Visual ZIP files: Viewers beat capacity limits by compressing redundant features across objects

Article *in* Journal of Experimental Psychology Human Perception & Performance · October 2020

DOI: 10.1037/xhp0000879

Hauke S. Meyerhoff

Leibniz-Institut für Wissensmedien

Mike Stieff

University of Illinois at Chicago

Mary Hegarty

University of California, Santa Barbara

Visual ZIP files: Viewers beat capacity limits by compressing redundant features across objects

Hauke S. Meyerhoff[1], Nicole Jardine[2], Mike Stieff[3], Mary Hegarty[4], and Steve Franconeri[2]

[1]Leibniz-Institut für Wissensmedien, Tübingen, Germany

[2]Northwestern University, Evanston, IL, USA

[3]University of Illinois at Chicago, IL, USA

[4]University of California Santa Barbara, CA, USA

Running Head : Visual ZIP files

Keywords : change detection, mental rotation, structure change, compression, visual zip file, working memory, capacity limitation

Address for correspondence : Hauke S. Meyerhoff
Leibniz-Institut für Wissensmedien
Schleichstr. 6
72076 Tübingen
Germany

Email : h.meyerhoff@iwm-tuebingen.de

Phone/Fax : +49 7071 29-75612 (Phone), +49 7071 979-115 (Fax)

Word count : 7794

Abstract

Given a set of simple objects, visual working memory capacity drops from 3-4 units down to only 1-2 units when the display rotates. But real-world STEM experts somehow overcome these limits. Here, we study a potential domain-general mechanism that might help experts exceed these limits: compressing information based on redundant visual features. Participants briefly saw four colored shapes, either all distinct or with repetitions of color, shape, or paired color+shape (e.g., two green squares among a blue triangle and a yellow diamond), with a concurrent verbal suppression task. Participants reported potential swaps (change/no change) in a rotated view. In experiments 1A-1C, repeating features improved performance for color, shape, and paired color+shape. Critically, Experiments 2A-2B found that the benefits of repetitions were most pronounced when the repeated objects shared both feature dimensions (i.e. two green squares). When color and shape repetitions were split across different objects (e.g., green square, green triangle, red triangle), the benefit was reduced to the level of a single redundant feature, suggesting that feature-based grouping underlies the redundancy benefit. Visual compression is an effective encoding strategy that can spatially tag features that repeat.

186 words

Public Significance Statement

The ability to compare objects across a rotation is limited to extremely simple objects, yet STEM experts such as chemists appear to circumvent this limitation. Understanding the limits of visuospatial thinking and mechanisms of overcoming these limitations is crucial for developing training supporting STEM-relevant abilities. Here, we study a domain-general mechanism that might allow people to exceed known limitations – leveraging redundant feature information. Our results show that the ability to detect swaps between rotated views is higher when people leverage redundant feature information, but that this advantage is limited to a single group of objects that share the same set of redundant features.

Visual ZIP files: Viewers beat capacity limits by compressing redundant features across objects

Visuospatial abilities such as mental rotation (Shepard & Metzler, 1971) have been identified as a critical component of success in the field of science, technology, engineering, and mathematics (STEM; National Research Council, 2006; Wai, Lubinski, & Benbow, 2009). As many pupils and students face problems with visuospatial demands, one strategy is to train these skills early in life (Newcombe, 2016). Domain-specific training is capable of substantially improving performance and learning (e.g., Kellman, 2013). Moreover, there is now good evidence that specific spatial skills (such as mental rotation) can be trained (Uttal et al., 2013) and that this training can transfer to other spatial tasks. However, to date, training domain-general visuo-spatial skills has led to only small improvements in STEM performance (Cheng & Mix, 2014; Sorby, 2009; Stieff & Uttal, 2015), and these improvements further shrink when generalized beyond the trained task or over time (Miller & Halpern, 2013).

In the present experiments, we explore the domain-general strategy of exploiting redundant visual features (repetitions of colors and/or shapes), as they exist in multiple kinds of STEM representations. For example, the spatial structure of colors and shapes all convey critical information in molecular representations in chemistry (see Figure 1), a domain where expert chemists have a superior ability to detect changes to this structure between different views (Stieff, 2007).

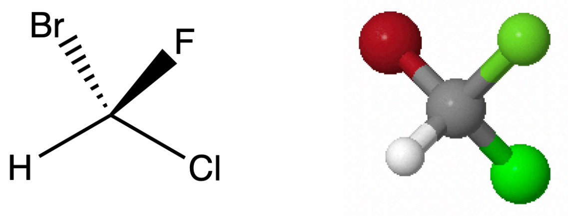

Figure 1. Illustration of molecular representations in organic chemistry. Dash-Wedge representations (left) mainly consist of shape information whereas Ball-and-Stick representations (right) consist of color and size information.

Given that these molecular structures typically include redundant feature information (repeated colors and shapes), we tested whether viewers would exploit these redundancies as a domain-general strategy to overcome capacity limitations. We translated this application-inspired question into a laboratory test that isolated distinct redundant visual features (colors and/or shapes). We also tested whether this redundancy advantage extended to displays where the redundant features were split across different objects. To anticipate our results, they show that human observers can compress redundant visual feature information to overcome their capacity limitations for detecting changes to the spatial structure of these features between rotated views, but that this ability is limited to a single set of objects sharing redundant feature information.

Detecting changes between rotated views

A central task for the visual system is to recognize objects as same or different even when they are presented from different viewing angles. However, the mental process underlying this ability has been controversial (see Peissig & Tarr, 2006, for a review). Shepard and Metzler (1971) asked participants to indicate whether two sets of concatenated cubes presented from different viewpoints were identical or mirror images of each other. They observed a linear relationship between increased angular disparity and increased response times, suggesting that their participants were continuously mentally rotating one of the objects in order to solve the comparison task. But such linear performance degradation with greater angular disparity does not necessarily indicate continuous mental rotation, because this effect can also be observed in object recognition tasks that do not appear to involve mental rotation (Edelmann & Bülthoff, 1992; Cheung, Hayward, Gauthier, 2009; Hayward & Williams, 2000, Jolicœur, 1985; Tarr & Pinker, 1989; Tarr, Williams, Hayward, & Gauthier, 1998). These object recognition tasks revealed distinct patterns of neural activation compared to rotation (Gauthier et al., 2002) showing that distinct mental processes could lead to a comparable linear decline in performance with increasing angular disparity. With the present experiments, we study the accuracy with which human observers can detect changes in a spatial structure of four objects following a rotation of the display. Our experiments are not intended to differentiate the potential mechanisms of continuous mental rotation from other object recognition processes that are similarly adversely affected by angular changes. Therefore, we will use the agnostic term 'structure change detection' to describe our task in which participants detect changes in a spatial structure of objects and their features between two rotated views, rather than the more theoretically loaded terms 'mental rotation' or 'object recognition'.

Capacity Limitations in Detecting Structure Changes between Rotated Views

In order to refer to the amount of information that can be rotated "at once" we have borrowed the term 'capacity limitation' from research on working memory (e.g. Cowan, 2001). Although this term is typically not used in mental rotation studies (but see Just & Carpenter, 1985; Shah & Miyake, 1996), asking whether complete objects or structures can be rotated at once can adapt the definitions of capacity limitations used in working memory research. Most of the work addressing these limitations has focused on the envelope of 3D- and 2D objects such as block figures (e.g., Shepard & Metzler, 1971), drawings (e.g., Pylyshyn, 1979), and polygons (e.g. Cooper & Podgorny, 1976). For connected shapes, some classic work has argued in favor of a virtually unlimited capacity resulting in so-called holistic rotation patterns (i.e. the entire shape at once; see Cooper & Podgorny, 1976), whereas other work suggested piecemeal rotations of parts of the full shape (i.e. sequential; Folk & Luce, 1987; Just & Carpenter, 1985; Yuille & Steiger, 1982).

