If You Watch It Move, You'll Recognize It in 3D:

Transfer of Depth Cues Between Encoding and Retrieval

Frank Papenmeier

University of Tübingen, Germany

Stephan Schwan

Leibniz-Institut für Wissensmedien, Tübingen, Germany

Author Note

Correspondence concerning this article should be addressed to Frank Papenmeier,

University of Tübingen, Schleichstr. 4, D-72076 Tübingen, Germany. E-mail:

frank.papenmeier@uni-tuebingen.de

Word count: 4944 (Abstract: 146)

Abstract

Viewing objects with stereoscopic displays provides additional depth cues through binocular disparity supporting object recognition. So far, it was unknown whether this results from the representation of specific stereoscopic information in memory or a more general representation of an object's depth structure. Therefore, we investigated whether continuous object rotation acting as depth cue during encoding results in a memory representation that can subsequently be accessed by stereoscopic information during retrieval. In Experiment 1, we found such transfer effects from continuous object rotation during encoding to stereoscopic presentations during retrieval. In Experiments 2a and 2b, we found that the continuity of object rotation is important because only continuous rotation and/or stereoscopic depth but not multiple static snapshots presented without stereoscopic information caused the extraction of an object's depth structure into memory. We conclude that an object's depth structure and not specific depth cues are represented in memory.

Keywords: object recognition; stereoscopic displays; memory; depth perception; depth from motion; binocular disparity

Imagine you finally get your hands on your new smart phone. What would you do? Certainly as one of the first things, you would turn it around, look at the back of it, its color, the thickness, and so on, thereby building up a mental model of the new phone. Once you put it down on the table, go away and come back, you would still recognize the object on the table as your new phone. Thus, you must have built up a memory representation of it. But what is the nature of this representation and what information does it contain, in particular regarding an object's depth structure? With the present manuscript, we investigate this question by studying transfer effects of depth from motion during encoding, to depth from binocular disparity during retrieval. In particular, we examine whether rotating an object during encoding supports the extraction of an object's depth structure into memory that can subsequently be accessed by stereoscopic information. Anticipating the results of our present experiments, we found evidence for transfer between continuous object rotation as a depth cue during encoding and stereoscopic information as a depth cue during retrieval, thus showing that the memory representation supporting object recognition is not flat but contains depth information and that it represents the depth structure of objects instead of specific depth cues.

The nature of memory representations supporting object recognition has been the subject of a long standing debate, in particular regarding the mechanisms by which memory representations support the recognition of objects from novel viewpoints (e.g., Biederman & Gerhardstein, 1993; Tarr & Bülthoff, 1995). There are two main lines of theories. On the one hand, there are theories considering object memory as a representation of object structure, be it complete three-dimensional models of objects (Marr & Nishihara, 1978) or the composition of viewpoint-independent primitives (Biederman, 1987; Hummel & Biederman, 1992). On the other hand, there are theories that consider object memory as a representation of objects in the

form of multiple 2D view-specific representations as seen during encoding (Bülthoff, Edelman, & Tarr, 1995; Ullman & Basri, 1991). Due to the opposing predictions regarding the viewpoint dependence of object recognition, much work was devoted to determining the conditions for either the viewpoint-dependent or viewpoint-independent representation of objects (e.g., Biederman & Bar, 1999; Biederman & Gerhardstein, 1993; Edelman & Bülthoff, 1992; Farah, Rochlin, & Klein, 1994; Hayward & Tarr, 1997; Rock & DiVita, 1987; Tarr & Bülthoff, 1995; Tarr & Pinker, 1990; Tarr, Williams, Hayward, & Gauthier, 1998). As it turns out, both lines of theories found a way to explain the spectrum of viewpoint-dependent and viewpoint-independent findings by either arguing that the degree of viewpoint dependence depends on the amount that the structural description of objects changes by object rotation (Biederman & Gerhardstein, 1993) or the extent to which interpolation and extrapolation processes can operate on the 2D view-specific representations (Bülthoff et al., 1995). Therefore, this line of research provided the important insight that object recognition is viewpoint-dependent under many conditions; however, it still remains unclear what information observers represent in memory, in particular regarding the depth structure of objects.

Studies investigating the role of depth on object memory often used stereoscopic presentations and therefore binocular disparity as a depth cue (e.g., Bennett & Vuong, 2006; Burke, 2005). When presenting an object stereoscopically, two slightly different images are presented to the left and right eye of the observer, resembling the different views of the eyes when viewing real objects. Binocular disparity resulting from those views acts a as strong depth cue during perception (Dövencioğlu, Ban, Schofield, & Welchman, 2013; Landy, Maloney, Johnston, & Young, 1995; Marr & Poggio, 1979). If observers use this depth information during the formation of memory representations, recognizing objects or detecting changes between

successive presentations of objects should be easier with stereoscopic presentations than without

stereo information (same image presented to both eyes), which is what was found in multiple

studies for object recognition and face recognition (Bennett & Vuong, 2006; Burke, 2005; Burke,

