VISUAL COGNITION, 2003, *10* (1), 101–127

Detection of three types of changes to novel objects

Simone K. Keane

Department of Psychology, University of Wollongong, NSW, Australia

William G. Hayward

Department of Psychology, Chinese University of Hong Kong, Shatin, NT, Hong Kong

Darren Burke

Department of Psychology, University of Wollongong, NSW, Australia

In this study a change-detection paradigm is used to explore the nature of the information used to recognize three-dimensional, novel objects. In particular, whether we are sensitive to changes in part identity or configuration information. Experiments 1 and 2 showed that configural changes made to the object parts were significantly easier and quicker to detect than changes made to the shape or arrangement of object parts. Variance due to the total change in pixels did not predict performance in either of the experiments. The results of Experiment 3 showed the same pattern as Experiments 1 and 2, even though the objects used were altered to make part identity information more salient. Experiment 4 demonstrated that the physical size of the changes is not the crucial variable for this pattern of results. These findings are discussed in relation to the nature of visual representations and theories of object recognition.

The phenomenon in which we are unable to detect changes to details of a scene from one view to the next is called change blindness (see Simons & Levin, 1997, for a review). Change blindness is studied using a change-detection paradigm, where the aim is to identify the difference(s) between two images. The rationale of this methodology is that the types of changes detected reflect the information that is effectively utilized by the visual system. Note that the change-detection task is the opposite of the identity-detection task often employed in object recognition studies, where the objective is to decide whether two images are the same.

Please address all correspondence to: Department of Psychology, University of Wollongong, Northfields Ave, Wollongong 2522, NSW, Australia. Email: skk04@uow.edu.au

The authors wish to thank Steve Roodenrys and Joseph Ciarrochi for assistance in statistical analyses.

Darren Burke is now at the Department of Psychology, Macquarie University, NSW, Australia.

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280143000014

Change blindness has been found to occur in different media using a number of techniques. Studies have shown that detection of a variety of changes to photographs of natural scenes (Grimes, 1996), and to the visual form of objects (Henderson, 1997), was difficult when those changes were made during a saccade. Using a ''flicker'' paradigm, in which two alternating views of a scene are separated by a brief retention interval, Rensink, O'Regan, and Clark (1997) found that changes to the scene were rarely detected during the first cycle and that some changes took nearly a minute to detect. Levin and Simons (1997) demonstrated change blindness across ''cuts'' in motion pictures and, further, Simons and Levin (1998) found that change blindness occurred in real-world occlusion situations.

Results of recent scene perception and change blindness research have been used to argue that little visual information survives from one view to the next (Aginsky & Tarr, 2000; McConkie & Currie, 1996; Rensink et al., 1997; Simons, 1996). Simons and Levin (1997) suggest that detailed representations are not a necessary feature of the human visual system. They argue that we abstract the meaning or gist (and possibly the spatial configuration and movement direction) of a scene from each visually rich fixation. If the gist is the same at each fixation, our perceptual system assumes that the details are the same. Thus, changes to visual details that do not violate the gist or spatial configuration of a scene go unnoticed.

Simons (1996) investigated the role of spatial configuration in change blindness. The study examined changes in object identity and location in arrays. If visual representations capture detailed information about the properties and arrangements of objects, then subjects should be sensitive to such changes. Subjects were shown two object arrays, separated by a blank white screen. The arrays were either identical or differed in one of three ways: (1) The identity of an object, where one object in the array was replaced with a new object; (2) a switching of objects, where two objects in the array switched positions; and (3) the configuration of the array itself, where one of the objects was moved to a previously unoccupied location, thus changing the spatial configuration of the array. Five experiments were conducted using arrays of photographs of common objects and of black-filled two-dimensional novel shapes.

The results of Simons' (1996) experiments showed that subjects had difficulty in detecting the switch and identity changes, but memory for the spatial configuration of objects in the arrays remained nearly perfect across all experimental manipulations. Simons contends that these results suggest a fundamental difference between representations of spatial configuration and object properties. Specifically, Simons argues that because detection of changes to spatial layout was accurate, fast, and not influenced by verbal interference, this information is probably represented visually and does not require complex processing. Object properties, however, may be assumed by the perceptual system to remain stable over time, allowing the world itself to serve as a memory store to be accessed when required (O'Regan, 1992). Thus, detection of changes to object properties requires that an effortful abstraction of detail is made.

Change blindness studies have focused on different types of changes for scenes (Hollingworth & Henderson, 1997; Rensink et al., 1997) and for arrays of objects, both common and novel (Simons, 1996); however, few studies have investigated changes made to displays of single, central objects. Williams and Simons (2000) have looked at the effect of the magnitude of changes made to single, three-dimensional objects and found that smaller changes (in terms of the number of object parts changed) were more difficult to detect than larger changes. Specifically, one-part changes were more difficult to detect than two-part changes, which in turn were more difficult to detect than three-part changes. That is, in terms of the magnitude of change, change blindness is not an ''all-or-nothing'' effect. Similarly, we can look at different *types* of changes made to single objects. Perhaps different types of changes vary in their ability to be detected. For example, based on the results of Simons (1996), one might expect part configuration and part shapes to be encoded differently in object representations. Such a finding would hold implications for models of object recognition.

One general approach to object recognition is to represent objects in terms of their structural components. For example, the Geon-Structural-Descriptions (GSD) approach (Biederman, 1987; Hummel & Biederman, 1992), proposes that our visual system represents an object on the basis of a restricted class of volumetric primitives known as ''geons''. The object representation is called a ''structural description'', and consists of the geons that denote each part of the object, as well as the qualitative spatial relations between geons. Thus, for any given object, the same structural description is derived, regardless of viewpoint, as long as the same general configuration of features is in the image.

