Shape and motion interactions at perceptual and attentional levels during processing of structure from motion stimuli

A. Miskiewicz  
CNRS, UPR640 LENA, Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, Paris, France, & Université Pierre et Marie Curie-Paris6, Paris, France  
Institut de Médecine Aérospatiale du Service de Santé des Armées, Brétigny sur Orge, France  
S. Buffat  
CNRS, UPR640 LENA, Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, Paris, France, & Université Pierre et Marie Curie-Paris6, Paris, France  
A.-L. Paradis  
CNRS, UPR640 LENA, Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, Paris, France, & Université Pierre et Marie Curie-Paris6, Paris, France  
J. Lorenceau  
CNRS, UPR640 LENA, Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, Paris, France, & Université Pierre et Marie Curie-Paris6, Paris, France

This study uses a rapid-serial-visual-presentation (RSVP) paradigm to test the extent to which shape and motion direction can be independently accessed and processed during the perception of structure-from-motion (SFM) stimuli. Subjects reported the number of occurrences of shape or motion direction during RSVP sequences of 3D-SFM stimuli. Overall, performance was better for motion than shape. In the motion task, observers were less accurate when the motion direction was repeated revealing a repetition blindness (RB) effect. In addition, the repetition of shape, although irrelevant to the motion task, resulted in increased performance, without change in RB rate. In contrast, there was no RB at the group level in the shape task and the repetition of the irrelevant motion direction had no effect on the performance. A closer look at the data showed that observers fall in two statistically distinct groups for the shape task. Some observers (N = 6) show a repetition advantage (RA) while the others (N = 5) show a repetition blindness (RB) effect. No behavioral differences between groups could be found for the motion task. The implications of these results for models of SFM processing are discussed in the light of the type/token theory (N. Kanwisher, 2001).

Keywords: 3D structure from motion, shape, motion, attention, RSVP, type, token


Introduction

Numerous studies (Kanwisher, Driver, & Machado, 1995; O’Craven, Downing, & Kanwisher, 1999; Treisman & Kanwisher, 1998) have investigated whether attention can be used to selectively process a specific dimension or feature of objects, such as their shape or color. Several studies (Fox & de Fockert, 2001; Harris & Dux, 2005; Kanwisher et al., 1995) have used an RSVP paradigm to study the dynamics and selectivity of attentional selection of a cued dimension through measures of repetition blindness (RB), which refers to a reduced capacity to report both occurrences of a repeated item during an RSVP sequence. One account of RB relies on the concept of “object file” (Kahneman, Treisman, & Gibbs, 1992; Kanwisher, 1991). While the recognition of an object would activate a “type,” successive occurrences of the same item would, or not in the case of RB, be recorded as different “tokens” or “object files.” RB can therefore be used to assess when two stimuli are considered as “the same object” by the visual system, providing insights into what defines an object file. Studies with letters, words, simple shapes, and colors, for which the cooccurrence of two dimensions were arbitrarily chosen (e.g., a red letter), have found that attention can select a single feature or dimension with little or no interference from other dimensions irrelevant to the task at hand (Kanwisher & Driver, 1992). On the other hand, studies with moving photographs of faces, houses, or drawings of objects have concluded that objects—and their intricate dimensions—are the units of attentional selection (Kanwisher...
and Wojciulik, 2000; O’Craven et al., 1999; see also Boucart, Humphreys, & Lorenceau, 1995).

One possible reason for these discrepancies is that whether attentive selection operates at the feature or at the object level may depend on the nature of the cued dimension (shape, color, motion) and on its status relative to the target object. Some features may be considered “intrinsic” to the object itself, as modifying them may disrupt or alter the activation of a type, resulting in poor categorization or recognition (e.g., Biederman, 1987). Other features can be modified without consequences on object categorization or recognition and can thus be considered “extrinsic” to the target object and not linked to the object type (Boucart & Humphreys, 1997; Delorme, Richard, & Fabre-Thorpe, 2000; Ecker, Zimmer, & Groh-Bordin, 2007).

In this regard, the status of motion relative to an object is ubiquitous. At first sight, motion may be considered an extrinsic object property, as an object can be static or moving, yet this may not influence its recognition. For instance, in O’Craven et al.’s (1999) study, a back and forth translation is arbitrarily assigned to a picture and any other motion could have been used without changing much their conclusions.

On the other hand, biological motion studies suggest that motion can be an intrinsic object property providing in itself an access to object identity (Jansson & Johansson, 1973; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). Similarly, it is well known that motion parallax provides a powerful cue to object recognition, as in structure-from-motion (SFM) stimuli (Wallach & O’Connell, 1953). With SFM stimuli, an arbitrary rotation imposed to a three-dimensional cloud of dots attached to the surface or volume of an object can elicit a vivid perception of its three-dimensional structure. In this case, motion is both necessary and sufficient to determine the identity of a shape. Previous studies (Hildreth, Grzywacz, Adelson, & Inada, 1990; Ullman, 1984) have further suggested that recovering shape structure in SFM develops over time by incrementally refining an initially crude 3D model through successive iterations.