Crucially, however, whether or not an object is rotated holistically or in a piecemeal manner is not entirely determined by the rotated objects themselves, but also depends on the spatial abilities and strategies of the observers. For instance, Khooshabeh, Hegarty, and

Shipley (2013) observed that participants with poor spatial abilities tend toward piecemeal strategies whereas observers with good spatial abilities are more likely to employ holistic strategies. However, other research has identified flexibility in the selection of strategies associated with good spatial abilities (Botella, Peña, Contreras, Shih, & Santacreu, 2009; Nazareth, Killick, Dick, & Pruden, 2019). Whether people use a piecemeal or holistic strategy also depends on their familiarity with the stimuli (Bethel-Fox & Shepard, 1988). Another correlate of individual differences in many of these mental rotation tasks is sex (with a male advantage in most cases; Hyde, 2005; Voyer, Voyer, & Bryden, 1995) although the causal origin for such sex difference are far from understood (Voyer, Saint-Aubin, Altman, & Doyle, in press)[1].

Many of the mental rotation studies have relied on detecting changes to the shapes of objects, so that the features of parts of these objects (e.g. colors) and their spatial structure of where those colors occurred ('bindings') are not tested. But many real-world STEM rotation tasks require maintaining these bindings, and this requirement leads to severe capacity limitations. Xu and Franconeri (2015) reported a set of experiments in which participants monitored a cross-like object consisting of four distinctly colored legs for color changes between rotated views. A capacity analysis revealed that they were only able to maintain a single color attached to its corresponding leg across a 90 degree rotation of the layout between the views (for similar results, see Saiki, 2003). This capacity limitation for location-feature bindings across rotations contrasts with the observation that chemists are relatively accurate in detecting changes between rotated molecular structures (Stieff, 2007).

We use the metaphor of a 'Zip File' to reflect the fact that noticing and compressing redundancies in a representation is a general information-theoretic strategy for fitting information into a limited capacity storage system. In a computer, an algorithm detects

[1] In the present project, we focus on general cognitive processes which we consider to be present in participants of *both sexes*. In terms of sample size as well as male/female composition, our experiments are not designed to investigate sex differences such as differently strong manifestations of effects. Nevertheless, we screen all our analyses for sex differences and report them in the few cases where we found any.

redundant information which is stored only once, and then points to the original positions of the redundant copies. In our examples, sources of redundant information are color and shape information. Observers might use such redundant information to encode compressed versions of the stimuli. Following display rotations, the stimulus could be decompressed in order to map the unrotated representation with the actual test display. The Zip File analogy generates an intriguing prediction that we will test in Experiment 2. If the redundancy is at the level of an entire repeated object, with a pointer to the spatial positions of those same objects, participants should have difficulty leveraging redundant features that are *split* across multiple objects (e.g., a red square, a red triangle, and a blue triangle).

Indeed, research on visual search has demonstrated that a unique object sharing multiple feature dimensions can be found more efficiently than objects sharing only one feature dimension (Krummenacher, Müller, & Heller, 2001; Nothelfer, Gleicher, & Franconeri, 2017; Wolfe, Cave, Franzel, 1989). Further, recent research on visual short-term memory (i.e., pure recall of feature-location bindings) has revealed beneficial effects of redundant stimulus information. For instance, Brady and Tenenbaum (2013) showed that participants were more likely to detect color changes within a briefly memorized layout of squares when neighboring objects were of the same color (see also Peterson & Berryhill, 2013, for a similar finding). Such findings show that short-term memory does not store each object in a visual display independently, but instead stores information hierarchically, taking advantage of redundancies and other statistical summary information (for a review see Brady, Konkle, & Alvarez, 2011). In fact, in some studies, the compression of redundant visual information saved memory resources so much that memory performance did not only increase for the items with redundant information but also spilled over to the remaining items (e.g., Morey, 2019; Thalmann, Souza, & Oberauer, 2018). Such spillover effects have been observed for color (Quinlan & Cohen, 2012; Lin & Luck, 2008; Morey, Cong, Zheng, Price,

& Morey, 2015) as well as for shape information (Mate & Baques, 2009; but see Quinlan & Cohen, 2012).

The Present Study

In the present manuscript, we explore how leveraging such redundancies can lead to improved change detection performance for a structure consisting of four objects (bindings between colors/shapes to particular objects within a spatial layout). The present studies also differ from previous work by asking participants not to detect new colors or shapes, but to detect *swaps* after a spatial transformation (a display rotation). Such structure change detection is a much more difficult operation, and one that forms a challenging problem for STEM thinking (Stieff, 2007).

We test three hypotheses. First, we propose that redundant visual information leads to improved structure change detection performance between the display rotations. We argue that it is reasonable to assume beneficial effects of redundancy for structure change detection as previous research has demonstrated clear links between spatial working memory – which benefits from redundancy (see above) - and change detection performance across rotated views (Shah & Miyake, 1996). We will refer to this possibility as the redundancy-boost hypothesis. Second, we investigate the hypothesis that the benefit of redundant objects during structure change detection spills over to the remaining unique items in the display (i.e. the reduced demand of processing two redundant rather than two distinct objects improve performance for the remaining non-redundant objects). In working memory experiments, such spill-over effects arise from the reduced demand of redundant stimuli on a limited working memory resource, which remains more available for other objects. Given the strong connection between working memory and performance at detecting changes across rotated views (Shah & Miyake, 1996), it therefore is plausible to expect such spill-over effects for structure change detection too. However, since processing rotated views reflects an active internal process (depending on executive control, Baddeley, 1986; or controlled attention, Engle, Tuholski, et al., 1999) rather than passive memory, which comes along with more severe capacity limitations, it cannot be taken for granted that such a boost for non-redundant objects (i.e. a spill-over) also applies to structure change detection across rotated views. We will refer to this possibility as the spill-over hypothesis.

Third, in order to explain the origin of redundancy-boost effects in structure change detection across rotated views, we test the possibility that the benefits for redundant stimuli arise from a feature-based grouping mechanism (i.e. joint attending to a set of objects sharing a basic feature such as color or shape). This feature-based grouping mechanism would limit the number of redundant objects that participants can take advantage of simultaneously. The rationale for this hypothesis arises from recent research arguing that grouping objects by similar features can only occur for a single group at a time, by isolating a set of objects that have a common set of feature values (e.g., just the red ones, or just the squares, or red squares; Huang, & Pashler, 2007; Huang, Treisman, & Pashler, 2007; Yu, Tam & Franconeri, 2019; Yu, Xiao, et al, 2019). In our experiments we test this redundant feature information which is either shared between the same objects or split across distinct subgroups of objects. We refer to this possibility as the feature-based grouping hypothesis.

Rationale of Experiment 1a-c

In this set of three experiments, we test the first two of our hypotheses, the redundancy-boost and the spill-over hypotheses. We test this with redundant color (with constant shape; Experiment 1a), shape (with constant color; Experiment 1b), and combined color and shape information (Experiment 1c). If the redundancy-boost hypothesis is correct, we should observe increased structure change detection accuracy for stimuli including feature redundancies relative to stimuli without such redundancies. If the spill-over hypothesis is correct, not only objects with redundant features should elicit improved structure change detection performance. Instead – in trials in which redundant objects are present – swaps involving objects without these redundant features should also be detected more accurately as they should benefit from the reduction in processing demand for the redundant objects.

Experiment 1a

Experiment code of all reported experiments, all raw data, and all analysis scripts are available at https://osf.io/n82hj/.

Methods

Power Considerations

The most relevant statistical tests in our study focus on the comparison between displays containing redundant feature information and displays without such redundancies. We were not aware of previous experiments investigating the impact of redundant feature information on structure change detection across rotated views. Prior to testing participants, we piloted authors HM and SF as well as two naïve research assistants with the displays of Experiment 1a. Introspectively, the beneficial effect of redundant color information could be appreciated within a single trial. On the performance level, we observed strong increases in raw accuracy (more than 10%) as well as a matching strong increase in sensitivity (d') for detecting swaps between the two views (this also matches with data from working memory displays studying the beneficial effects of redundant color information, e.g., Experiment 2a in Brady & Tenenbaum, 2013). From these piloting results, we expected a large effect (i.e. $d_z >$ .8). Such an effect size would require 15 participants to achieve an acceptable level of statistical power of $(1- \beta) > .8$ at $\alpha = .05$ (G*Power, Faul, Erdfelder, Lang, & Buchner, A.). In order to compensate for potential exclusions of data (e.g. due to poor dual task performance) as well as a potentially weaker manifestation of the effect of redundant color information in real participants than highly motivated piloting subjects, we decided to test 24 participants in each experiment.