Taubert, & Higman, 2007; Edelman & Bülthoff, 1992; Humphrey & Khan, 1992; Lee &

Saunders, 2011; Liu & Ward, 2006). Whereas this stereo advantage was specific for new

viewpoints in some studies (Bennett & Vuong, 2006; Burke, 2005; Burke et al., 2007) it also

occurred without viewpoint changes in others (Edelman & Bülthoff, 1992; Lee & Saunders,

2011). Importantly, this stereo advantage cannot be explained by the additional two-dimensional

information present in two views as opposed to one view because only the stereoscopic

presentation of the left and right eyes views but not a side-by-side presentation of both views

results in an increased performance (Burke, 2005). Taken together, the presence of stereoscopic

information supports object recognition suggesting the representation of depth information in

memory. However, the nature of this depth representation is still unspecified, in particular its

specificity to stereoscopic depth and potential transfer effects between depth cues usually found

during visual perception (e.g., Nawrot & Blake, 1989).

Besides binocular disparity, monocular depth cues are available to the visual system such

as shading, texture, or depth from motion (Landy et al., 1995). For example, viewing a rotating

object with one eye only would also provide depth information, as is known from studies

investigating the kinetic depth effect or structure from motion (Braunstein, 1962; Ullman, 1979;

Wallach & O'Connell, 1953). During visual perception, these depth cues are integrated into a

combined depth percept (Dövencioğlu et al., 2013; Ichikawa & Saida, 1996; Landy et al., 1995),

however, there is some evidence to suggest that this is not necessarily the case (Tittle & Perotti,

1997). The integration of depth cues into a combined depth percept is particularly true for depth

from motion and depth from binocular disparity as has been shown by many studies showing that visual adaptation to one of the two depth cues causes transfer effects on the other depth cue (Bradshaw & Rogers, 1996; Nawrot & Blake, 1989, 1991, 1993; Poom & Börjesson, 1999). Furthermore, in perceptual tasks observers can adjust the depth structure perceived through motion to match a successively presented stereoscopic display (Perotti, Todd, Lappin, & Phillips, 1998; Todd & Perotti, 1999). If the representation of depth in object memory resembles the visual perception of depth, we should observe similar transfer effects when presenting depth from motion during object encoding and depth from binocular disparity during object retrieval.

On the other hand, however, mental imagery, and therefore possibly also object memory, has been shown to differ from the visual perception of 3D objects under some circumstances (Lobmaier, Mast, & Hecht, 2010). In their study, Lobmaier et al. (2010) asked participants to view a real 3D object from different viewpoints and to judge the direction where the object pointed to in space. They also had a mental imagery condition in which participants first memorized the object from one perspective and then performed the same direction judgement task with the exception that no physical object was present but participants vividly imagined the object being placed at its original location. In a second experiment, they replaced the real 3D object with a 2D photograph of the object. They found that the pointing errors of participants in the mental imagery condition were akin to the 2D photographs condition and differed from the 3D objects condition, suggesting that the mental imagery of 3D objects differed from the visual perception of 3D objects. Therefore, we might find no transfer between depth cues in memory because the representation of depth in object memory might differ from the visual perception of depth.

With the present set of three experiments, we investigated the nature of depth representations in object memory by studying transfer effects of depth from motion during encoding to depth from binocular disparity during retrieval. After establishing such transfer effects in Experiment 1, we further investigated the contribution of continuous object motion and binocular disparity during encoding on the representation of a depth structure in object memory (Experiments 2a and 2b).

Experiment 1

Experiment 1 had two objectives: First, we investigated whether object rotation alone results in the formation of three-dimensional memory representation that can subsequently be accessed by stereoscopic information during retrieval. Second, we were interested in whether adding stereo information to object rotation during learning provides additional depth information used in the formation of object memory representations, thus causing a larger stereo effect during retrieval.

Method

Participants

Thirty-two students (24 female; age: 19 – 30 years; mean age: 23.47 years) participated in exchange for monetary compensation. Fourteen participants reported normal vision, fifteen participants reported corrected-to-normal vision and three participants reported near-sightedness without correction. Our experiments were approved by the institutional review board and we gained informed consent from the participants in all experiments.