The tight dependence of shape perception on motion analysis in SFM stimuli thus provides an interesting and relevant tool to evaluate whether attention can, or cannot, select a particular dimension or feature of an object. To shed light on this issue, we therefore used 3D structure-from-motion (3D-SFM in the following) stimuli presented in RSVP sequences and measured identification performance and RB rate in a shape and a motion task. In this way, we aimed at testing whether attention can be deployed so as to selectively and independently process these features and whether repetition of the attribute irrelevant to the current task influences performance and RB rate. In the following, we assume that performance, the percentage of correct identification, more specifically reflects the processing of a type, whereas RB rate is an estimate of how accurately tokens can be registered. If RB previously found for shape also exists for motion direction, and given that shape can only be accessed through motion parallax, we reasoned that reporting the occurrence of shapes (repeated or not) should be impaired by RB to motion direction. In addition, if the registration of tokens is facilitated by a change of the irrelevant motion dimension, RB to shape should be less when it is

![Figure 1. Working hypotheses. Each circle represents an occurrence of a 3D-SFM stimulus, with its shape (S) and motion (M) attributes. Small arrows from M to S remind that shape perception depends on motion. C1 stands for “first Critical target,” D for “Distractor,” and C2 for “second Critical target” (see General methods section). The colored part pointed by the “Task” arrow (red for shape, green for motion) indicates which attribute is attended to during the task. Empty arrows point toward two processes that are expected to take place for the attended attribute. These are the type identification (i.e., identification of the attended attribute) and the token attribution (i.e., the counting of the occurrences for each shape or motion direction). RB rate, corresponding to a failure of token attribution, is the difference of performance between repeated and non-repeated attended attribute. The mean performance for repeated and non-repeated attended attribute (% correct) is an index of type identification. (A) Motion task: motion repetition is expected to impair motion “tokenization,” resulting in a significant RB rate. (B) Shape task: shape repetition is expected to impair shape “tokenization,” resulting in a significant RB rate. Because shape perception depends on motion cues in 3D-SFM stimuli, motion repetition—although irrelevant to the task—is expected to alter shape processing (curved green arrows) either at the identification level—assessed by the % correct—or at the tokenization level—assessed by the RB rate.](http://jov.arvojournals.org/ on 04/20/2018)
associated to different motion directions in RSVP sequences. In contrast, processing direction in the motion task should not be modulated by shape repetition if processing 3D-SFM stimuli involves a hierarchical motion-to-shape flow (see Figure 1).

In the following, we first describe an experiment designed to determine the minimal durations required for identifying motion directions and shapes before presenting the main experiment using RSVP sequences and measures of RB rate in both a motion and a shape task.

**General methods**

The 3D-SFM stimuli (Figure 2) were made of 140 white discs (0.19° of visual angle, 40.3 cd/m²) randomly distributed on a black background (0.1 cd/m²) and projected using parallel projection on one of three 3D virtual surfaces: a hyperbolic paraboloid, a paraboloid, and a parabolic cylinder (Figure 2A). The virtual surface, subtending 28° of visual angle, could rotate around one of three fronto-parallel axes tangent to the surface, tilted at either −60°, 60°, or 180° from vertical and passing through the screen center. To avoid any association between shapes and dot configuration, an identical dot arrangement was used for all 3D surfaces when the sequence comprised one item only. When the stimulus sequence comprised more than one item (i.e., in the RSVP experiment), each item of the sequence had a different dot arrangement from the preceding one. A cardboard mask fixed to the screen delimited a circular viewing window (20° of visual angle) so as to hide the boundaries of the 3D-SFM stimuli. Under these conditions, no contour, texture, configuration, or density information could be used to distinguish one shape from the others, preventing shape identification from static cues. Also note that because motion duration was very short, the changes in local dot density resulting from rotational motion were minimal and could not be used to discriminate the different shapes. Dot speed varied with eccentricity from 0°/s at the screen center to a maximum of 7°/s at the edge of the circular mask. Stimuli were displayed on an Iiyama HM204DT (22 inch, 1024 × 768 × 8 bits per pixel, refreshed at 150 or 60 Hz). Head-screen distance was maintained at 57 cm by means of a chin rest (Figure 2B).