There has been much evidence supporting the role of parts in human shape perception (e.g., Biederman & Cooper, 1991; Biederman & Gerhardstein, 1993, Hoffman & Richards, 1984); however, comparatively little is known about how we perceive the spatial relations among an object's features or parts. Hummel and Stankiewicz (1996) examined the types of spatial relations used in representing objects and found that the configuration of an object's parts was perceived in terms of categorical, rather than coordinate, relations. Saiki and Hummel (1998) found that relations among parts that are connected (i.e., parts of one object) are processed differently to relations among parts that are not connected. That is, our visual system appears particularly sensitive to the spatial relations among object parts. However, neither of the studies directly compares perception of object parts with perception of spatial relations.

One study that has investigated perceptual relations between configural and component properties was conducted by Kimchi and Bloch (1998). They found that performance in discrimination and classification tasks was dominated by the configural properties of the objects (novel line drawings), regardless of the discriminability of the component properties. Component properties can be used to make these judgements; however, this incurs significant costs in time. Kimchi and Bloch use these results to argue that the human perceptual system may be more sensitive to configural rather than component properties and that therefore configural properties may be available earlier than component properties.

Cave and Kosslyn (1993) conducted experiments investigating the role of parts and spatial relations in object identification. Using line drawings of everyday objects, the spatial relations and segmentation of object parts were manipulated. They found that the way an object is sectioned into parts affects object identification only under impoverished viewing conditions. Further, having the correct spatial relations among object parts was found to be critical for easy identification. Cave and Kosslyn argue that their results suggest that the global shape or overall spatial relations of an object is encoded first, with parts being analysed subsequently. Given the results of Cave and Kosslyn (1993) and Kimchi and Bloch (1998), one might expect that changes made to the spatial configuration of an object would be detected faster and more easily than changes made to the shape of the object parts.

Simons (1996) operationalized the spatial configuration of object arrays as the overall layout of the objects in the array. That is, the spatial configuration is the locations in the array grid that are occupied. Different array configurations would have different grid locations occupied. Importantly, configuration is not dependent on the components in the array. Similarly, here we shall define spatial configuration of three-dimensional objects as locations in space occupied by object parts. Different object configurations would have different locations occupied in the object space. The configural properties of an object depend on the relations between the components not on the components themselves (Kimchi & Bloch, 1998). We are sensitive to spatial configuration changes in scenes and object arrays, are we sensitive to configural changes in objects? Is it the case that we are generally sensitive to spatial configuration information in visual processing? Because of the importance afforded to the spatial configuration or structure in many models of object recognition, we investigated sensitivity to changes to this type of object information and directly compared it to sensitivity of object part information.

The present study looks at different types of changes made to three-dimensional novel objects, including changes to spatial and object properties. Similar to Simons' (1996) experiments, the objects could differ in one of three ways (see Figure 1): (1) Part identity, where one randomly chosen object part was replaced with a new object part; (2) a switching of parts, where two randomly selected object parts switched positions; and (3) spatial configuration, where one randomly selected object part moved to a previously unoccupied position on the object body, changing the spatial configuration of the object. In doing so, we have two fundamental concerns. First, we wish to investigate whether change blindness is caused by the complexity of scene information, or whether it can be

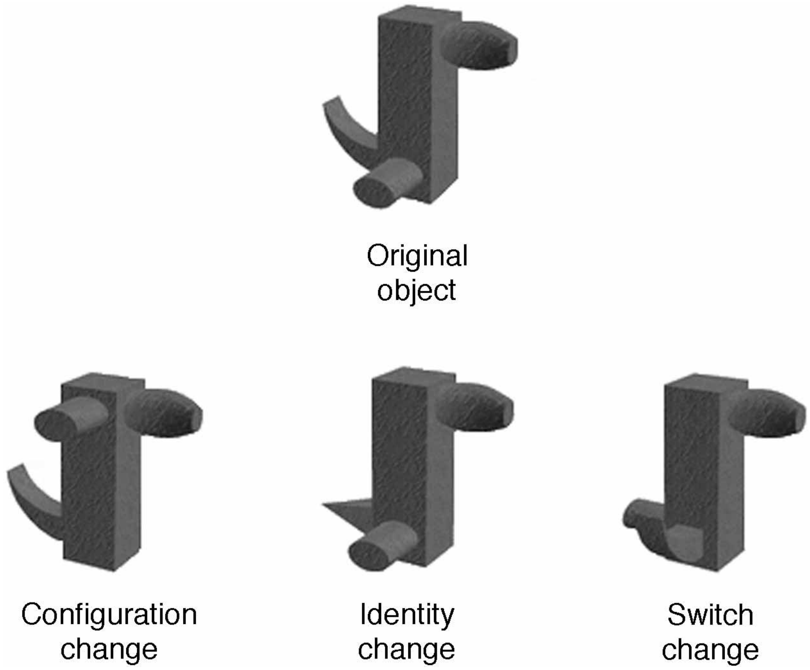

Figure 1. Examples of a configuration change, part identity change, and switch change made to the objects.

found for single objects. Second, we wish to examine the role of parts and spatial relations in object perception.

EXPERIMENT 1

Spatial relations and part identity information are two important components of some object recognition theories (e.g., Biederman, 1987; Hummel & Biederman, 1992). The present set of experiments aims to investigate what role each of these components play. If spatial relations information is as accessible or important as identity information in visual object recognition, then there should be no difference in detection performance between the three change types. However, if, as Simons (1996) found, spatial configuration is used in preference to other types of information available, then detection of spatial configuration changes should be easier and quicker than part arrangement and identity changes.

Method

Subjects. A total of 26 undergraduate students participated and were tested individually. Subjects received course credit for participating.