Participants

We collected data from 24 participants. For all experiments, the participants were students recruited via the local Sona system. The participants received course credit in return for approximately one hour of their time. After exclusion of three participants performing at chance level, the final samples consisted of 21 students (12 female, 18-22 years). The experimental procedure was approved by the institutional review board at a large Midwestern University.

Apparatus, Stimuli, and Procedure

The experiment was run in Python using the PsychoPy libraries (Pierce, 2007). The stimuli were presented on a 23-inch LCD monitor (60 Hz, 1920 x 1080 pixels) controlled by a MacMini at an unrestricted viewing distance of approximately 60 cm.

Our participants performed a structure change detection task with four colored geometrical shapes (square, cross, octagon, star; see Figure 2) each covering an area of 10 deg^2, which were presented against a black background. The shapes were placed 8 deg from the center of the screen (top, bottom, right, and left) and their colors were randomly drawn (without replacement) from the set of red, green, blue, yellow, magenta, and cyan. In this experiment, all objects had the same shape (drawn randomly on a trial-to-trial basis) but varied in their color (see Figure 3a). In half of all trials, one of the colors was replaced by one of the other three colors in the display, creating redundant color information.

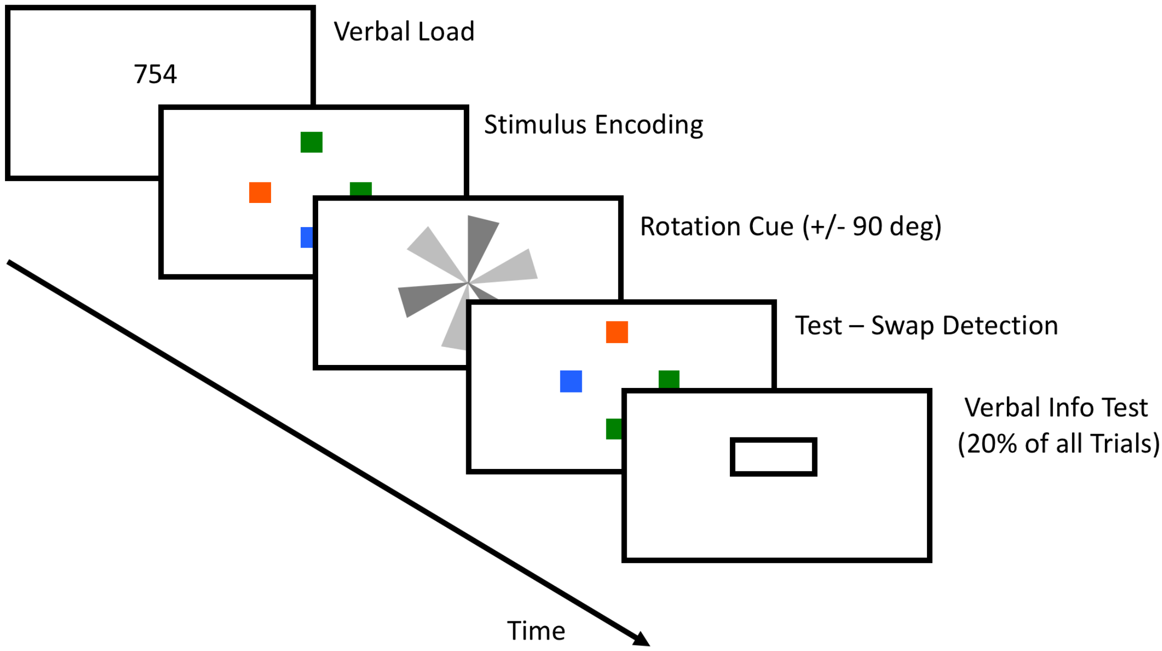

Figure 2. Illustration of the experimental task. Our participants perform a structure change detection task with a concurrent verbal load which is probed in 20% of the trials. They encode a display consisting of four colored geometrical shapes, some of which may include redundant feature information. Their task is to detect potential swaps between two of the shapes between two views. During the retention interval a rotating windmill indicates whether the entire layout is rotated 90 deg clockwise or counterclockwise. Please note that the objects were displayed against a black background while verbal load was presented in white.

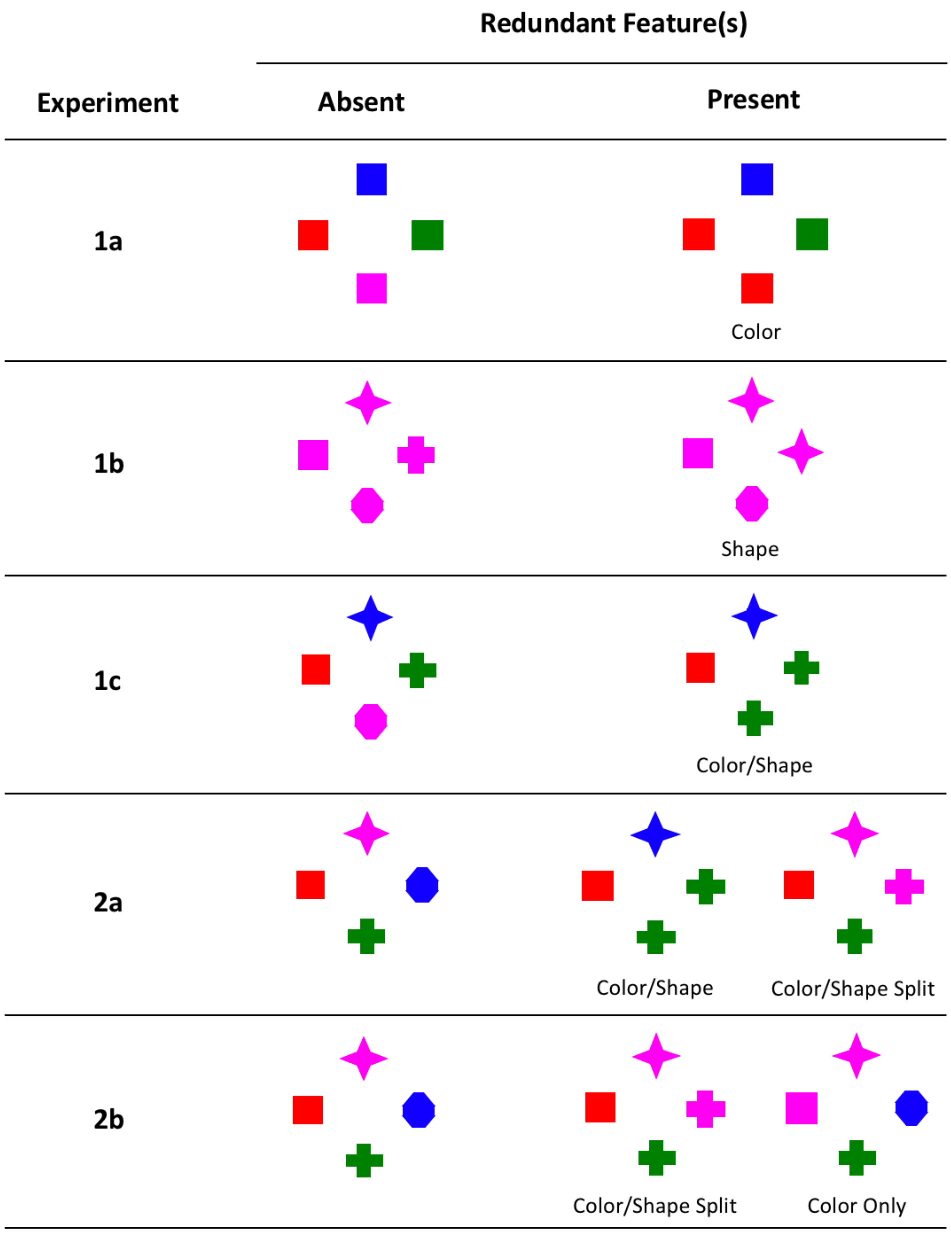

Figure 3. Illustration of the stimuli used in the five experiments. The left column illustrates displays without redundant feature information whereas the right column illustrates displays with redundant feature information. Please note that there are two conditions with different feature redundancies in Experiments 2a and 2b.