During the experiments, the sequence of 3D-SFM stimuli was flanked by two masks made of incoherently moving dots in order to control the processing time and to impose a temporal sequencing during a trial (Figure 2C). A short mask, made of 140 dots (0.19° of visual angle;...
40.2 cd/m²), was displayed for 400 ms before and after a 3D-stimulus sequence. A long mask, made of 500 dots (0.19° of visual angle; 31.64 cd/m²), was presented for 833 ms before and after the short mask. To compensate for the larger number of dots and to avoid strong luminance transitions, dot luminance was lowered to 8.58 cd/m² in this latter mask. The duration and the number of 3D-SFM stimuli within a sequence varied in the different experiments (see below). Subjects were asked either to identify the shapes (“shape task”) or to identify the directions of motion (“motion task”) in blocked-task runs. Participants were instructed to attend to the whole screen to ease identification.

**Experiment 1: Duration thresholds**

The aim of this experiment was twofold:

i. to measure the minimum duration required to process a 3D-SFM stimulus or its direction and

ii. to determine if these thresholds are compatible with the use of 3D-SFM stimuli in RSVP sequences.

A number of studies suggest that recovering the motion direction and shape of SFM stimuli involves a motion measurement stage—presumably V1—followed by moving surface reconstruction stage, involving area MT (Bradley, Chang, & Andersen, 1998; Treue, Andersen, Ando, & Hildreth, 1995). This may entail that less time is needed to process motion as compared to the processing of shape, but we are not aware of experiments investigating this possibility. If true, we may expect shorter duration thresholds for identifying the direction of motion of 3D-SFM stimuli than for identifying their shape. We tested this prediction using a unique 3D-SFM stimulus flanked by a sequence of two dynamic masks (see General methods section). Minimal durations for identifying the direction of motion of 3D-SFM stimuli were measured using a three-alternative forced choice (3-AFC) method coupled with a staircase procedure. One of three intermingled staircases (one for each of the 3 shapes or for each of the 3 motion directions) was chosen at random on each trial. At the end of a trial, a static image of 3 shapes or 3 arrows corresponding to the 3 different possible responses appeared. Observers signaled which shape (or direction) had been presented by pressing the corresponding key. In the shape task, a single motion direction was used in a block of trials and each observer ran 3 blocks, one for each direction; similarly, a single shape was used in each of the 3 runs in the motion task. The staircase procedure varied stimulus duration so as to converge onto 79% correct identification (3 consecutive correct responses: 1 step down; 1 error: 1 step up). At the refresh rate of 150 Hz used in this experiment, the duration step was 6.66 ms. The initial duration was 200 ms. A session ended after 10 reversals; thresholds were computed as the mean of the last 6 reversals. The whole procedure was repeated 3 times, resulting in 9 measures of duration thresholds for each dimension. Four right-handed subjects with normal or corrected-to-normal vision participated in this experiment.

**Results**

The results for the 3 directions of motion and the 3 shapes are presented in Figure 3. The mean duration threshold for the motion task (56 ms ± 9 ms) is significantly shorter (by 40 ms ± 17 ms; $F(1,3) = 16.91$, $p = 0.026$) than mean duration threshold for the shape task (97 ms ± 17 ms; Figure 3). The differences between the three directions ($F(2,10) = 1.93$, $p = 0.19$) or the three shapes are not significant ($F(2,10) = 1.12$, $p = 0.36$), although there exists a trend toward lower duration thresholds for the hyperbolic paraboloid.

The lower duration threshold in the motion task supports the hypothesis of a sequential processing of motion and shape in SFM. It remains however possible that shape and motion are processed along parallel—ventral and dorsal—pathways (Bullier & Nowak, 1995). In this case, the difference between the thresholds could reflect the contributions and relative speed of the parvocellular and magnocellular pathways (Yabuta &
Callaway, 1998). Another account of the different duration thresholds could be a difference in task difficulty between the two conditions. Although one cannot exclude the possibility that the motion task was easier because observers could rely on the local motion of a single dot, this is unlikely. The dots in the region of fixation (at the center of the screen) were either static or moving slowly and opposite directions of motion were present in different locations. Furthermore, because the time available to perform the task was so short it is unlikely that observers could direct their attention toward a single dot.

Although the present results are not sufficient to determine the origin of this difference, they nevertheless show that motion information is available sooner than shape information. Importantly, duration thresholds in both tasks are compatible with the temporal constraints of a rapid serial visual presentation paradigm. However, the differential temporal sensitivity found for motion and shape may modulate the ability to report their occurrences in an RSVP sequence, a point addressed further in the next experiment.

Experiment 2: RSVP with 3D-SFM stimuli

The aim of this experiment was to measure performance and RB rates when attention is directed to the shape or to the direction of motion of 3D-SFM stimuli and to further investigate if performance and RB rate are influenced by the repetition of the irrelevant unattended dimension. Because recovering the shape from 3D-SFM stimuli is difficult with short durations, in particular when using RSVP sequences, observers were first trained until they reach a fixed performance level for each task (see below). Eleven subjects participated in this experiment. All were right handed and had a normal or corrected-to-normal vision.