Apparatus and Stimuli

Stimuli were presented on a 55-inch passive polarized stereoscopic display (horizontal interleaved) using the software Blender 2.63 and custom code written in Python 3.2. Participants

were standing and placed at a marking line in order to ensure that they kept a viewing distance of about 190 cm to the screen. We presented molecular-like objects consisting of seven spheres connected by six edges (see Figure 1). We created 176 such objects using an iterative algorithm that started with one sphere and then connected a random number of spheres to each newly created sphere until a total of seven connected spheres was reached. Sphere positions were restricted to prevent overlaps with other spheres. Spheres had a radius of 0.1 units within our virtual coordinate system (distance of virtual camera to object center: 10 units, simulated focal length: 10 units, simulated eye-separation in stereo trials: 0.4 units). Inter-sphere distance was 0.5 units. The edges had a length of 0.3 units and a radius of 0.02 units. Spheres were colored to support orientation in 3D space but the colors were never changed and were irrelevant to the task. Sphere colors were assigned randomly with replacement: red, green, blue, yellow, cyan, purple, white. We used a Nintendo Wii remote control with the MotionPlus extension as input device. Stimuli were rendered either stereoscopically or not according to the experimental conditions and participants wore circularly polarized glasses. We also measured mental rotation abilities (Vandenberg & Kuse, 1978) and stereopsis (Randot SO-002 stereotest, Stereo Optical Co., Inc.).
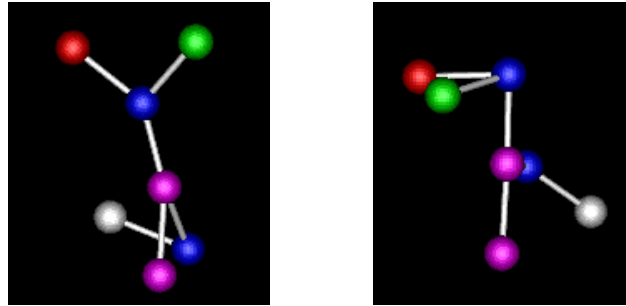
Fig. 1.

Example of the stimuli (left) used in our experiments. Change trials (right) were constructed by rotating all but one edge by at least 30 degrees.

Procedure and Design

We used a change detection paradigm: encoding for 7 s, black screen for 1 s, probe image (retrieval) for 3 s, black screen until response. During encoding, the stimulus automatically rotated around each of its three main axes once. The probe image was a static snapshot showing the stimulus in a random orientation. It depicted a change in one half of the trials. Change trials were constructed by deforming the object (all but one edge were rotated by at least 30 degrees). Participants decided whether the probe image depicted the same object seen during encoding and received feedback after each response. Participants started the next trial by pressing two buttons on the remote control while a fixation cross was presented at zero disparity in the center of the screen.

Experimental conditions were equally distributed across four blocks (44 trials each) and were presented in a random order within each block. The first four trials of each block served as practice trials. We manipulated the stereo mode during encoding and retrieval resulting in a 2 (encoding stereo mode: stereo, non-stereo) x 2 (retrieval stereo mode: stereo, non-stereo) x 2 (change present: yes, no) x 20 (repetitions) within-subjects design.

At the beginning of the experiment, we applied the mental rotation test. Following each block, participants answered the NASA-TLX rating sheet (Hart & Staveland, 1988) to provide participants with a break between blocks. Following all blocks, we tested participants' stereoscopic acuity using the Randot stereotest.

Results and Discussion

We analyzed change detection performance based on the signal detection theory using the sensitivity measure d' as dependent variable. All responses with response times larger than 8 s (i.e., response occurred more than 5 s after probe offset) were considered invalid and removed from the data set (5 trials, 0.10% of the data).

We analyzed the effects of stereo mode during encoding and retrieval on change detection performance (see Figure 2) using a repeated measures ANOVA with the independent variables encoding stereo mode and retrieval stereo mode and the dependent variable d'. This revealed a significant main effect of retrieval stereo mode, $F(1, 31) = 12.17$, $p = .001$, but neither the main effect of encoding stereo mode, $F(1, 31) = 0.88$, $p = .354$, nor the interaction of encoding stereo mode and retrieval stereo mode, $F(1, 31) = 0.20$, $p = .655$, were significant. The higher performance with stereo information during retrieval demonstrates the representation of the objects' depth structure in memory. Importantly, the presence of object rotation alone (non-stereo encoding phase) was sufficient to produce a stereo effect during retrieval; that is, the depth information extracted from object rotations during encoding could be accessed by stereo information during retrieval, indicating the representation of an object's depth structure instead of specific depth cues in memory. The non-significant main effect of encoding stereo mode and non-significant interaction of encoding stereo mode and retrieval stereo mode are, at a first glance, surprising; that is, the presence of stereoscopic information during encoding did not

provide any additional depth information (in addition to object rotation) relevant for the

formation of memory representations. Possibly, the fast rotation of our objects during encoding

provided enough depth information rendering stereoscopic information ineffective. Indeed,

similar effects have previously been observed for the integration of motion information and

binocular disparity during depth perception (Ichikawa & Saida, 1996).
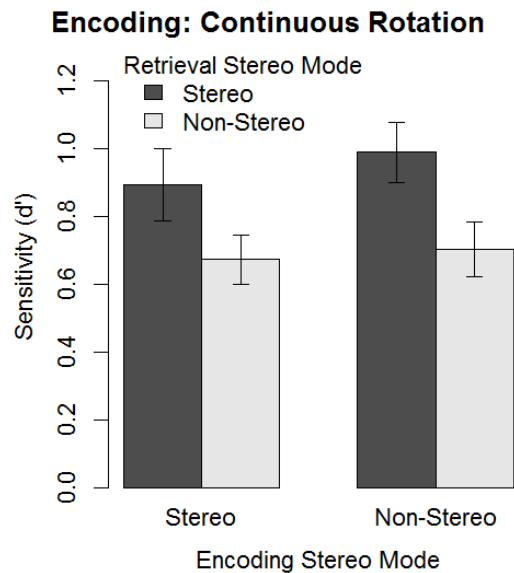

Fig. 2.

