya, T., and Sohmiya, K. (1992). Where does an anorthoscopic image appear? *Percept. Mot. Skills* 75, 707–714. doi: 10.2466/pms.1992.75.3.707
- Sohmiya, T., and Sohmiya, K. (1994). What is a crucial determinant in anorthoscopic perception? *Percept. Mot. Skills* 78, 987–998. doi: 10.2466/pms.1994.78.3.987
- Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29.
- Sun, J.-S., and Irwin, D. E. (1987). Retinal masking during pursuit eye movements: implications for spatiotopic visual persistence. *J. Exp. Psychol.* 13, 140–145. doi: 10.1037/0096-1523.13.1.140
- Ternus, J. (1926). Experimentelle Untersuchung über phänomenale Identität. *Psychol. Forsch.* 7, 81–136. doi: 10.1007/BF02424350

- Theeuwes, J., Mathôt, S., and Grainger, J. (2013). Exogenous object-centered attention. *Atten. Percept. Psychophys.* 75, 812–818. doi: 10.3758/s13414-013-0459-4
- Vandenbroucke, A. R. E., Sligte, I. G., Barrett, A. B., Seth, A. K., Fahrenfort, J. J., and Lamme, V. A. F. (2014). Accurate metacognition for visual sensory memory representations. *Psychol. Sci.* 25, 861–873. doi: 10.1177/0956797613516146
- van der Heijden, A. H. C., Bridgeman, B., and Mewhort, D. J. K. (1986). Is stimulus persistence affected by eye movements? A critique of Davidson, Fox, and Dick (1973). *Psychol. Res.* 48, 179–181. doi: 10.1007/BF00309166
- van Moorselaar, D., Olivers, C. N. L., Theeuwes, J., Lamme, V. A. F., and Sligte, I. G. (2015). Forgotten but not gone: Retro-cue costs and benefits in a double-cueing paradigm suggest multiple states in visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1755–1763. doi: 10.1037/xlm0000124
- von Helmholtz, H. (1867). *Handbook of Physiological Optics*. New York, NY: Dover. Reprint in 1962.
- Weichselgartner, E., and Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science* 238, 778–780. doi: 10.1126/science.3672124
- Westerink, J. H. D. M., and Teunissen, K. (1995). Perceived sharpness in complex moving images. *Displays* 16, 89–97. doi: 10.1016/0141-9382(95)91178-5
- White, C. W. (1976). Visual masking during pursuit eye movements. *J. Exp. Psychol.* 2, 469–478. doi: 10.1037/0096-1523.2.4.469
- Wolfe, W., Hauske, G., and Lupp, U. (1978a). How pre-saccadic gratings modify post-saccadic modulation transfer functions. *Vision Res.* 18, 1173–1179. doi: 10.1016/0042-6989(78)90101-3
- Wolfe, W., Hauske, G., and Lupp, U. (1978b). Interaction of pre- and post-saccadic patterns having the same co-ordinates in visual space. *Vision Res.* 20, 117–125. doi: 10.1016/0042-6989(80)90153-4
- Wutz, A., and Melcher, D. (2014). The temporal window of individuation limits visual capacity. *Front. Psychol.* 5:952. doi: 10.3389/fpsyg.2014.00952
- Zöllner, F. (1862). Über eine neue art anorthoskopischer zerrbilder. *Ann. Phys. Chem.* 117, 477–484. doi: 10.1002/andp.18621931108

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ögmen and Herzog. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.