Training

The stimulation sequence was similar to that of Experiment 1 but with fixed durations. Participants underwent blocked-task runs of 81 trials each. The order of the tasks was randomized across participants: six participants began the training with the shape task. Participants were first trained in runs with “long” durations (250 ms, 166.6 ms, 116.6 ms), until they reach 80% correct for a duration of 166.6 ms. They were then trained with “short” durations (166.6 ms, 116.6 ms, 83.3 ms). Training stopped once participants reached 80% correct response at 116.6 ms. Hence, participants performed at least two runs per task. One subject who participated in Experiment 1 did not need training sessions. Note that we used a 60-Hz refresh rate to maintain compatibility with ongoing MEG experiments.

For both the shape and the motion tasks, all observers were able to reach 80% of correct identification at 116.6-ms stimulus duration, although with different initial performances and different amounts of training (Figure 4). Training was performed over one or two days depending on subjects’ performances. On debriefing, participants reported the motion task to be easier than the shape task, which is consistent with the observed performances (see the respective numbers of training runs,

![Figure 4](http://jov.arvojournals.org/) Results of the training sessions at 116.6-ms stimulus duration. Left: performance for the shape task. Right: performance for the motion task. Observers were ranked depending on their initial performance (small symbols). Performances at the end of training (large symbols) are all above the 80% correct criterion. Gray bars indicate the total number of training runs performed by each subject. Subjects keep the same label across figures.
in Figure 4). Note that the number of runs needed to achieve training is inversely correlated with initial performances (shape task: $r = -0.87$, $p = 0.001$; motion task: $r = -0.74$, $p = 0.014$). As will become clear in the following, the existence of “fast” and “slow” learners, resulting from different initial performances, is relevant to account for the pattern of results for the shape task in the RSVP experiment, presented below.

**Methods and stimuli**

3D-SFM stimuli were presented in a rapid serial visual presentation (RSVP) with a duration of 133 ms. Dynamic masks were displayed before and after the sequence (see General methods section and Figure 2C). In the “Test trials” (80% of the trials), three stimuli were displayed sequentially. The rest of the trials, “Catch trials,” comprised two 3D-SFM stimuli, in order to limit the bias of systematically reporting three objects. As already stated in the General methods section, dot configuration did not vary with the underlying shape but depended on the position of the stimulus in the sequence: one configuration was used for all 3D surfaces presented as the first or third item, another one was used for the second item.

In test trials, the first critical item (C1) and the second critical item (C2) were either identical (“repeated” condition) or different (“non-repeated” condition). The distractor (D), presented in between C1 and C2, was always different from C1 and C2. Since each stimulus was studied along 2 dimensions, test trials were divided in 4 categories: repeated motion direction and shape; non-repeated motion direction and shape; repeated shape with non-repeated motion direction; non-repeated shape with repeated motion direction. For each task, there were 216 repeated trials, 216 non-repeated trials, and 108 catch trials.

Subjects performed blocked-task runs of 90 trials. Before each run, they were instructed to process one dimension of the 3D-SFM stimuli, shape, or motion—the “relevant” dimension—and to ignore the other attribute—“irrelevant” dimension. They were also informed that three or less 3D-SFM stimuli could be presented in a trial. At the end of a trial, a response screen representing the 3 shapes or the 3 directions was displayed, depending on the current task. In order to limit the involvement of semantic or lexical information, each item of the response screen was highlighted by a yellow border in a random order and observers indicated whether the selected item appeared 0, 1, or 2 times during a trial. To allow resting breaks during a run, we used a self-paced procedure whereby observers initiated a new trial by pressing the “space key.”

The experiment was completed over 3 days, subjects performing 4 runs per day (two runs of shape task, two runs of motion task). The delay between the first and the last sessions of experiments ranged from 3 to 45 days. Task order was randomized across subjects. In addition, subjects who began the training with the shape task started this experiment with the motion task to minimize a possible priming effect at the individual level.

Note that the duration of all 3D stimuli in the RSVP sequence (133 ms) is one more frame than the 116.6-ms duration used in the training phase and thus well above the shape and motion duration thresholds measured in Experiment 1. Although identical stimulus duration for both tasks makes motion task easier than shape task, durations depending on observers’ sensitivity to shape and motion may prevent shape processing during the motion task, such that a potential effect of the irrelevant shape would go undetected. The choice of a common duration that accommodates both the shape and motion thresholds allows to evaluate the influence of the relevant and irrelevant dimensions in all conditions.