Following an initial encoding duration of 2s, the display disappeared for a retention interval of 2.2s. During the first and the last 300 ms of the retention interval, the participants saw an empty screen. During the second and the second-last 300 ms, a stationary windmill was visible. During the central 1000 ms of the retention interval, the windmill rotated 90 deg clockwise or counterclockwise, indicating the direction and extent of the display rotation to be performed. We counterbalanced the direction of the rotation to prevent participants from anticipating the direction beforehand[2]. The windmill had a radius of 10 deg of visual angle and consisted of 6 isosceles triangles (alternatingly colored light and dark grey) with an inner angle of 30 deg. Following the retention interval, the initial display reappeared with the shape layout rotated as indicated by the windmill. In half of all trials, two objects in the layout were swapped between the two views and the task of the participants was to indicate whether both views showed the same layout despite the rotation.

In order to prevent participants from encoding the layouts verbally, they performed a concurrent dual task of verbally repeating three randomly selected digits across each trial. A new random sequence was presented for 2 seconds before every trial. An experimenter was present in the room to confirm compliance with the dual task. Additionally, there was a 20% chance that participants had to enter the repeated digits after the trial. As feedback, the entered digits turned green or red for 500 ms after the response. Participants who performed the dual task below 80% were excluded from the analyses as this could potentially arise from the implementation of a verbal strategy. Average dual task performance was otherwise near ceiling between M = 97.1% and M = 98.7% across the five experiments of this report.

Prior to the experimental trials, our participants completed 8 practice trials. Thereafter, the participants completed 240 trials which fully counterbalanced the absence and presence of swaps, the direction of the rotation as well as any potential swaps within the layout. Subsequent trials were separated by a 1.5 seconds inter-trial-interval. Following blocks of 10 trials, the participants had the chance to take breaks. The entire experiment took approximately 1 hour to complete.

[2] When screening for effects of rotation direction, there were none with regard to performance measures. However, in Experiments 1b and 1c, there was a significant main effect on response bias indicating that participants were more inclined to indicate the absence of swaps for counterclockwise than clockwise rotations. As we do not interpret response bias in our study but rather report it for the sake of completeness, we will not discuss this any further.

Results

Redundancy-boost hypothesis. In order to test the redundancy-boost hypothesis, we analyze the overall effect of the presence of redundant feature information in our displays. As visible from the raw accuracy values in Table 1, there was a positive response bias in our experiments indicating that participants responded "no swap" rather than guessing between both response alternatives when they were not certain. To compensate for this response bias, we calculated sensitivity value d' as well as response bias c from signal detection theory (for this analysis the hit rate was the proportion of correctly indicating the presence of a change and the false alarm rate was the proportion of incorrectly indicating the presence of a swap). By definition, sensitivity d' and response bias c aggregate data from all trials of one participant into one value on a continuous scale. Therefore, we compared performance for these values with *t*-tests for paired samples. The presence of redundant color information improved the sensitivity for detecting swaps between the two views, $t(20) = 6.76$, $p < .001$, $d_z = 1.48$, 95%-CI [1.03, 2.54][3] (see Figure 4, upper panel; see also Table 1 for the corresponding accuracy values). There were no differences in the response criterion c between trials with and without redundant color information, $t(20) = 1.19$, $p = .25$, $d_z = 0.26$, 95%-CI [0, 0.68]. Additional exploratory analyses investigating the effect of the configuration of the redundant objects before and after the display rotation on the detectability of swaps (i.e. hits) are available in the supplementary materials.

[3] We used a bootstrapping approach with 10,000 iterations to calculate the 95%-CI of the effect size d_z.

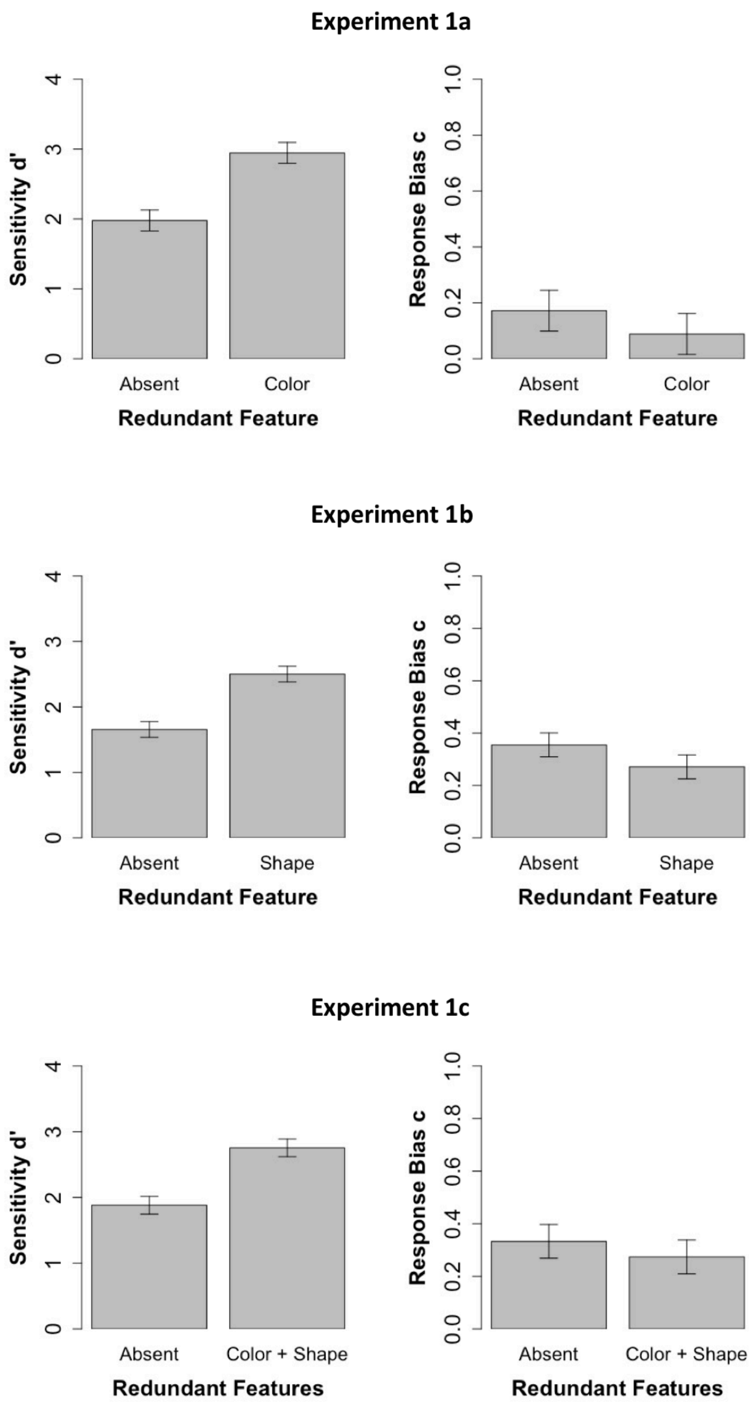

Figure 4. Results of the signal detection analysis of Experiments 1a-c. The left column displays sensitivity d'. The right column displays response bias c. The error bars indicate within-subject confidence interval.

Spill-over hypothesis.

With the spill-over hypothesis, we analyze whether the beneficial effect of redundant feature information stems from a shift towards the redundant items (i.e. selectively encoding the objects with redundant feature information) or whether it spills over to the two non-redundant objects in the display. As this hypothesis focuses on differences in the detection of different kinds of swaps, we only analyzed accuracy (i.e. hits) within those trials (see Figure 5, left panel). Given the restricted range of the dependent variable hit rate (i.e. 0-1) as well as potential restrictions in the variance of the conditions with redundant objects (i.e. performance above .8), we fit generalized linear mixed effect models with the logit as a link function to our data using the R-packages lme4 (Bates et al., 2020) The model included the intercepts of individual participants as random effects. We analyzed the differences between the conditions using Type II Wald chi-square tests (R-package car, Fox et al., 2020). Swaps that included one of the redundant objects were detected more often than swaps that included the two unique objects, $\chi^2(1) = 120.93$, $p < .001$, $R^2_m = .16$, 95%-CI [.09, .26][4]. When compared with swaps in the condition without any redundant objects (all-unique), swaps including the redundant objects were also detected more often, $\chi^2(1) = 111.91$, $p < .001$, $R^2_m =.15$, 95%-CI [.07, .26], however, swaps between the two unique objects within the redundant display were detected less often than swaps in the baseline condition without redundant objects (all-unique), $\chi^2(1) = 10.64$, $p = .001$, $R^2_m = .01$, 95%-CI [<.001, .03] (i.e. the opposite from what would be predicted by the spill-over hypothesis).