Results of Experiment 1 (objects were continuously rotating during encoding). Change detection

performance was higher with stereoscopic cues during retrieval even when no stereo information

was present in addition to object rotation during encoding, indicating the representation of the

depth structure of objects in memory (transfer between depth cues). Error bars indicate the SEM.

Experiment 2a

Continuous object rotation seems to act as a depth cue supporting the representation of an

object's depth structure in memory that transfers to stereoscopic information during retrieval. If

this is true, then removing depth from motion during encoding should eliminate the stereo effect during retrieval. Studies on structure from motion showed that the effectiveness of motion as depth cue is strongly determined by the size of angular rotation between successive views of an rotating object (Dick, Ullman, & Sagi, 1991; Todd, Akerstrom, Reichel, & Hayes, 1988). The larger the rotation between two successive views of a rotating object, the less apparent motion is perceived and the less depth is provided by the motion cue. Therefore, presenting multiple views separated by 120 degrees of rotation during encoding should effectively eliminate motion as depth cues in our experiments. In Experiment 2a, we either presented continuous object rotations during encoding or we presented multiple views not connected by continuous rotations but separated by 120 degrees of rotation. We predicted that the stereo benefit during retrieval would only be present in the former but not latter conditions.

Method

Participants

A new set of thirty-two students (20 female; age: 20 – 32 years; mean age: 24.00 years) participated in exchange for monetary compensation. Eighteen participants reported normal vision, nine participants reported corrected-to-normal vision and five participants reported near-sightedness without correction.

Apparatus and Stimuli

Apparatus and stimuli were the same as in Experiment 1.

Procedure and Design

The procedure was similar to Experiment 1 (non-stereo encoding conditions). The memory object either rotated around each of its three main axes continuously as in Experiment 1, or nine static snapshots along the same rotation path in jumps of 120 degrees were shown for 778

ms each (presentation mode blocked and counter-balanced across participants). In addition, we varied the stereo mode during retrieval resulting in a 2 (rotation continuity: continuous, multiple snapshots) x 2 (retrieval stereo mode: stereo, non-stereo) x 2 (change present: yes, no) x 20 (repetitions) within-subjects design.

Results and Discussion

We excluded one participant from the analysis who did not see any depth information in the Randot stereotest. All responses with response times larger than 8 s were considered invalid and removed from the data set (3 trials, 0.06% of the data). Because d' is not defined for hit rates and false alarm rates of 1.0 and 0.0, we replaced such values with half a trial incorrect/correct, that is $(1-1/(2N))$ and $1/(2N)$ respectively.

We analyzed the effects of rotation continuity and retrieval stereo mode on change detection performance (see Figure 3) using a repeated measures ANOVA with the independent variables rotation continuity and retrieval stereo mode and the dependent variable d'. The interaction of rotation continuity and retrieval stereo mode was significant, $F(1, 30) = 13.81, p < .001$, and the main effects for rotation continuity, $F(1, 30) = 2.93, p = .097$, and retrieval stereo mode, $F(1, 30) = 0.45, p = .510$, were not significant. Two post-hoc paired t-tests that were uncorrected for multiple comparisons and restricted to each rotation continuity condition revealed a significant retrieval stereo mode effect with continuous object rotation, $t(30) = 2.80, p = .009$, but not with multiple snapshots, $t(30) = -1.80, p = .090$; that is, participants could only benefit from the presence of stereoscopic depth cues during retrieval if continuous object motion was present during encoding. Therefore, continuous object rotation acted as a depth cue during encoding, supporting the representation of an object's depth structure in memory. In the multiple snapshots condition, however, no motion cues were present preventing participants from

encoding the object's depth structure to memory. Also note that the absence of the stereo advantage in the multiple snapshots condition shows that the stereo benefit during retrieval found in our experiments is not an artifact occurring under any encoding situation.