**Analysis**

Correct responses for non-repeated trials correspond to C1 and C2 both being reported once, and D being reported once at most. Correct responses for repeated trials correspond to C1 being reported twice exactly and D being either correctly identified or omitted. Any other pattern of responses is considered an error. This measure is referred to as “performance” in the following. The difference between percentages of correct responses for non-repeated and repeated trials is called “repetition effect:” a positive difference corresponds to an RB effect, while a negative difference corresponds to a repetition advantage (RA) effect. Catch trials are not used in estimating the repetition blindness rate. With these constraints, chance level performance is equal to 11.1% correct. This analysis uses a strict criterion—no false alarms are allowed—to analyze RB (Arnell & Jolicoeur, 1997).

**Results**

We first conducted a $2 \times 2 \times 2$ repeated-measures ANOVA, with factors Task (motion/shape) × Relevant dimension (repeated/non-repeated) × Irrelevant dimension (repeated/non-repeated) on the mean percentage of correct responses for test trials (see data in Figure 5). This ANOVA indicates that performances significantly depend on the task ($F(1, 10) = 5.298, p = 0.044$), the motion task yielding better performances than the shape task. The repetition of the relevant and irrelevant dimensions have a significant effect on performances (relevant dimension: $F(1, 10) = 20.96, p = 0.001$; irrelevant dimension: $F(1, 10) = 6.64, p = 0.027$). There are also significant crossed effects with the Task (Task × Relevant dimension: $F(1, 10) = 15.05, p = 0.003$ and Task × Irrelevant dimension: $F(1, 10) = 8.85, p = 0.014$), reflecting the fact that repetition effects are mainly found in the motion task. We hence analyzed
separately the data for the motion and shape tasks using a $2 \times 2$ repeated measures ANOVA (Relevant dimension × Irrelevant dimension).

In the motion task, performance is significantly lower when the direction of motion is repeated than when it is not ($F(1, 10) = 34.77, p = 0.00015$). This corresponds to a mean RB rate of 25.38% (Figure 5A) with individual RB rates ranging from 3.08 to 55.85%. Unexpectedly, performance is significantly better when the irrelevant 3D shape is repeated than when it is not ($F(1, 10) = 9.07, p = 0.013$; Figure 5B). This effect of the irrelevant shape does not interfere with the RB rate, which is the same whether shape is repeated or not (Relevant × Irrelevant dimension: $F(1, 10) = 0.136, p = 0.72$).

In the shape task, the effect of the shape repetition is not significant at the group level ($F(1, 10) = 0.317, p = 0.59$), suggesting a lack of RB (Figure 5C). The repetition of the irrelevant motion direction has no significant effect either ($F(1, 10) = 0.072, p = 0.79$; Figure 5D). Plotting the individual repetition effect in Figure 6A however indicates large interindividual differences (range −14.9% to +23.4%). Overall, 5 subjects show a positive rate (“RB” subjects) while the remaining 6 subjects show a negative rate indicative of a repetition advantage (“RA” subjects). This unexpected distribution of the repetition effects presumably accounts for the lack of a significant RB effect at the group level. In the following, we further explore the interindividual variability.

Although the repetition effect seems to follow a bimodal distribution (Figure 6C), this distribution is not significantly different from normal (Kolmogorov–Smirnov test: $D = 0.193; p > 0.2$), presumably because of the limited number of data points ($N = 11$). The homogeneity of the population was hence tested with a single linkage dendrogram. This hierarchical clustering analysis uses Euclidian distances between individual data and is thus not biased by the sign of each value. The analysis indicates that observers fall in two distinct groups (Figure 6D) that nicely fit the classification in RA and RB subjects. A $t$ test confirms that the effect of repetition is significantly different between the two groups ($t(1, 9) = −7.9, p < 0.0001$). ANOVAs conducted separately on both groups reveal a significant effect of repeating the relevant dimension (RA group $F(1, 5) = 22.48, p = 0.0051$; RB group $F(1, 4) = 37.26, p = 0.0036$) corresponding to RA and RB effects, respectively (Figure 7). The effect of repeating the
Figure 6. Population analysis for the shape (red) and motion (green) tasks. (A) Individual repetition effects in the shape task ranked as a function of the repetition effect. Subjects showing a positive difference of performance between non-repeated and repeated shape trials are classified as RB, subjects showing a negative difference are classified as RA. (B) Learning effect plotted as a function of the repetition effect for each observer. The continuous blue line represents the correlation between ‘learning index’ and repetition effect. (C) Distribution of the individual repetition effects. (D) Hierarchical clustering of repetition effects for the shape task. Vertical axis indicates distances between nodes. Intermediate nodes are formed using single linkage aggregation. The resulting dendrogram tree reveals two distinct and homogenous groups, with large cumulative distances between groups and small distances within each group. This clustering in two groups fits the segregation between RA and RB rates. (E) Individual repetition effects for the motion task. The positive differences of performance between non-repeated and repeated shape trials for all observers indicate RB. Subjects have been ranked depending on their RB rate. (F) Correlation graph between the ‘learning index’ and the individual repetition effect ($r = 0.051; p = 0.88$). (G) Distribution of the individual repetition effect (Kolmogorov–Smirnov test: $d = 0.169; p > 0.2$). (H) Hierarchical clustering of repetition effects for the motion task. The resulting dendrogram tree reveals a homogenous population. One participant, trained previously, is not shown in these figures.
irrelevant dimension is however not significant (RA group: $F(1, 5) = 2.7$, $p = 0.16$; RB group: $F(1, 4) = 4.01$, $p = 0.12$).