[4] We calculated the marginal R^2 as effect size for logistic mixed models (see Johnson, 2014; Nakagawa & Schielzeth, 2013) using the R-package MuMIn (Barton, 2020). This R^2_m-value expresses the variance in the data explained by the fixed factor (i.e. the type of redundancy within our study). The 95%-CIs were calculated using a bootstrapping procedure with 10.000 iterations.

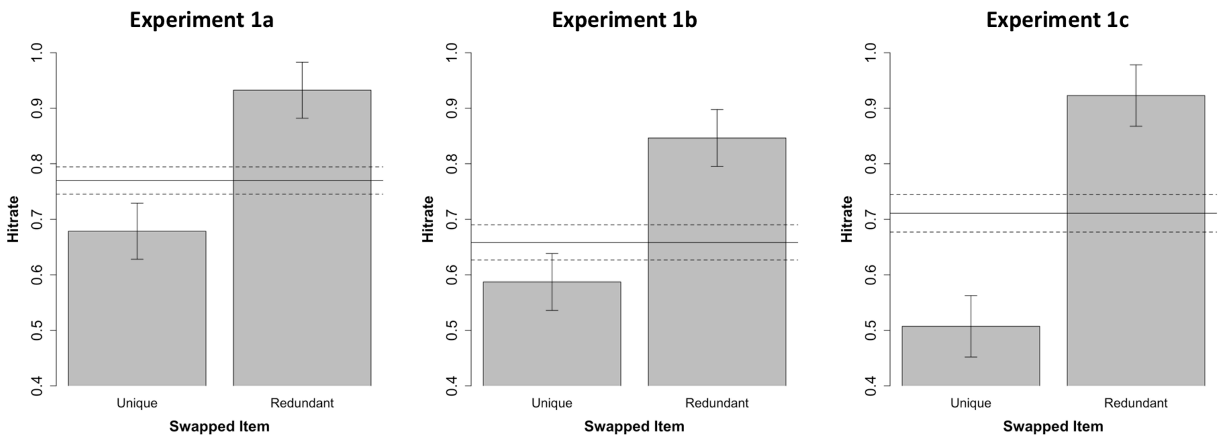

Figure 5. Hit rates for different types of swaps within the conditions with redundant objects across Experiments 1a-c. The solid lines refer to the means of the control conditions without redundant objects (i.e. the baseline), and the dashed lines refer to the corresponding within-subject confidence intervals. The error bars indicate within-subject confidence intervals.

Table 1: Proportions correct of all experiments

	swap *M (SD)*	no swap *M (SD)*
Experiment 1a		
Color Redundancy	88.2 (15.3)	92.1 (6.7)
No Redundancy	77.0 (14.9)	85.6 (9.3)
Experiment 1b		
Shape Redundancy	79.5 (15.7)	90.6 (12.8)
No Redundancy	65.8 (15.7)	85.8 (11.2)
Experiment 1c		
Joint Color + Shape Redundancy	84.0 (9.4)	92.54 (9.8)
No Redundancy	71.1 (12.7)	85.72 (15.3)
Experiment 2a		
Joint Color + Shape Redundancy	94.0 (6.6)	83.0 (15.2)
Split Redundancy	91.6 (8.4)	76.2 (10.8)
No Redundancy	87.2 (11.1)	72.3 (19.0)
Experiment 2b		
Split Reundancy	73.6 (12.3)	88.0 (10.9)
Color Redundancy	73.4 (13.0)	86.7 (11.8)
No Redundancy	67.8 (14.1)	85.7 (11.1)

M = mean; SD = standard deviation

Experiment 1b

Methods

Participants

We collected data from 24 new participants who did not participate in any of the other experiments. Four additional participants who failed to comply with instructions were replaced during the data collection. After exclusion of one participant performing at chance level and an additional participant with a dual task performance below 80% correct, the final samples consisted of 22 students (13 female, 18-21 years).

Apparatus, Stimuli, and Procedure

All apparatus, stimuli, and procedures were identical to Experiment 1a except the redundant feature information. In this experiment, all objects had the same color (drawn randomly on a trial-to-trial basis) but varied in their shape (see Figure 3b). In half of all trials, one of the shapes was replaced by one the other three shapes in the display, creating redundant shape information.

Results

Redundancy-boost hypothesis. As in Experiment 1a, we calculated the sensitivity value d' as well as the response bias c from signal detection theory for trials with and without redundant objects. The presence of redundant feature information improved the sensitivity for detecting swaps between the two views, $t(21) = 7.37$, $p < .001$, $d_z = 1.57$, 95%-CI [1.14, 2.45] (see Figure 4, middle panel; see also Table 1 for the corresponding accuracy values). There were no differences in the response criterion c between trials with and without redundant shape information, $t(21) = 1.90$, $p = .07$, $d_z = 0.40$, 95%-CI [0.03, 0.82].

Spill-over hypothesis. We analyzed hits for different types of changes using logit mixed effect models. Replicating Experiment 1a, swaps that included one of the redundant objects were detected more often than swaps that included the two unique objects in displays with redundant objects, $\chi^2(1) = 90.93$, $p < .001$, $R^2_m = .09$, 95%-CI [.05, 13]. When compared with swaps in the condition without any redundant objects (all-unique), swaps including the redundant objects were also detected more often, $\chi^2(1) = 114.06$, $p < .001$, $R^2_m = .08$, 95%-CI [.05, .11]. Swaps between the two unique objects within the redundant display were detected less often than swaps in the baseline condition without redundant objects (all-unique), $\chi^2(1) = 5.52$, $p = .019$, $R^2_m = .004$, 95%-CI [<.001, .02]. Note that this difference is opposite to the direction that would be predicted by the spill-over hypothesis (see Figure 5, middle panel).

Experiment 1c

Methods

Participants

We collected data from 24 new participants who did not participate in any of the other experiments. After exclusion of one participant performing at chance level, the final samples consisted of 23 students (10 female, 18-22 years).

Apparatus, Stimuli, and Procedure

All apparatus, stimuli, and procedures were identical to Experiments 1a and 1b except the redundant feature information. In this experiment, all Objects had a unique color with a unique shape (drawn randomly on a trial-to-trial basis; see Figure 3c). In half of all trials, one of the objects was replaced by a second instance of one of the other three objects in the display, creating redundant combinations of color and shape information.

Results

Redundancy-boost hypothesis. As in Experiments 1a and 1b, we calculated the sensitivity value d' as well as the response bias c from signal detection theory for trials with and without redundant objects. The presence of redundant color and shape information improved the sensitivity for detecting swaps between the two views, $t(22) = 6.75$, $p < .001$, $d_z = 1.41$, 95%-CI [1.12, 1.95][5] (see Figure 4, lower panel; see also Table 1 for the corresponding accuracy values). There were no differences in the response criterion c between trials with and without redundant color and shape information, $t(22) = 0.96$, $p = 0.35$, $d_z = 0.20$, 95%-CI [-0.21, 0.70].

Spill-over hypothesis. We analyzed hits for different types of changes using logit mixed effect models. Replicating Experiments 1a and 1b, swaps that included one of the redundant objects were detected more often than swaps that included the two unique objects in displays with redundant objects, $\chi^2(1) = 220.96$, $p < .001$, $R^2_m = .23$, 95%-CI [.01, .49]. When compared with swaps in the condition without any redundant objects (all-unique), swaps including the redundant objects were also detected more often, $\chi^2(1) = 163.91$, $p < .001$, $R^2_m = .16$, 95%-CI [.10, .24].[6] Nevertheless, swaps between the two unique objects within the redundant display were detected less often than swaps in the baseline condition without redundant objects (all-unique), $\chi^2(1) = 44.65$, $p < .001$, $R^2_m = .03$, 95%-CI [.01, .07] (i.e. the opposite from what would be predicted by the spill-over hypothesis; see Figure 5, right panel).