Materials. Stimuli were rendered images of three-dimensional novel objects. Each object was composed of a main body with three appendage parts. The appendages or ''limbs'' were attached to the body at three of six possible positions (see Figure 1 for an example). There were 10 ''base'' objects, each having a configuration, identity, and switch change made to them, giving a total of 40 different object exemplars used in the current experiment. All objects were photorealistically rendered with the same colour and texture. Objects were shown at the same orientation and magnification. The entire background screen was white.

The experiment was controlled by RSVP software (Williams & Tarr, 1998) on Macintosh computers. The objects were all of similar size, with the average dimensions of each object being 200 pixels wide and 230 pixels high. The mask used in this experiment consisted of elements from a variety of object images.

Procedure. The procedure and timings used were similar to that of Simons (Exp. 1, 1996). The experiment consisted of three blocks of 60 randomly ordered trials in which subjects viewed objects on a computer monitor. Each object was randomly placed at a position 25 pixels in any direction from the centre of the screen. Each trial began with a fixation cross appearing for 500 ms at the centre of the screen, followed by the first object for 2 s, immediately followed by a mask appearing on the screen for 4.3 s, and finally another object which remained on the screen until the subject responded. The next trial began 1000 ms after the subject made a response. The second object was either identical to the first or different in one of three ways: (1) Part identity; (2) a switching of parts; or (3) spatial configuration (see Figure 1). Participants were asked to indicate whether the two objects presented to them were the ''same'' or ''different'' by pressing corresponding keys on a keyboard. Half of the trials were ''same'' trials and half ''different''. The different trials were split equally into the three change conditions.

Results and discussion

Figure 2 shows that participants were most accurate at detecting a spatial configuration change, less accurate at detecting a switch change and least accurate at detecting an identity change. A 4×3 repeated measures ANOVA including the within-subjects factors of change type (none, configuration, identity, or switch) and block (1, 2, or 3) was conducted on accuracy rates and on reaction time (RT).

The ANOVA on accuracy rates showed a significant variation between change type, $F(3, 75) = 23.569$, $p < .01$. Post hoc Scheffé contrasts showed that accuracy for both identity and switch change conditions were significantly worse than for the configuration change condition (both $p < .01$). Further, accuracy for the identity change was significantly worse than for the switch condition ($p < .01$).

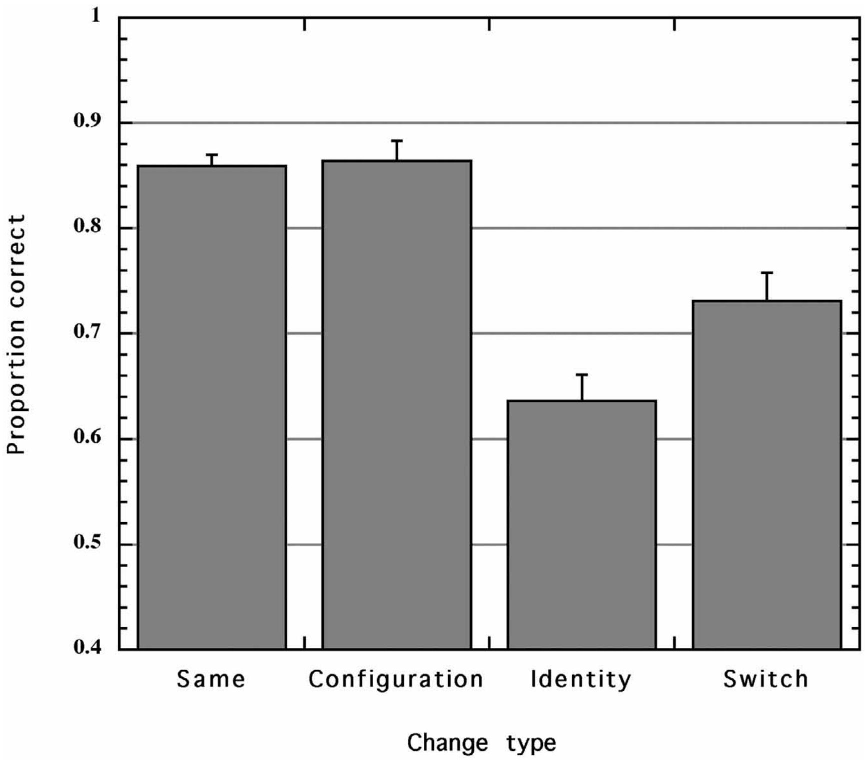

Figure 2. Mean proportion correct in each condition of Experiment 1.

As shown in Figure 3, the RT results follow the accuracy pattern shown previously. Participants were quickest at detecting a spatial configuration change and slower at detecting an identity change and a switch change. The ANOVA on RTs for change type showed a significant variation between conditions, $F(3, 75) = 3.304$, $p = .03$. Post hoc Scheffé contrasts showed that RT for both identity and switch change conditions were significantly slower than for the configuration change condition ($p = .02$ and $p = .01$, respectively). RT for the identity change condition was not significantly different from the switch condition ($p = .85$).

The main effect of block for accuracy was significant, $F(2, 50) = 18.613$, $p < .01$. Not surprisingly, subjects were less accurate in block 1 and became more accurate in blocks 2 and 3. Post hoc Scheffé contrasts show that accuracy in block 1 was significantly worse than accuracy in blocks 2 and 3 (both $p < .01$). There was no significant difference in accuracy between blocks 2 and 3 ($p = .30$). Further, there was a significant interaction between block and change type, $F(6, 150) = 2.658$, $p = .03$. Figure 4 shows that accuracy for the switch condition improved considerably over blocks, whereas the other conditions did not.