Experiment 2b

One potential problem in the design of Experiment 2a is the fact that the multiple snapshots condition does not only remove the apparent motion cues but also presents a reduced number of views as compared to the continuous rotation condition during learning. Therefore, it remains possible that the altered response pattern observed in the multiple snapshots condition is not related to any encoded depth structure in memory but to the fact that participants encoded fewer views in memory. The reduced number of views in memory could in turn reduce the ability to generalize to the novel viewpoint of the static probe (Bülthoff & Edelman, 1992; Tarr & Pinker, 1989) and also reduce the ability of participants to integrate the learned views into a common object representation (Wallis & Bülthoff, 2001). Therefore, we conducted Experiment 2b which was identical to Experiment 2a with the exception that we enabled stereoscopic rendering during the encoding phase. If it is indeed the reduced number of views in memory that changes the response pattern in the multiple snapshots condition, we should observe the same interaction that was present in Experiment 2a. If, however, participants failed to encode the object's depth structure in the multiple snapshots condition of Experiment 2a, enabling stereoscopic rendering during encoding in the present experiment should act as depth cue allowing participants to encode the object's depth structure despite the reduced number of views. Thus, participants should encode the object's depth structure in both the continuous rotation condition and multiple snapshots condition of the present experiment resulting in a main effect of retrieval stereo mode irrespective of rotation continuity (no interaction).

Method

Participants

A new set of thirty-two students (22 female; age: 19 – 33 years; mean age: 25.69 years) participated in exchange for monetary compensation or course credit. Seventeen participants reported normal vision, fourteen participants reported corrected-to-normal vision and one participants reported near-sightedness without correction.

Apparatus, Stimuli, Procedure and Design

The experiment was identical to Experiment 2a with the only exception that the encoding phase was presented in stereo.

Results and Discussion

All responses with response times larger than 8 s were considered invalid and removed from the data set (8 trials, 0.16% of the data). Because d' is not defined for hit rates and false alarm rates of 1.0 and 0.0, we replaced such values with half a trial incorrect/correct, that is (1-1/(2N)) and 1/(2N) respectively.

We analyzed change detection performance (see Figure 3) using a repeated measures ANOVA with the independent variables rotation continuity and retrieval stereo mode and the dependent variable d'. As predicted, the interaction of rotation continuity and retrieval stereo mode was not significant, $F(1, 31) = 2.69*10^{-5}$, $p = .996$. Instead, we observed a main effect of retrieval stereo mode, $F(1, 31) = 5.13$, $p = .031$, indicating that the presence of stereoscopic information during the retrieval phase supported change detection irrespective of the continuity of object rotation during encoding in the present experiment. The main effect of rotation continuity was not significant, $F(1, 31) = 1.88$, $p = .180$.

The results of this experiment indicate that the interaction effect observed in Experiment 2a was not an artefact of the reduced number of views in the multiple snapshots condition. If this was true, the same interaction effect should have occurred in the present experiment. Instead, the presence of stereoscopic information during encoding enabled participants to extract the objects' depth structure into memory irrespective of rotation continuity in the present experiment.
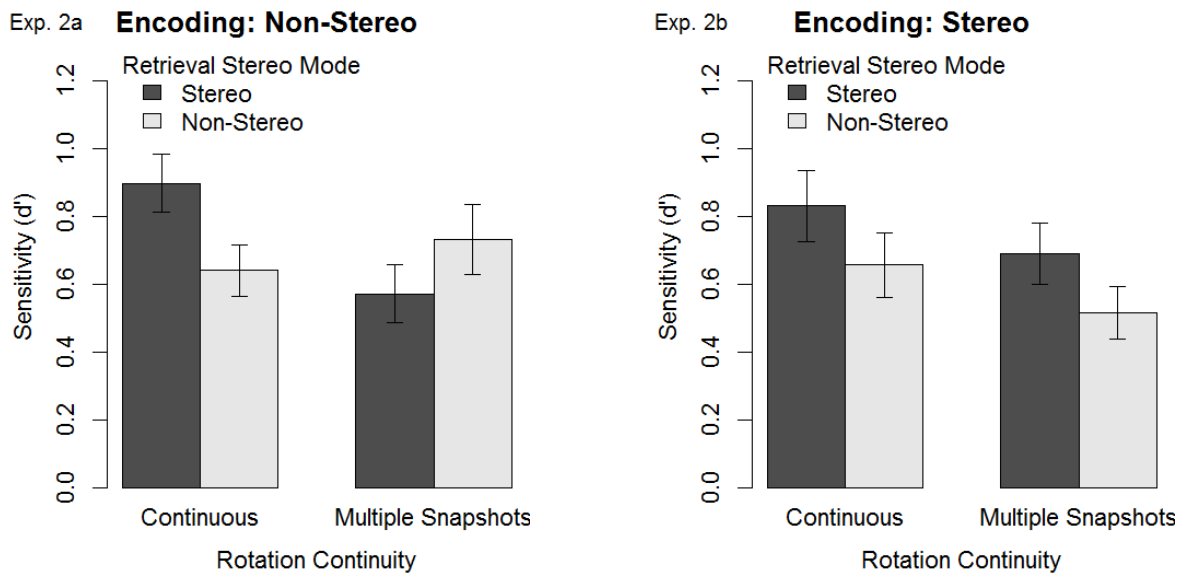

Fig. 3.