[5] Exploratory screening for effects of sex revealed a significant interaction here. Numerically, this interaction arises from females being more sensitive than males in the condition without redundancy whereas there are no differences in the condition with redundancy. Critically, the beneficial effect of redundancy was present for females, $t(9) = 3.58$, $p = .006$, $d_z = 1.13$, 95%-CI [0.83, 1.92], as well as males, $t(12) = 6.60$, $p < .001$, $d_z = 1.83$, 95%-CI [1.46, 2.91].

[6] Exploratory screening for effects of sex revealed a significant interaction here. Numerically, this interaction arises from females revealing more hits than males for swaps in the baseline, but less hits when the swaps involved the redundant objects. Critically, however, the difference between both conditions is significant for the subgroup of females $\chi^2(1) = 43.35$, $p < .001$ $R^2_m = .09$, 95%-CI [.03, .23] as well as males, $\chi^2(1) = 118.85$, $p < .001$ $R^2_m = .22$, 95%-CI [.15, .33].

Intermediate Discussion of Experiments 1a-c

Across Experiments 1a-c, we observed clear evidence in favor of the redundancy-boost hypothesis and clear evidence against the spill-over hypothesis in Experiments 1a-c. Redundant feature information increased the sensitivity for detecting swaps in the conditions including redundant feature information relative to the conditions with unique objects. Within the conditions with redundancies, however, only swaps including one of the redundant objects elicited more accurate swap detections. Swaps among unique objects in displays including redundant objects were detected less accurately than in the baseline condition without the presence of redundancy.

Rationale for Experiments 2a-b

Experiments 1a-c have confirmed the redundancy-boost hypothesis that participants can leverage redundant feature information in a structure change detection task. The next experiments test whether this benefit stems from feature-based grouping among entire objects by testing whether participants can take advantage of multiple redundant features (color and shape) that are split across groups of objects. There are two key predictions from work in feature-based perceptual grouping that we test in the remaining experiments. The first prediction derived from Nothelfer et al. (2017) is that feature-based grouping should be more pronounced for objects that share all features (e.g., color and shape) than objects that share only a subset of features (e.g., same color but distinct shapes). The second prediction is that people group objects of similar color and/or shape by jointly attending to objects of the same color and/or shape simultaneously, such that the feeling of objects belonging together stems from the fact that they are attended together (Huang, 2019, Huang, & Pashler, 2007; Huang, Treisman, & Pashler, 2007; Yu, Tam & Franconeri, 2019; Yu, Xiao, Bemis, & Franconeri, 2019).

But this means that only one group can be created at a time. If a viewer sees a green square, green triangle, and a red triangle, they cannot create both the green groups and the triangle groups at once. They must choose to group according to one feature or the other. If so, leveraging redundant feature information for structure change detection should also be limited to a single feature-based group. In Experiment 2a, we compare structure change detection performance for two redundant features which are either bound to the same objects or split between separate sets of objects. If the feature-based grouping hypothesis is correct, structure change detection performance should be more accurate when the redundant features are combined within the same objects than when they are split between objects (see Nothelfer et al., 2017). In Experiment 2b, we test an additional prediction by comparing structure change detection performance for the condition with two redundant features that are split between separate subsets of objects relative to a condition in which only a single redundant feature is present. If the feature-based grouping hypothesis is correct, the benefit arising from redundant features that are split between separate groups of objects should not exceed those arising from a single redundant feature because grouping is limited to a single group of objects.

Experiment 2a

Methods

Participants

We collected data from 24 new participants who did not participate in any of the other experiments. After the exclusion of one participant who performed below 80% correct in the dual task, the final samples consisted of 23 students (13 female, 18-21 years).

Apparatus, Stimuli, and Procedure

All apparatus, stimuli, and procedures were identical to Experiments 1a-c with the following exceptions. We compared three conditions. We repeated the redundancy condition of Experiment 1c in which redundant objects combined the same color as well as shape information; and the no redundancy condition. Additionally, we introduced a new split redundancy condition in which there were also two redundant features, but these features were distributed between three or all four objects (e.g., a green circle, a green triangle, and a yellow triangle; see Fig. 3).

Results

We conducted repeated measures ANOVAs with the redundancy condition as independent variable and the sensitivity d' as well as the response bias c as dependent variables in addition to planned subsequent post-hoc comparisons.

We observed that the redundancy condition altered the sensitivity d' for swaps between the two views, $F(2, 44) = 15.69$, $p < .001$, $\eta_G^2 = .15$, 95%-CI [.07, .30][7] (see also Fig. 6, upper panel). The post-hoc comparisons revealed that sensitivity on the condition with two joint redundant features (i.e. on the same objects) was higher than in the condition with two redundant features distributed across more objects, $t(22) = 4.02$, $p < .001$, $d_z = 0.84$, 95%-CI [0.45, 1.45], as well as in the condition without redundancy, $t(22) = 5.01$, $p < .001$, $d_z = 1.04$, 95%-CI [0.70, 1.60]. The comparison of the condition with the distributed redundant features and the baseline without redundancy trended toward significance, $t(22) = 2.07$, $p = .05$, $d_z = 0.43$, 95%-CI [0.5, 0.92]. In contrast, the redundancy condition (combined: $M = 0.27$, $SD = 0.38$; distributed: $M = 0.39$, $SD = 0.27$; no: $M = 0.28$, $SD = 0.40$) had no effect on the response bias c, $F(2, 44) = 1.62$, $p = .20$, $\eta_G^2 = .02$, 95%-CI [.002, .14]. This finding signals the redundancy-boost in structure change detection is most pronounced when objects share all color as well as shape information facilitating feature-based grouping (please note that in Experiments 1a and 1b all features were redundant as one was constant across all

[7] We calculated η_G^2 (R-package "ez", Lawrence, 2016) as effect size to facilitate comparability to studies with different designs (Bakeman, 2005). The confidence intervals were calculated using a bootstrapping procedure with 10,000 iterations.

objects in the display). The results show that it is feature-based grouping rather than the number of redundant features per se which drives the benefit of redundant feature information.

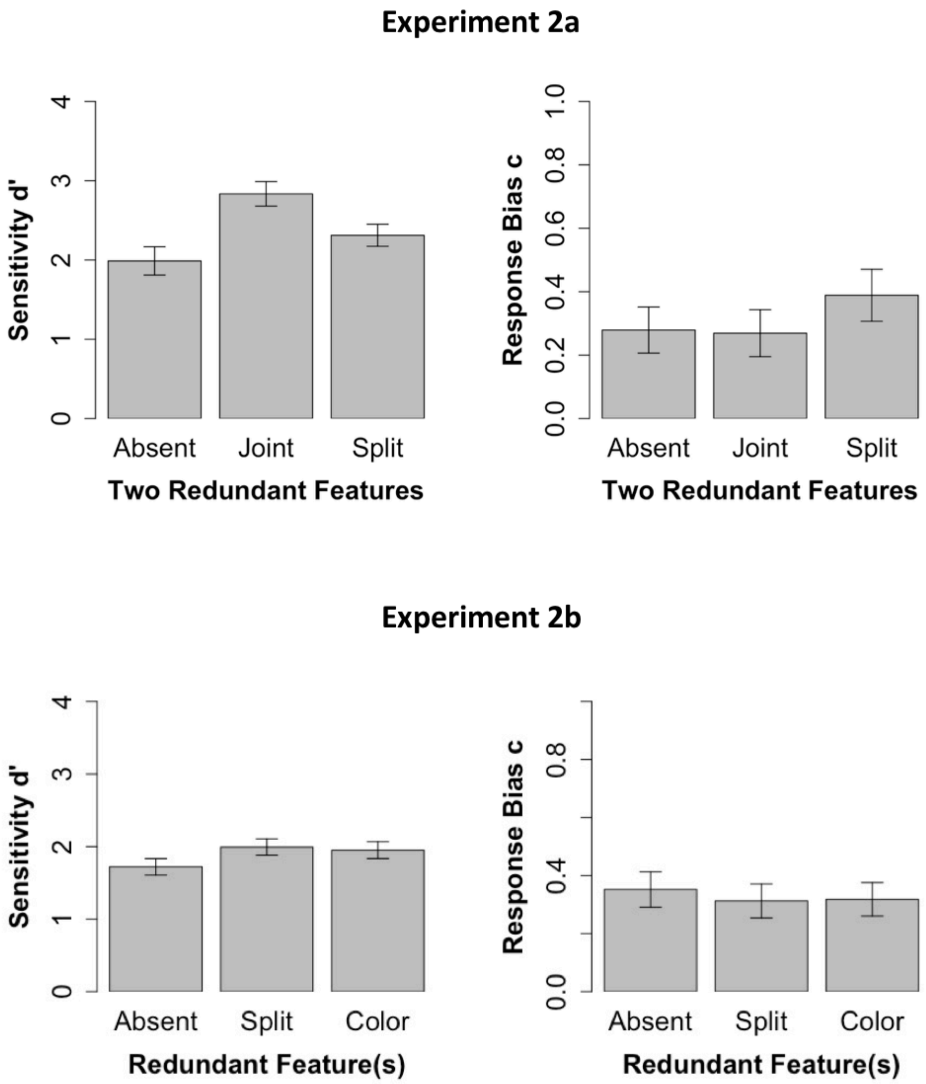

Figure 6. Results of the signal detection analysis of Experiments 2a and 2b. The left column displays sensitivity d'. The right column displays response bias c. The error bars indicate within-subject confidence interval.