These different behaviors are unlikely to be accounted for by different performances in the shape task, as all observers reached 80% correct at a 116.6-ms stimulus duration during the training phase. They, however, did so after different numbers of training sessions. Seeking a link between the observation of 2 groups in the RSVP experiment and the time taken to reach this level of performance, we calculated a “learning index” contrasting the initial and final performances of the training phase ([Final − Initial] / [Final + Initial] at the 116-ms duration, see Figure 4). Interestingly, this learning index is negatively correlated with the repetition effect: subjects with highest learning indices show RA; subjects with lowest learning indices show RB (Figure 6B). Different measures of “learning”—e.g., the number of runs required to reach the criterion—were tested with similar outcomes. Overall, subjects with initial performance close to the 80% correct criterion show RB; subjects with initial performances far from this criterion show RA. Note that task order does not account for this group classification as 3 of the 6 RA subjects and 3 of the 5 RB subjects began with the motion task.

Whether a similar differential effect exists in the motion data, specifically when the irrelevant shape is repeated, was tested by analyzing further the results of the motion task. We first tested whether a similar bimodal distribution of the repetition effect would exist (Figure 6). Despite the repetition effect being more variable across observers for the motion than for the shape task (Figure 6E), the histogram and dendrogram analyses (Figures 6G and 6H) do not reveal any clustering of the data into sub-groups. Moreover, the learning index and the repetition effect are not correlated (Figure 6F; $r = 0.051; p = 0.88$).

We then analyzed the results of the motion task using the RA and RB groups identified for the shape task as a categorial factor. Neither simple ($F(1, 9) = 0.19; p = 0.67$) nor crossed (group × motion repetition: $p = 0.46$; group × shape repetition: $p = 0.74$) group effect was found in the motion task. Overall, these results assert that the population is homogenous with respect to the motion task and validate the whole group analysis performed for this task.

To summarize, in the shape task, repeating the relevant shape yields either RB or RA depending on the identified groups, with no effect of the irrelevant motion attribute. In the motion task, motion repetition yields an RB effect and...
the repetition of the irrelevant shape attribute increases performance, independently of the identified groups (see Figure 8).

Discussion

The aim of the present study was to determine whether attentional selection of shape and motion information during rapid serial visual presentation of 3D-SFM stimuli is influenced by the repetition of the attended and/or the unattended attribute. Briefly, we found an influence of both the relevant attended and irrelevant unattended attributes in the motion task. For the shape task, repeating the shape or the motion had no effect at the group level. Analyzing the results of two populations distinguished on statistical grounds as well as on their training performance, we found either repetition blindness (RB) or repetition advantage (RA) when the relevant shape was repeated, with no effect of the irrelevant motion dimension.

This pattern of results is unexpected and may provide insights into the type/token dichotomy and the deployment of attention toward 3D-SFM stimuli characterized by a tight coupling of shape and motion information. In the following, we first discuss the effect of repeating the relevant dimension separately for the motion task and shape task. We then discuss the results for the irrelevant dimension and analyze how these asymmetrical effects may shed light on the possible interactions between the processing of both dimensions.

Repetition blindness for motion

To our knowledge, the finding of a significant RB effect for motion direction has not been reported before. The paradigm and results of this study make us confident that this effect neither results from a physical continuity of the motion percept for the repeated motion nor reflects a physiological adaptation to local motion. First, the three successive SFM stimuli of the RSVP sequence were generated from different spatial dot distributions and the starting position of the dots was never in motion continuity with the end of the previous display. It is thus unlikely that observers processed an RSVP sequence as a single motion flow or event. Second, if physiological adaptation was to account for the results, one would expect lower performance when stimuli are physically alike, i.e., when both attributes are repeated. Contrary to this prediction, the lowest performance was found for unRepeated shape. Hence, decreased performance for repeated motion likely corresponds to a genuine repetition blindness effect for motion.

No repetition blindness for shape?