Results of Experiment 2a (left, encoding phase: without stereoscopic information) and Experiment 2b (right, encoding phase: with stereoscopic information). Stereo information during retrieval supported change detection only if object rotation was continuous during encoding (Experiment 2a) and/or stereoscopic cues were available during encoding (Experiment 2b). If however neither stereoscopic cues nor a continuous rotation were available during encoding (Experiment 2a, multiple snapshots), no stereo benefit during retrieval occurred. Thus, both the continuity of object rotations and stereoscopic information caused the extraction of an object's depth structure to memory. This depth structure was accessible by stereoscopic information

during retrieval independent of the depth cue originally presented during encoding. Error bars

indicate the SEM.

Cross-Experimental Analysis: Experiment 2a vs. Experiment 2b

We conducted a final cross-experimental analysis in order to ensure that the result

patterns of Experiment 2a and Experiment 2b differed significantly from one another. While we

observed a significant interaction of rotation continuity and retrieval stereo mode in Experiment

2a (non-stereoscopic encoding phase), we observed no such interaction in Experiment 2b

(stereoscopic encoding phase). We conducted a mixed ANOVA with the within factors rotation

continuity (continuous vs. multiple snapshots) and retrieval stereo mode (stereo vs. non-stereo)

and the between factor experiment (Experiment 2a vs. Experiment 2b) and the dependent

variable d'. Most important, the three-way interaction of rotation continuity, retrieval stereo

mode and experiment was significant, $F(1, 61) = 6.92$, $p = .011$. That is, the result patterns of

Experiment 2a and Experiment 2b did indeed differ significantly from one another supporting

our claim that the presence of continuous rotation and/or stereoscopic depth during encoding

supported the extraction of an object's depth structure to memory while this was not the case

when neither depth cue was present (multiple snapshots condition in Experiment 2a).

Furthermore, the main effect of experiment was not significant, $F(1, 61) = 0.17$, $p = .680$,

indicating that the overall performance did not differ significantly between the two experiments.

The remaining effects were as follows: significant main effect of retrieval stereo mode, $F(1, 61)$

$= 4.49$, $p = .038$, significant main effect of rotation continuity, $F(1, 61) = 4.32$, $p = .042$,

significant interaction of rotation continuity and retrieval stereo mode, $F(1, 61) = 6.74$, $p = .012$,

non-significant interaction of experiment and rotation continuity, $F(1, 61) = 0.04$, $p = .848$, non-significant interaction of experiment and retrieval stereo mode, $F(1, 61) = 1.41$, $p = .240$.

General Discussion

Recognizing objects and detecting changes in objects is easier when viewed stereoscopically than when viewed without stereo information (e.g., Bennett & Vuong, 2006; Burke, 2005; Burke et al., 2007; Edelman & Bülthoff, 1992). This suggests the representation of depth information stemming from binocular disparity in memory. However, the nature of this representation was unspecified. In particular, it remained unclear whether these findings were specific to stereoscopic depth cues or whether a more general depth structure of objects was represented in memory. We examined this question by asking whether depth information stemming from object rotation during encoding could facilitate retrieval with stereoscopic information. With our experiments, we present evidence for this transfer between continuous object rotation as a depth cue during encoding and stereoscopic information as a depth cue during retrieval. In Experiment 1, continuous object rotations supported the encoding of an object's depth structure into memory independent from the presence of additional stereoscopic information. Experiments 2a and 2b demonstrated that the continuity of object rotation and/or stereoscopic depth acted as a depth cue facilitating the extraction of an object's depth structure to memory. Our results do, therefore, provide evidence for the representation of an object's depth structure in memory and that this depth structure is not specific to certain depth cues.

Having shown that an object's depth structure is represented in memory, this introduces the question of the format of its representation. Considering the numerous findings showing viewpoint-dependent object recognition (e.g., Edelman & Bülthoff, 1992; Rock & DiVita, 1987;

Tarr et al., 1998), it seems unlikely that depth information supports the formation of a viewpoint-

independent 3D models of memorized objects. This is particularly unlikely because object

recognition remains viewpoint-dependent even when performed with stereoscopic depth

information (e.g., Bülthoff & Edelman, 1992; Burke, 2005; Lee & Saunders, 2011). Nonetheless,

viewpoint effects are reduced by stereo information (Bennett & Vuong, 2006; Burke, 2005;

Burke et al., 2007), suggesting that depth information helps to generalize across viewpoints to

some extent. This was also the case in our experiments because participants saw the probe image

from novel viewpoints. Depth information might, therefore, be represented as a local depth map

extracted during perception (Landy et al., 1995) or as 2.5D sketch as proposed by Marr and

Nishihara (1978). But if so, it remains puzzling why stereo information was only beneficial with

viewpoint changes but not without viewpoint changes in some studies (Bennett & Vuong, 2006;

Burke, 2005; Burke et al., 2007). When taking research on the representation of object color and

shape into account (e.g., Wheeler & Treisman, 2002), the following representational structure

seems feasible. Depth information obtained from object rotation or stereoscopic information

might not be integrated with other visual information obtained from an object. Instead, depth

information might be stored on a layer separate from other two-dimensional object information

in memory, such as object color being stored separately from object shape, requiring binding

mechanisms to combine different sources of information to a single object representation