Experiment 2b

Methods

Participants

We collected data from 24 new participants who did not participate in any of the other experiments. After exclusion of one participant who performed at chance level in the structure change detection task and another participant who performed below 80% correct in the dual task, the final samples consisted of 22 students (12 female, 18-23 years).

Apparatus, Stimuli, and Procedure

All apparatus, stimuli, and procedures were identical to Experiments 2a with the following exceptions. We compared three conditions: the distributed redundancy condition of Experiment 2a, a color redundancy only condition in which two of four uniquely shaped objects share the same color (e.g. a green circle, a green triangle, a yellow square, and a red star, see Fig. 3), and a no redundancy condition.

Results

We again observed that the redundancy condition altered the sensitivity d’ for swaps between the two views, $F(2, 42) = 3.56$, $p = .04$, $\eta_G^2 = .02$, 95%-CI [.004, .08] (see also Fig. 6, lower panel). The post-hoc comparisons revealed that sensitivity in the condition with two split redundant features was higher than in the condition without redundant features, $t(21) = 2.51$, $p = .020$, $d_z = 0.54$, 95%-CI [0.19, 0.94]. The comparison of the condition with only redundant color information and the condition without redundancy was at the border of significance, $t(21) = 2.08$, $p = .050$, $d_z = 0.44$, 95%-CI [0.04, 0.92]. Importantly, the condition with two distributed redundant features and only color redundancy did not differ from each other, $t(21) = 0.39$, $p = .70$, $d_z = 0.08$, 95%-CI [- 0.34, 0.56]. As in all previous experiments, the redundancy condition (Split: $M = 0.31$, $SD = 0.22$; Color only: $M = 0.32$, $SD = 0.31$; no:

M = 0.35, *SD* =0.30) had no effect on the response bias c, $F(2, 42) = 0.28$, $p = .76$, η_G^2 = .004, 95%-CI [<.001, .07]. This finding shows a strong limit in the number of effective redundant features. In line with the feature-based-grouping hypothesis, our participants were unable to take advantage of more than one redundant feature when they are not part of the same object. Splitting multiple redundant features across more than a single pair of objects did not improve performance beyond a single redundant feature.

Discussion of Experiments 2a-b

Across Experiments 2a-b, we observed further evidence in favor of the redundancy-boost hypothesis, as redundant feature information again improved the identification of swaps between views. In line with the feature-based-grouping hypothesis, we observed that this benefit was more pronounced when two redundant features were combined in the same objects than when they were split between more objects. In fact, in the case of split redundant features, performance does not improve beyond performance with a single redundant feature, suggesting that the benefit of redundancy is limited to a single redundant feature, or a single group of redundant objects.

General Discussion

The current series of experiments explored how redundant feature information might be leveraged to improve structure change detection performance. In line with the redundancy-boost hypothesis, we observed that the presence of redundant feature information improved the ability to detect swapped objects between two views separated by a 90 degree rotation. The presence of redundant feature information also apparently encouraged participants to preferentially encode those objects, as they more reliably detected changes that involved those objects, relative to objects with no redundancy. However, contradicting the spill-over hypothesis, this benefit came at the cost of the remaining objects in the display. When redundant feature information was present, swaps between objects not carrying redundant features were detected less often than in the conditions without redundant feature information. Finally, in line with the feature-based-grouping hypothesis, we observed strict limitations with regard to leveraging multiple redundant features. When two redundant features were split across more than two objects, performance was less accurate than when the two redundant features formed a pair of fully redundant objects. Indeed, splitting redundant features did not improve performance beyond the level of a single redundant feature.

Feature-based grouping as the mechanism of redundancy benefits

Although the capacity in maintaining feature-location bindings across rotated views is limited to a single object (Xu & Franconeri, 2015), leveraging redundant feature information allows observers to perform beyond this limitation, by maintaining that binding across an entire homogenous group. Such a mechanism would allow observers to encode more objects within the same restricted capacity (see also Brady & Tenenbaum, 2013). Feature-based grouping appears likely to underlie these redundancy benefits. In particular, this idea is in line with the results of Experiments 2a and 2b, which have revealed limitations in leveraging redundant visual features. The beneficial effects of the presence of two redundant features was remarkably reduced when different combinations of redundant features were shared by distinct groups of objects. In these cases, a pair of objects with redundant features shares only one of the two features, but not the other (e.g., a red square and a red triangle).

These limitations are congruent with a recent account from the perceptual grouping literature, where visual similarity grouping (e.g., color or shape) is limited to a single group at a time (Huang, 2019, Huang, & Pashler, 2007; Huang, Treisman, & Pashler, 2007; Yu, Tam & Franconeri, 2019; Yu, Xiao, et al, 2019). If this strict limitation transfers the visual compression that aids structure change detection across rotated views, this would explain why mental zip files cannot leverage across redundancies that is split across multiple objects, as in Experiments 2a and 2b.

Using leveraged feature information

Following the initial feature-based grouping, there appear to be two plausible mechanisms that might explain how observers can preserve redundant information across the display rotation: perceptual averaging and multiple object tracking. Research on the perceptual averaging of to-be-grouped objects (Ariely, 2001; Alvarez, 2011; Haberman, Brady, & Alvarez, 2015) has shown that observers represent mean values of object-based features which could change dynamically over time (Albrecht & Scholl, 2010). This includes the average location of two or more objects (Alvarez & Oliva, 2008). It is possible that our participants maintain the binding of the average location with the color binding across the rotation or the display. In this case, if the transformed average locations do not match with the perceived average location in the final display, the participants are able to identify the trial as a swap trial.

In contrast, multiple object tracking does not require assumptions about averaging. Instead, this account states that participants track the spatial locations of simultaneously selected objects (Franconeri, Alvarez, & Enns, 2007) in parallel (see Meyerhoff, Papenmeier, & Huff, 2017). Critically, these tracked objects could be maintained across brief intervals of object invisibility (Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2005) including rotations (Meyerhoff, Huff, Papenmeier, Jahn, & Schwan, 2011). In agreement with our observation that benefits from feature redundancies are limited to one group of objects, temporarily invisible features information also cannot be maintained during object tracking (Pylyshyn, 2004). Similar to maintaining feature-location bindings across views, however, the participants in our experiments may have been able to maintain only a single feature (i.e., the redundant one), but were able to point it to multiple object locations (see Huang et al., 2007).

In this case, a mismatch between the maintained color and actual color at the tracked locations in the final image would signal that a swap had occurred between the two views.

The current set of experiments is compatible with both explanations; however, future research might aim at disentangling them by addressing the question of whether observers maintain information about distinct objects or only the average location across the display rotation. Critically, however, and independent of the exact mechanism, we propose that an initial stage of feature-based grouping strongly limits any processes that arise subsequently.

Why is there no spill-over effect?