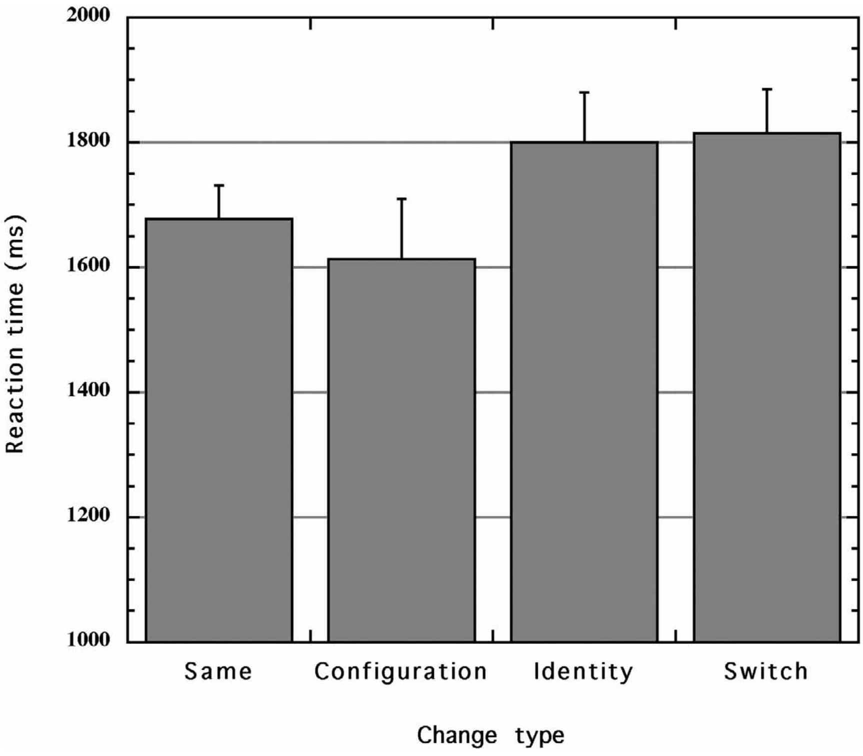

Figure 3. Mean reaction time in each condition of Experiment 1.

The main effect of block for reaction time was significant, $F(2, 50) = 5.77$, $p < .01$; subjects were slowest in block 1 and faster in blocks 2 and 3. Post hoc Scheffé contrasts show that RTs for block 1 were significantly slower than for blocks 2 and 3 ($p = .02$ and $p = .01$, respectively) and there was no significant difference in RT between blocks 2 and 3 ($p = .51$). The interaction between block and change type was not significant $F(6, 150) = 1.037$, $p = .40$, indicating that subjects improved equally in RT across change types.

Further analyses were conducted on individual object changes. Because each of the objects were quite different in shape, perhaps something other than the *type* of change influenced our results. A quantitative measure of the change between two objects is the number of pixels that change from black to white or vice versa in the two silhouetted images (Williams & Simons, 2000). The average percentage of pixels changing in each condition is 5.5% pixels for a configuration change, 4.2% pixels for a switch, and 2.3% pixels for an identity change.

To examine the effect of pixel change on performance, we analysed the main effect of condition after partialling out the variance due to the amount of pixels

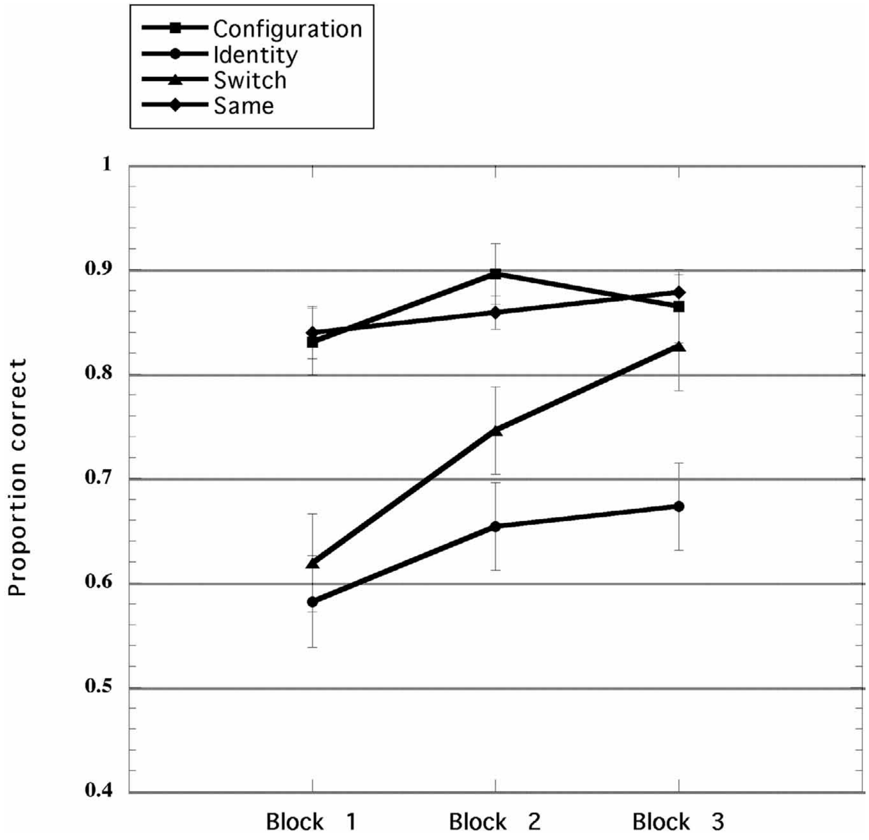

Figure 4. Mean proportion correct in each condition across blocks of Experiment 1.

changing. To do this we conducted an analysis of covariance (ANCOVA) on accuracy, with pixel change as the covariate. This showed a main effect for condition, $F(2, 17) = 4.80$, $p = .02$, indicating that pixel change does not entirely account for the variance between conditions. An analysis of covariance requires that the homogeneity of regression be tested, that is, that the effect of pixel change does not differ significantly across condition. Tests for this were not significant (all $p > .05$).