(Wheeler & Treisman, 2002). This could explain the absence of stereo effects without viewpoint

changes in some studies (Bennett & Vuong, 2006; Burke, 2005; Burke et al., 2007) by stating

that two-dimensional object information gives a perfect match without changes in viewpoint thus

not requiring any additional information from the depth layer. Assuming that depth information

is less viable to viewpoint-changes than two-dimensional information, it becomes important once

the viewpoint changes thus reducing viewpoint costs. Certainly, based on current empirical findings, this proposal remains speculative. But we are confident that further research in this direction can provide valuable insights on both the structure of object memory in general and the representation of an object's depth structure in particular.

The present set of experiments was designed to investigate whether the benefit of stereo information on change detection (Bennett & Vuong, 2006; Burke, 2005; Burke et al., 2007) is caused by a specific representation of depth from binocular disparity in memory or whether the depth structure of objects is represented in a cue-independent manner. We chose to study this question by investigating transfer effects of depth from motion to depth from binocular disparity. Our results indicate that such transfer effects exist thus supporting the conclusion that the representation of objects' depth structure in memory is cue-independent. Future studies should address the opposite direction of transfer, which is transfer of depth from binocular disparity to depth from motion. Based on our conclusions and the fact that such transfer effects have previously been observed during depth perception (Bradshaw & Rogers, 1996; Nawrot & Blake, 1989; Poom & Börjesson, 1999), we would predict that such transfer effects also exist for object memory.

Our finding that an object's depth structure is represented independent from the depth cues present during encoding poses a problem for the interpretation of possible null effects in object recognition and change detection tasks involving manipulations of depth cues. The manipulation of a particular depth cue might sometimes be ineffective because another depth cue present in an image might be sufficient to extract an objects depth structure, as was the case with the addition of stereoscopic information to continuous object rotations in our first experiment.

Such null effects are, therefore, not a reliable indicator of the absence of depth representations in memory as long as alternative depth cues are available during encoding.

Summarizing, we showed that the depth structure of objects is represented in memory and that this representation is not tied to the specific depth cues present during encoding. Instead, depth information extracted from object rotations during encoding could subsequently be accessed by stereoscopic information during retrieval. We suggested that depth information might be stored on a layer separate from other two-dimensional object information in memory in similar ways as object color is stored separately from object shape requiring binding processes to form a single representation (Wheeler & Treisman, 2002). We are confident that further research testing this idea can provide valuable insights on the structure of object memory.

Author Contributions

Both authors developed the study concept. F. Papenmeier was responsible for testing and data collection, and performed the data analysis and interpretation. F. Papenmeier drafted the manuscript, and S. Schwan provided critical revisions. Both authors approved the final version of the manuscript for submission.

References

Bennett, D. J., & Vuong, Q. C. (2006). A stereo advantage in generalizing over changes in viewpoint on object recognition tasks. *Perception & Psychophysics*, *68*, 1082–1093.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, *39*, 2885–2899.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1162–1182.

Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research*, *36*, 3457–3468. http://doi.org/10.1016/0042-6989(96)00072-7

Braunstein, M. L. (1962). The perception of depth through motion. *Psychological Bulletin*, *59*, 422–433.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, *89*(1), 60–64.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*, 247–260. http://doi.org/10.1093/cercor/5.3.247

Burke, D. (2005). Combining disparate views of objects: Viewpoint costs are reduced by stereopsis. *Visual Cognition*, *12*, 705–719. http://doi.org/10.1080/13506280444000463

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces. *Vision Research*, *47*, 2164–2169. http://doi.org/10.1016/j.visres.2007.04.018

Dick, M., Ullman, S., & Sagi, D. (1991). Short- and long-range processes in structure-from-motion. *Vision Research*, *31*, 2025–2028. http://doi.org/10.1016/0042-6989(91)90197-D

Dövencioğlu, D., Ban, H., Schofield, A. J., & Welchman, A. E. (2013). Perceptual integration for qualitatively different 3-D cues in the human brain. *Journal of Cognitive Neuroscience*, *25*, 1527–1541. http://doi.org/10.1162/jocn_a_00417

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, *32*, 2385–2400.

Farah, M. J., Rochlin, R., & Klein, K. L. (1994). Orientation invariance and geometric primitives in shape recognition. *Cognitive Science*, *18*, 325–344. http://doi.org/10.1016/0364-0213(94)90004-3

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1511–1521.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517. http://doi.org/10.1037/0033-295X.99.3.480

Humphrey, G. K., & Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology*, *46*, 170–190.