We did not find RB for shape at the group level, which may appear surprising in the light of previous studies (Arnell & Jolicoeur, 1997; Harris & Dux, 2005). The observation that a sub-population shows an RB effect partly reconciles our data with these previous studies. That some observers experienced RB with 3D-SFM stimuli suggest that processing shape in SFM stimuli is not fundamentally different from processing static shape. The finding of another sub-population showing an RA effect is more unexpected.

Although not often documented in the RSVP literature, RA effects have been described for nonwords as opposed to RB for words (Campbell, Fugelsang, & Saskatchewan, 2002; Coltheart & Langdon, 2003). Coltheart and Langdon (2003) were able to improve the level of recall of nonwords with training but found no effect of training on RA rate. Harris and Morris (2004) found an RB effect for both words and nonwords, although smaller for the latter, suggesting a diminished capability to encode nonwords. Interestingly, Morris and Still (2008) obtained both RA and RB with either words and nonwords, depending on the task or instruction given to the subjects. These studies suggest that RA and RB phenomena may depend on observers’ ability to encode stimuli in a particular task context.

Paradoxical relationship between tokenization and task-related encoding ability

In line with the above studies, our data point toward a close relationship between observers’ ability to identify the relevant stimuli and the “tokenization” process, estimated herein through measures of RB rate. This can be seen not only through the behavioral differences between the motion and shape tasks (RB only for the “easier” motion task vs. RA and RB for the shape) but also through individual differences in the shape task (RA vs. RB depending on the learning index). We are not aware of other studies showing such relationship between RB rate and subjects’ general performance in the task. The inverse correlation found here may however appear paradoxical, as one would expect that the easier it is to process a particular dimension, the easier it should be to report its successive occurrences. One explanation of this seemingly paradoxical effect relies on a functional interpretation of the role of the tokenization process.

In our environment, it is necessary to distinguish between the presence of a single object seen twice in the same location and the presence of similar objects in two locations (Kahneman et al., 1992). Object tokenization is at the core of this process. Accordingly, an identical object seen twice at a single location likely corresponds to a unique object in the real world and should yield the record of one object file only, the second occurrence of
this object being ignored. However, the outcome of this process may depend on processing depth and accuracy as slight differences in the object or its location may, or not, go unnoticed. Omitting the second occurrence of the same object should occur only if both occurrences of an object have been reliably processed and a specific type unambiguously activated. Conversely, if the first and/or second occurrence of a stimulus is only coarsely processed, observers may identify both occurrences as belonging to the same type but fail to notice the stimulus is exactly “the same” and hence attribute a second token to the same object, thus turning the RB rate into an RA effect.

From RA to RB: A continuous phenomenon?

It is noteworthy that RA and RB rates do not only correspond to separate subject categories or different tasks but also correlate with a continuous measure (learning index). This correlation, reminiscent of the shift from facilitatory priming effects for subliminal stimuli to suppressive masking effects for suprathreshold stimuli, suggests the existence of a continuum between RA and RB effects. In this regard, exposure duration relative to the individual sensitivity could be critical. Indeed, observers experience RB whenever the duration of an event in an RSVP sequence is long relative to their duration threshold. This is clear for the motion task where the duration of a 3D-SFM motion event (133 ms) is much longer than the duration threshold (~50 ms). This might also be true in the shape task for those fast learners having a small learning index. Although it remains unclear whether these learning indices reflect different sensitivities to spatial and temporal characteristics of SFM stimuli or different learning skills, they likely reflect variations in the way observers perform the task (see Coltheart & Langdon, 2003; Harris & Morris, 2004). Although speculative, the question of exposure duration relative to subjects’ sensitivity could be tested in future experiments by correlating individual thresholds with repetition effects while varying the duration of each event of an RSVP sequence.

No effect of irrelevant motion repetition in the shape task

As observers experience repetition blindness for motion, we assumed that repeated motion would impair the buildup of a reliable motion-dependent response to shape, resulting in increased RB rate or overall decreased performance in the shape task (see Figure 1B). That the repetition of the irrelevant motion has no effect on either the RB rate or the percentage of correct responses in the shape task has two implications. One is that RB occurs when attention is drawn toward the attended dimension but does not occur for the unattended dimension. This is in line with the study by Kanwisher et al. (1995) showing that RB occurs only when attention is allocated to a cued dimension. A second implication, specific to the 3D-SFM stimuli used here, is that repetition of motion-direction information does not interfere with object identification or shape tokenization (Figure 8B). Whether interferences would be found by repeating the local motion direction rather than the global 3D object motion remains to be explored.

Unexpected effect of irrelevant shape repetition in the motion task

As processing 3D-SFM stimuli supposedly involve a hierarchical motion-to-shape flow, we did not expect, and did not find, that repeating the irrelevant shape influences the rate of RB to motion (see Figure 1A). Although this indicates that shape repetition does not alter token attribution, we found that the percentage of correct responses is higher when an irrelevant shape is repeated (see Figure 8A). This could reflect a relative advantage for repeated shape trials or a relative disadvantage of non-repeated shape trials. In the following, we examine whether attentional and perceptual mechanisms could account for this effect.