It is possible that subjects were able to verbally encode some or all of the object parts (for example, ''cone'', ''curved cylinder'', ''rectangular block''). Further, if the subjects were verbally encoding the object parts in a particular order, the increase in accuracy for the switch condition may be explained. As the body of the object always stayed the same for each trial, the subject need only encode two of the object parts to successfully detect a switch, as one of those parts must change in the switch condition. However, to successfully detect an identity change using this strategy, the subject must encode all three of the object parts because only one of them will change. This strategy, however, cannot be used to explain the strong performance in the configuration change condition.

Simons (1996, Exp. 5) used a verbal shadowing task to interfere with the verbal encoding of arrays of photographs of common objects. The pattern of results was similar to those of his previous experiments, that is, spatial changes were easier to detect than identity and switch changes. Simons found that performance for the configuration change condition was unchanged by the verbal shadowing task, whereas performance in the identity and switch conditions was significantly lower than in his Experiment 1 (same experiment without the verbal shadowing task).

EXPERIMENT 2

To investigate whether performance in the switch condition was due to verbal encoding, we conducted a second experiment which was similar to Experiment 1, but used articulatory suppression to prevent subjects from verbalizing the object parts. If, in fact, the arrangement and identity of object parts are represented verbally, differences between the switch and identity change conditions should disappear and overall performance should become worse. Following the results of Simons (1996, Exp. 5), we expected no change for performance in the configuration change condition.

Method

Subjects. A total of 31 undergraduate students participated and were tested individually. Subjects received course credit for participating. None had participated in Experiment 1.

Materials. The materials were the same as for Experiment 1.

Procedure. The procedure for Experiment 2 was the same as for Experiment 1 except that participants were required to perform an articulatory suppression task. Participants were required to say ''1, 2, 3, 4, 5'' out loud and constantly repeat the sequence for the period during which the sequential matching task was being performed. Participants were given a short break at 15 trial intervals.

Results and discussion

A 4×3 repeated measures ANOVA including the within-subjects factors of change type (none, configuration, identity, or switch) and block (1, 2, or 3) was conducted on accuracy rates and on RT. As shown in Figure 5, participants were

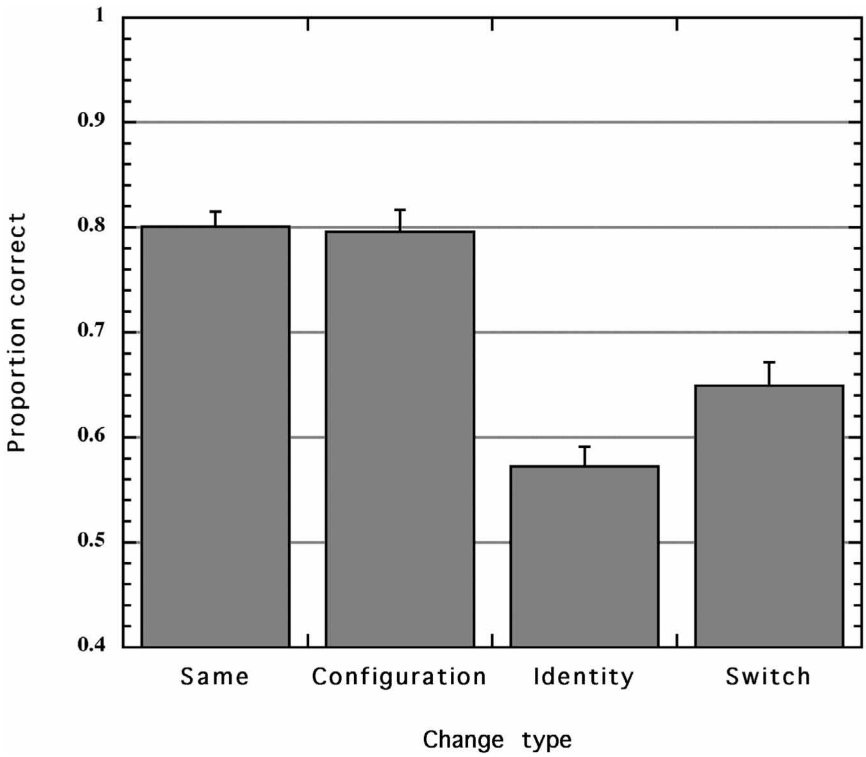

Figure 5. Mean proportion correct in each condition of Experiment 2.

equally accurate at detecting no change as at detecting a spatial configuration change, less accurate at detecting a switch change, and least accurate at detecting an identity change. The ANOVA on accuracy rates for change type showed a significant variation between conditions, $F(3, 90) = 40.971$, $p < .01$. Post hoc Scheffé contrasts showed the same pattern as Experiment 1, that accuracy for all change conditions (configuration, identity, and switch) were significantly different to one another (all $p < .01$).

The ANOVA on RTs for change type showed significant variation between conditions, $F(3, 90) = 5.281$, $p < .01$. As shown in Figure 6, RT for the switch condition was greatest. Post hoc Scheffé contrasts showed that RT for the switch condition was significantly slower than all other conditions (all $p < .05$). Further, there was no significant difference between the remaining conditions (all $p > .16$).