A difference between our experiments and related experiments on visuo-spatial memory for color-location bindings is that the increased sensitivity for changes involving redundant objects came at the expense of change sensitivity for non-redundant objects in the same display. Whereas performance for such unique objects among redundant objects also improved (e.g., Lin & Luck, 2008; Mate & Baques, 2009; Morey et al., 2015), or at least was unaffected (e.g., Quinlan & Cohen, 2012), detecting swaps among non-redundant objects was clearly impaired in our study. A plausible explanation for this difference could be provided by the different capacity limitations of the different tasks at hand. Whereas a capacity limitation of approximately four objects would free space for an additional object when compressing two redundant objects to the size of one, no comparable free space would emerge if the initial capacity limitation encompasses only one object, such as in tasks involving rotated views (Xu & Franconeri, 2015). As attentional disruptions – including internal spatial manipulations (Engle et al., 1999) - interfere with maintaining feature bindings (Fougnie & Marois, 2009; Wheeler & Treisman, 2002), any potential spill-over from the memory representation is obsolete after the display rotations in our study. We therefore argue that only passive memory tasks are capable of inducing spill-over effects, whereas tasks that require a mental manipulation of feature-location bindings cannot produce spill-over effects, but instead result in an attentional bias toward redundant objects. In return, this bias toward redundant objects lowers performance for the remaining unique objects in the display relative to conditions without redundant objects, for which unique objects are encoded more or less at chance level.

Implications for application and training

As trainings of domain-general spatial skills such as mental rotation have been largely ineffective in improving STEM outcomes (Stieff & Uttal, 2015), one aim of the present study was to identify other domain-general mechanisms that could be candidates for training. Our experiments show that visual features such as color and shape could be used to compress elements of a spatial layout in order to use strictly limited capacities more efficiently (i.e. one instead of two elements). Thus, explicitly encouraging observers to take advantage of redundancies could be an important step toward improving spatial training. However, this needs to be established in future research which should focus on the question of whether individual differences in visual compression actually predict structure change detection performance for in-context stimuli such as molecules. It will also be important to test whether it is possible to leverage grouping cues beyond color and shape (Kubovy & Van den Berg, 2008).

Conclusion

Redundant visual features improve structure change detection performance (e.g. identifying swaps between objects) across display rotations; however, this improvement is restricted to a single group of objects, rendering mental zip files as less flexible than their computer-based counterparts. Instead, this trick requires leveraging feature-based grouping mechanisms to reduce demands on working memory limitations. Leveraging redundant features therefore potentially could be a domain-general strategy to improve performance, for training of spatial abilities in STEM domains.

Acknowledgment

The experimental scripts (requires PsychoPy), data as well as the scripts for the analysis (requires R) are available at https://osf.io/n82hj/ (Meyerhoff, Jardine, Stieff, Hegarty, & Franconeri, 2020). This work was supported by NSF awards DRL-1661264, DRL-1661096 and DRL-1661151. We would like to thank Sneha Pamulapati and Jason Liou for their help with the data collection.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21,* 560-567.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15,* 122-131.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological science*, *19,* 392-398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12,* 157-162.

Baddeley, A. (1986). *Working memory*. New York, NY, US: Oxford University Press.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*, 379-384

Barton, K. (2020). MuMIn: Multi-model inference. R-package, https://cran.r-project.org/web/packages/MuMIn/

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H. ... Fox, J. (2020). lme4: Linear Mixed-Effects Models using 'Eigen' and S4. R-package, https://CRAN.R-project.org/package=lme4

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 12 - 23.

Botella, J., Peña, D., Contreras, M. J., Shih, P. C., & Santacreu, J. (2009). Performance as a function of ability, resources invested, and strategy used. *The Journal of General Psychology, 136*, 41-70.

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

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120,* 85-109.

Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15,* 2-11.

Cheung, O. S., Hayward, W. G., & Gauthier, I. (2009). Dissociating the effects of angular disparity and image similarity in mental rotation and object recognition. *Cognition*, *113,* 128-133.

Cooper, L. A., & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 503-514.

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

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.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, *128,* 309-331.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191.

Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 395-404.

Fougnie, D., & Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Visual Cognition*, *17,* 48-66.

Fox, J., Weisberg, S., Price, B, Adler, D., Bates, D., Baud-Bovy, G, … & R core team (2020). car: Companion to applied regression. R package, https://CRAN.R-project.org/package=car

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1003-1012.

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron, 34*, 161-171.

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General, 144,* 432-446.

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science, 11*, 7-12.

Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects?. *Psychonomic Bulletin & Review*, *13,* 516-523.

Huang, L. (2019). Unit of visual working memory: A Boolean map provides a better account than an object does. *Journal of Experimental Psychology: General.*

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114,* 599-631.

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, *317,* 823-825.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581-592.

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution, 5*, 944-946.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*, 289-303.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92,* 137.

Kellman, P. J. (2013). Adaptive and perceptual learning technologies in medical education and training. *Military Medicine*, *178,* 98-106.

Khooshabeh, P., Hegarty, M., & Shipley, T. F. (2013). Individual differences in mental rotation. *Experimental Psychology, 60,* 164-171.

Krummenacher, J., Müller, H. J., & Heller, D. (2001). Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Perception & Psychophysics, 63*, 901-917.

Kubovy, M., & Van Den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, *115,* 131-154.

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R-package, https://CRAN.R-project.org/package=ez

Lin, P. H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition*, *17,* 356-372.

Mate, J., & Baqués, J. (2009). Short article: Visual similarity at encoding and retrieval in an item recognition task. *Quarterly Journal of Experimental Psychology*, *62,* 1277-1284.

Meyerhoff, H.S., Huff, M., Papenmeier, F., Jahn, G., & Schwan, S. (2011). Continuous visual cues trigger automatic spatial target updating in dynamic scenes. *Cognition, 121,* 73-82.

Meyerhoff, H. S., Jardine, N., Stieff, M., Hegarty, M., & Franconeri, S. (2020). Raw data, scripts, and code for "Visual ZIP files: Viewers beat capacity limits by compressing redundant features across objects". https://doi.org/10.17605/OSF.IO/N82HJ

Meyerhoff, H.S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: a tutorial review. *Attention, Perception, & Psychophysics, 79,* 1255-1274.

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences*, *26,* 141-152.

Morey, C. C. (2019). Perceptual grouping boosts visual working memory capacity and reduces effort during retention. *British Journal of Psychology*, *110,* 306-327.

Morey, C. C., Cong, Y., Zheng, Y., Price, M., & Morey, R. D. (2015). The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory. *Archives of Scientific Psychology*, *3,* 18.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133-142.

National Research Council. (2006). Learning to think spatially. Washington, DC: The National Academies Press.

Nazareth, A., Killick, R., Dick, A. S., & Pruden, S. M. (2019). Strategy selection versus flexibility: Using eye-trackers to investigate strategy use during mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 232-245.

Newcombe, N. S. (2016). Thinking spatially in the science classroom. *Current Opinion in Behavioral Sciences*, *10,* 1-6.

Nothelfer, C., Gleicher, M., & Franconeri, S. (2017). Redundant encoding strengthens segmentation and grouping in visual displays of data. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1667-1676.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162,* 8-13.

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Reviews of Psychology, 58,* 75-96.

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review, 20,* 1282-1289.

Pylyshyn, Z. W. (1979). The rate of "mental rotation" of images: A test of a holistic analogue hypothesis. *Memory & Cognition*, *7,* 19-28.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11,* 801-822.

Quinlan, P. T., & Cohen, D. J. (2012). Grouping and binding in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38,* 1432-1438.

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, *3*, 6-21.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125,* 4-27.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171,* 701-703.

Sorby, S. A. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education*, *31,* 459-480.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17,* 219-234.

Stieff, M., & Uttal, D. (2015). How much can spatial training improve STEM achievement? *Educational Psychology Review*, *27,* 607-615.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*, 233-282.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, *1*, 275-277.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45,* 37-55.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin, 139*, 352-402.

Voyer, D., Saint-Aubin, J., Altman, K., & Doyle, R. A. (in press). Sex differences in tests of mental rotation: Direct manipulation of strategies with eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250-270.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817-835.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131,* 48-64.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419-433.

Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental rotation. *Psychological Science*, *26,* 1241-1251.

Yu, D., Tam, D., & Franconeri, S. L. (2019). Gestalt similarity groupings are not constructed in parallel. *Cognition*, *182,* 8-13.

Yu, D., Xiao, X., Bemis, D. K., & Franconeri, S. L. (2019). Similarity grouping as feature-based selection. *Psychological Science*, *30,* 376-385.

Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics*, *31*, 201–209.