When the shape is not repeated, the occurrence of a new shape could draw observers’ attention away from the ongoing motion task, which would entail decreased performance (see Figure 5B). According to the role of attention in RB discussed above, reducing the attentional resources allocated to the motion task should also reduce the RB rate. This is not what we observe, as RB rates are independent of shape repetition. It thus appears unlikely that attentional shifts can account for the results.

Another related possibility is that observers direct their attention toward both the relevant and irrelevant attributes. In explicit shared attention conditions, Kanwisher and Driver (1992) found that the repetition of either attribute of an object can induce RB. If observers shared their attention between motion and shape in our study, the repetition of the shape should yield decreased performances and increased RB rate. That we find just the opposite argues against the shared attention hypothesis. This conclusion is further supported by the finding of two groups with different performances in the shape task. If attention was shared between motion and shape, the differential processing of shape in these groups should also exist in the motion task. This is not the case, as no clustering in two groups could be found in this latter task. Hence, attention-related mechanisms cannot convincingly account for the effect of the irrelevant attributes.

An alternative interpretation, based on perceptual rather than attentional mechanisms, relies on the incremental rigidity scheme proposed by Hildreth et al. (1990) and Ullman (1984). These authors suggested that an accurate
perception of SFM stimuli is built up incrementally so that the perceived 3D structure depends on progressively refined internal 3D models. In RSVP sequences, the duration of each event may be too short to achieve this process for each item separately. The first target could however provide a model for the following events. In this view, inferences following the identification of the C1 shape could modulate the responses to C2, such that a mismatch between the inferred and actual C2 motion distribution would impair motion identification. This would explain why performance is better for repeated shapes in the motion task. Although speculative, this explanation could be formalized using the recently developed predictive coding scheme (Creutzig & Sprekeler, 2008; Kersten, Mamassian, & Yuille, 2004; Langley & Anderson, 2007; Montagnini, Mamassian, Perrinet, Castet, & Masson, 2007), according to which output signals from a cortical structure reflects the gap between inferences predicting how the stimulus should be and the actual distal stimulus. Additional experiments are however needed to test this hypothesis further.

3D-SFM processing and ventral and dorsal pathways

The above discussion considers motion and shape with regard to attentional selection and the type/token theory independently from the neural bases of 3D-SFM processing. Experimental evidence suggests that motion and shape are processed along distinct visual—dorsal and ventral—streams (DeYoe & Van Essen, 1988; Mishkin & Ungerleider, 1982), whether the activity is driven by visual stimulation or attentional selection (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Paradis, Droulez, Cornilleau-Pèrès, & Poline, 2008). In addition, electrophysiological (Treue & Maunsell, 1999) and fMRI (Kriegeskorte et al., 2003; Murray, Olshausen, & Woods, 2003; Orban, Sunaert, Todd, Van Hecke, & Marchal, 1999; Paradis et al., 2000) studies found that areas from both the ventral and dorsal streams are involved in the perception of 3D-SFM stimuli. How motion information gives rise to object recognition however remains an open issue. Because a 3D stimulus can be identified only after the retinal motion distribution has been processed, one possibility is that information from dorsal areas transfers to ventral areas. Our findings that shape repetition influences motion identification and that motion repetition does not influence shape processing do not easily fit with a functional scheme involving the sequential recruitment of dorsal to ventral areas. They rather suggest that projections from ventral areas to dorsal areas modulate motion processing, a view supported by the observation that static images with implied motion elicit responses in motion areas (Kourtzi & Kanwisher, 2000) or by the finding of strong influences of form on motion processing (Lorenceau & Alais, 2001). However, recent imaging studies with static stimuli question the notion of parallel processing of motion and form through dorsal and ventral pathways (Konen & Kastner, 2008). The present results suggest that 3D-SFM stimuli are well suited to study further the functional dependence of shape and motion processing.

Acknowledgments

We thank Jean-Claude Ardouin and Véronique Chastres for their assistance with data analysis. AM was supported by a doctoral grant from the Délégation Générale pour l’Armement. This work was supported by an MRIS-DGA grant (2006.34.059) to JL. Part of this work has been presented at the Annual Meeting of the Vision Sciences Society, Sarasota, Florida, May 2007.

Commercial relationships: none.
Corresponding author: Agnieszka Miskiewicz.
Email: miskiewicza@yahoo.fr.
Address: CNRS, UPR640, LENA 47 bd de l’hôpital 75651 Paris Cedex 13, France.

References


O’Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units...