The main effect of block for accuracy was significant, $F(2, 60) = 18.145$, $p < .01$. Post hoc Scheffé contrasts show that accuracy in block 1 was significantly worse than accuracy in blocks 2 and 3 (both $p < .01$). There was no significant difference in accuracy between blocks 2 and 3 ($p = .79$). The

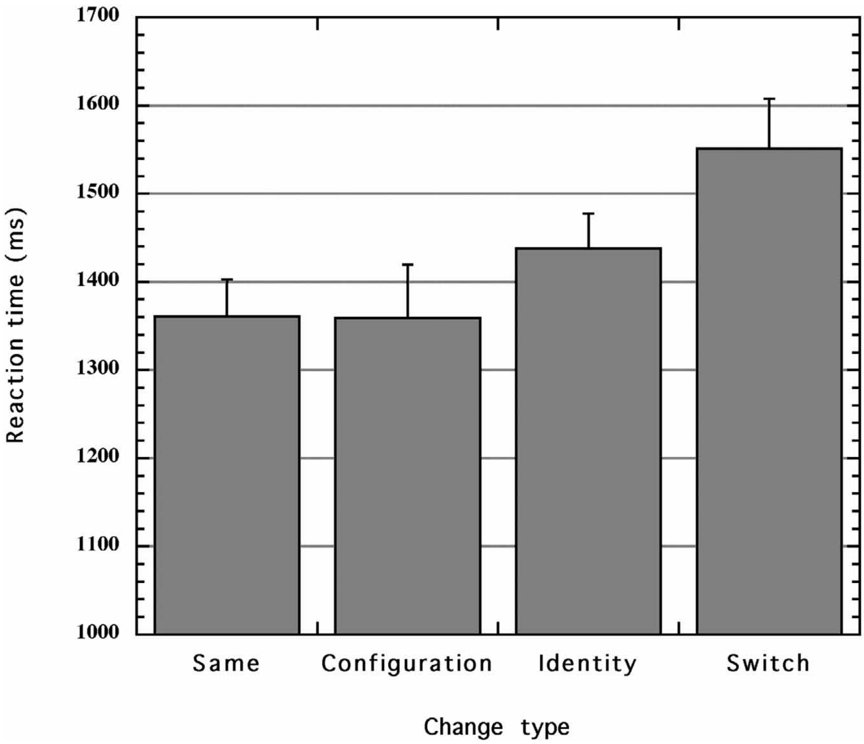

Figure 6. Mean reaction time in each condition of Experiment 2.

interaction between block and change type was approaching significance, $F(6, 180) = 2.071$, $p = .059$. From Figure 7 it can be seen that the pattern of accuracy across blocks seems to improve equally except for the same condition, which remains at a relatively constant level across blocks. The main effect of block for reaction time was significant, $F(2, 60) = 4.794$, $p = .012$. Post hoc Scheffé contrasts show that RTs for blocks 1 and 2 were significantly slower than for block 3 ($p < .01$ and $p = .02$, respectively) and there was no significant difference in RT between blocks 1 and 2 ($p = .54$). The interaction between block and change type for RT was not significant $F(6, 180) = 0.925$, $p = .479$.

Again, further analyses were conducted on individual object changes using pixel change as a quantitative measure of the size of the change between object pairs. An analysis of covariance (ANCOVA) on accuracy, with pixel change as the covariate, was conducted. Again, there was a main effect for condition, $F(2, 17) = 6.91$, $p < .01$, indicating that the main effect of change type is still significant after partialling out the variance due to pixel change. The tests for homogeneity of regression across conditions were not significant (all $p > .05$.

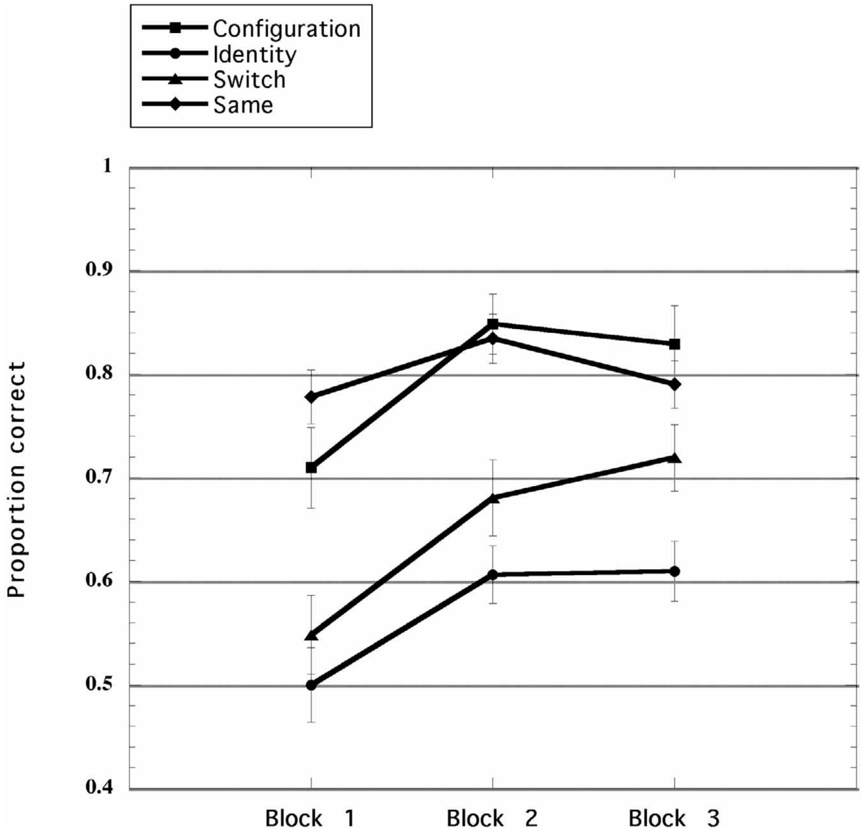

Figure 7. Mean proportion correct in each condition across blocks of Experiment 2.