Ichikawa, M., & Saida, S. (1996). How is motion disparity integrated with binocular disparity in depth perception? *Perception & Psychophysics*, *58*, 271–282.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*, 389–412. http://doi.org/10.1016/0042-6989(94)00176-M

Lee, Y. L., & Saunders, J. A. (2011). Stereo improves 3D shape discrimination even when rich monocular shape cues are available. *Journal of Vision*, *11*(9), 1–12. http://doi.org/10.1167/11.9.6

Liu, C. H., & Ward, J. (2006). The use of 3D information in face recognition. *Vision Research*, *46*, 768–773. http://doi.org/10.1016/j.visres.2005.10.008

Lobmaier, J. S., Mast, F. W., & Hecht, H. (2010). For the mind's eye the world is two-dimensional. *Psychonomic Bulletin & Review*, *17*, 36–40. http://doi.org/10.3758/PBR.17.1.36

Marr, D., & Nishihara, H. K. (1978). Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society of London B*, *200*, 269–294. http://doi.org/10.1098/rspb.1978.0020

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, *204*, 301–328.

Nawrot, M., & Blake, R. (1989). Neural integration of information specifying structure from stereopsis and motion. *Science*, *244*, 716–718. http://doi.org/10.1126/science.2717948

Nawrot, M., & Blake, R. (1991). The interplay between stereopsis and structure from motion.

    Perception & Psychophysics, *49*, 230–244. http://doi.org/10.3758/BF03214308

Nawrot, M., & Blake, R. (1993). On the perceptual identity of dynamic stereopsis and kinetic

    depth. *Vision Research*, *33*, 1561–1571. http://doi.org/10.1016/0042-6989(93)90149-Q

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature

    from optical motion. *Perception & Psychophysics*, *60*, 377–388.

Poom, L., & Börjesson, E. (1999). Perceptual depth synthesis in the visual system as revealed by

    selective adaptation. *Journal of Experimental Psychology: Human Perception and*

    Performance, *25*, 504–517. http://doi.org/10.1037/0096-1523.25.2.504

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive*

    Psychology, *19*, 280–293. http://doi.org/10.1016/0010-0285(87)90013-2

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon

    structural descriptions or by multiple views? Comment on Biederman and Gerhardstein

    (1993). *Journal of Experimental Psychology: Human Perception and Performance*, *21*,

    1494–1505.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape

    recognition. *Cognitive Psychology*, *21*, 233–282. http://doi.org/10.1016/0010-

    0285(89)90009-1

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered

    reference frame? *Psychological Science*, *1*, 253–256. http://doi.org/10.1111/j.1467-

    9280.1990.tb00209.x

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object

    recognition is viewpoint dependent. *Nature Neuroscience*, *1*, 275–277.

    http://doi.org/10.1038/1089

Tittle, J. S., & Perotti, V. J. (1997). The perception of shape and curvedness from binocular

    stereopsis and structure from motion. *Perception & Psychophysics*, *59*, 1167–1179.

    http://doi.org/10.3758/BF03214205

Todd, J. T., Akerstrom, R. A., Reichel, F. D., & Hayes, W. (1988). Apparent rotation in three-

    dimensional space: Effects of temporal, spatial, and structural factors. *Perception &*

    Psychophysics, *43*, 179–188. http://doi.org/10.3758/BF03214196

Todd, J. T., & Perotti, V. J. (1999). The visual perception of surface orientation from optical

    motion. *Perception & Psychophysics*, *61*, 1577–1589.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE*

    Transactions on Pattern Analysis and Machine Intelligence, *13*, 992–1006.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional

    spatial visualization. *Perceptual and Motor Skills*, *47*, 599–604.

    http://doi.org/10.2466/pms.1978.47.2.599

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental*

    Psychology, *45*, 205–217. http://doi.org/10.1037/h0056880

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory.

    Proceedings of the National Academy of Sciences, *98*(8), 4800–4804.

    http://doi.org/10.1073/pnas.071028598

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64. http://doi.org/10.1037//0096-3445.131.1.48

Figure Captions

Fig. 1.

Example of the stimuli (left) used in our experiments. Change trials (right) were constructed by rotating all but one edge by at least 30 degrees.

Fig. 2.

Results of Experiment 1 (objects were continuously rotating during encoding). Change detection performance was higher with stereoscopic cues during retrieval even when no stereo information was present in addition to object rotation during encoding, indicating the representation of the depth structure of objects in memory (transfer between depth cues). Error bars indicate the SEM.

Fig. 3.

Results of Experiment 2a (left, encoding phase: without stereoscopic information) and Experiment 2b (right, encoding phase: with stereoscopic information). Stereo information during retrieval supported change detection only if object rotation was continuous during encoding (Experiment 2a) and/or stereoscopic cues were available during encoding (Experiment 2b). If however neither stereoscopic cues nor a continuous rotation were available during encoding (Experiment 2a, multiple snapshots), no stereo benefit during retrieval occurred. Thus, both the continuity of object rotations and stereoscopic information caused the extraction of an object's depth structure to memory. This depth structure was accessible by stereoscopic information during retrieval independent of the depth cue originally presented during encoding. Error bars indicate the SEM.