The ANOVA on accuracy and RT compared performance across change type and blocks in Experiment 1 to Experiment 2. As shown in Table 1, subjects were significantly more accurate in Experiment 1, $F(1, 55) = 5.318$, $p = .02$; however, RT was significantly slower in Experiment 1, $F(1, 55) = 8.624$, $p < .01$. Further, none of the interactions (experiment $\times$ change type, experiment $\times$ block, experiment $\times$ change type $\times$ block) were significant (all $p > .05$), suggesting that the differences between the experiments was simply one of magnitude. That is, articulatory suppression in the second experiment results in an overall decrease in accuracy but also an overall decrease in reaction time. Although the pattern of results remained the same, spatial configuration changes were not found to be as impervious to verbal interference as Simons (1996) found. However, this may be a consequence of the present experiments using novel objects as opposed to photographs of common objects.

TABLE 1
Overall proportion correct and reaction time means for Experiments 1 and 2 (standard errors in parentheses)

	Experiment 1	*Experiment 2*
Proportion correct	0.773 (±0.012)	0.705 (±0.011)
Reaction time (ms)	1726.471 (±38.355)	1427.248 (±25.459)

The pattern of results from the second experiment is similar to that of the first, showing that subjects are more accurate at detecting spatial configuration changes than part identity or arrangement changes. These results are directly comparable to those of Simons' (1996) except for the switch condition. One possible explanation for this deviation is that the nature of the switch change differs somewhat for three-dimensional objects and two-dimensional arrays. A switch change for three-dimensional objects involves changes to the two-dimensional appearance of the parts themselves and perhaps a greater change to the appearance of the object overall. Thus, the switch change used in these experiments may not exactly correspond to Simons' (1996) change.

EXPERIMENT 3

One possible account of the results of the first two experiments is that the configuration changes are larger in some way than the other object property changes. We measured the physical differences between the different objects and found that such differences did not account for performance differences between the conditions. However, pixel change between silhouetted object pairs is a rather imprecise measure of ''information change''. Another method would be to attempt to equate the magnitude of the changes or to at least bias performance against a configural change. One way of increasing the salience of the part identity information would be to increase the size of the ''limb'' or appendage parts (Hoffman & Singh, 1997) and to decrease the height of the body. This should result in identity changes being more obvious due to the larger size of the parts involved in this type of change. In addition, parts would have to move a shorter distance along the body in the configuration change condition which should produce a less obvious change.

If it is the case that configural information is used quickly and more accurately than other kinds of object property information, then the pattern of results for this experiment should be similar to that of the first two experiments. Should this pattern appear even though the salience of part identity information has been increased, it would provide evidence that regardless of the physical size of the object change, configural information is processed more accurately and quickly than part identity information.

Method

Subjects. A total of 26 undergraduate students participated and were tested individually. Subjects received course credit for participating. None had participated in either Experiments 1 or 2.

Materials. The stimuli were based on those used in Experiments 1 and 2. Each object was the same in terms of the types of changes made, but the size of the body and parts was altered. The body of each object was shortened on the vertical axis by 20% and the size of each ''limb'' part was increased in all three dimensions by 30% (for example, see Figure 8).

Procedure. The procedure for Experiment 3 was the same as for Experiment 1.

Results and discussion

A 4 $\times$ 3 repeated measures ANOVA including the within subjects factors of change type (same/no, configuration, identity, or switch) and block (1, 2, or 3) was conducted on accuracy rates and on RT. As shown in Figure 9, participants

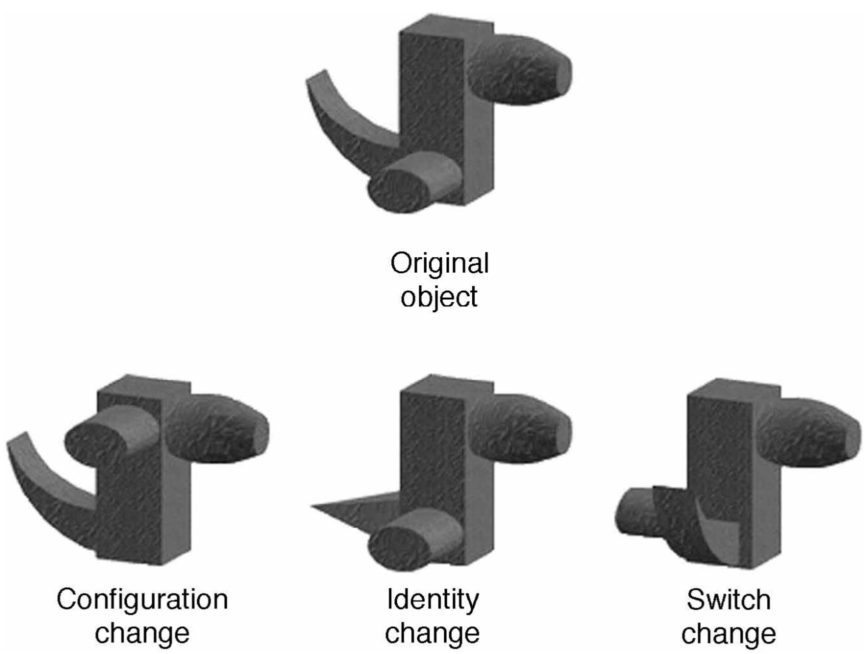

Figure 8. Examples of the changes made to objects used in Experiment 3.

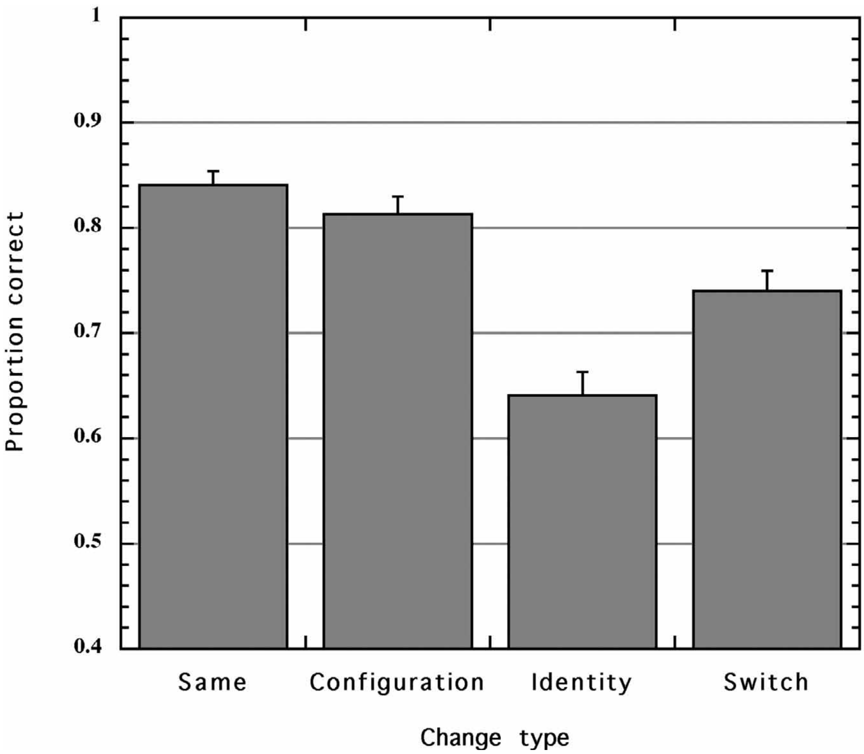

Figure 9. Mean proportion correct in each condition of Experiment 3.

were equally accurate at detecting no change as at detecting a spatial configuration change, less accurate at detecting a switch change and least accurate at detecting an identity change. The ANOVA on accuracy rates for change type showed a significant variation between conditions, $F(3, 75) = 28.449$, $p < .01$. Post hoc Scheffé contrasts showed the same pattern as Experiments 1 and 2, that accuracy for all change conditions (configuration, identity, and switch) were significantly different to one another (all $p < .01$).

The ANOVA on RTs for change type showed significant variation between conditions, $F(3, 75) = 3.953$, $p < .05$. As shown in Figure 10, RT for the configuration condition was smallest. Post hoc Scheffé contrasts showed that RT for the configuration condition was significantly faster than all other conditions (all $p < .05$). Further, there was no significant difference between the remaining conditions (all $p > .5$).

The main effect of block for accuracy was significant, $F(2, 50) = 9.913$, $p < .01$. Again, the same pattern as Experiments 1 and 2 emerged, post hoc Scheffé contrasts show that accuracy in block 1 was significantly worse than accuracy in

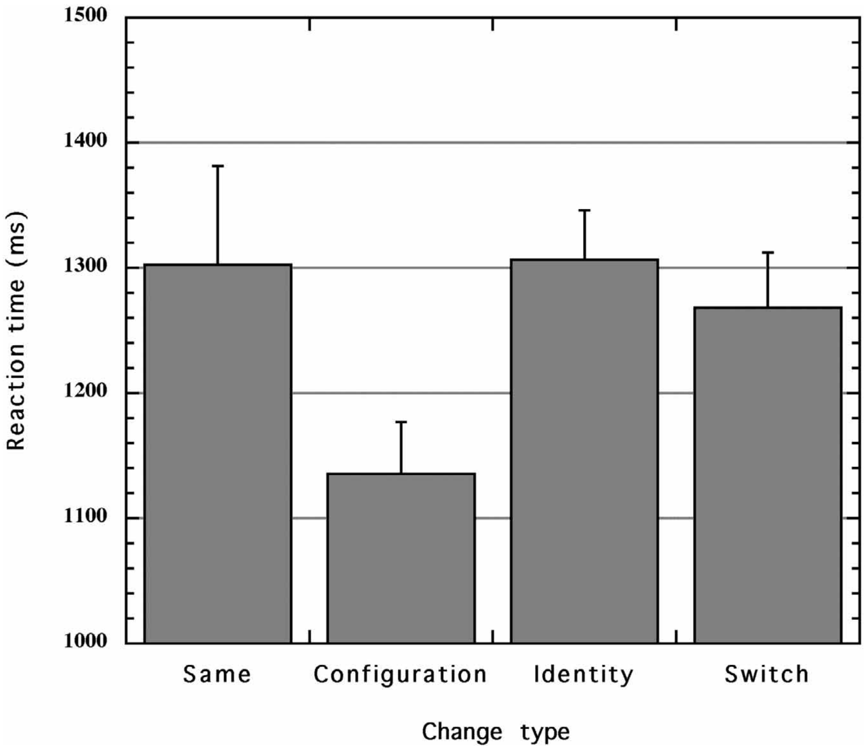

Figure 10. Mean reaction time in each condition of Experiment 3.

blocks 2 and 3 (both $p < .01$). There was no significant difference in accuracy between blocks 2 and 3 ($p = .79$). The interaction between block and change type was significant, $F(6, 150) = 3.027$, $p < .01$. From Figure 11 it can be seen that the pattern of accuracy across blocks seems to improve equally except for the same/no change condition which remains at a relatively constant level across blocks. The main effect of block for reaction time was not significant, $F(2, 50) = 2.059$, $p = .14$, neither was the interaction between block and change type for RT, $F(6, 150) = 1.656$, $p = .14$. Despite the salience of part information being increased, ability to detect changes to the identity of object parts did not increase relative to the other change conditions.

Experiment 3 was not designed to equalize the number of pixels. This experiment was designed to increase the salience of the part information of the objects by increasing the size of the parts and making configuration changes less distant (by shortening the bodies). As a consequence, more pixels change overall compared to Experiment 1 but we still get a similar

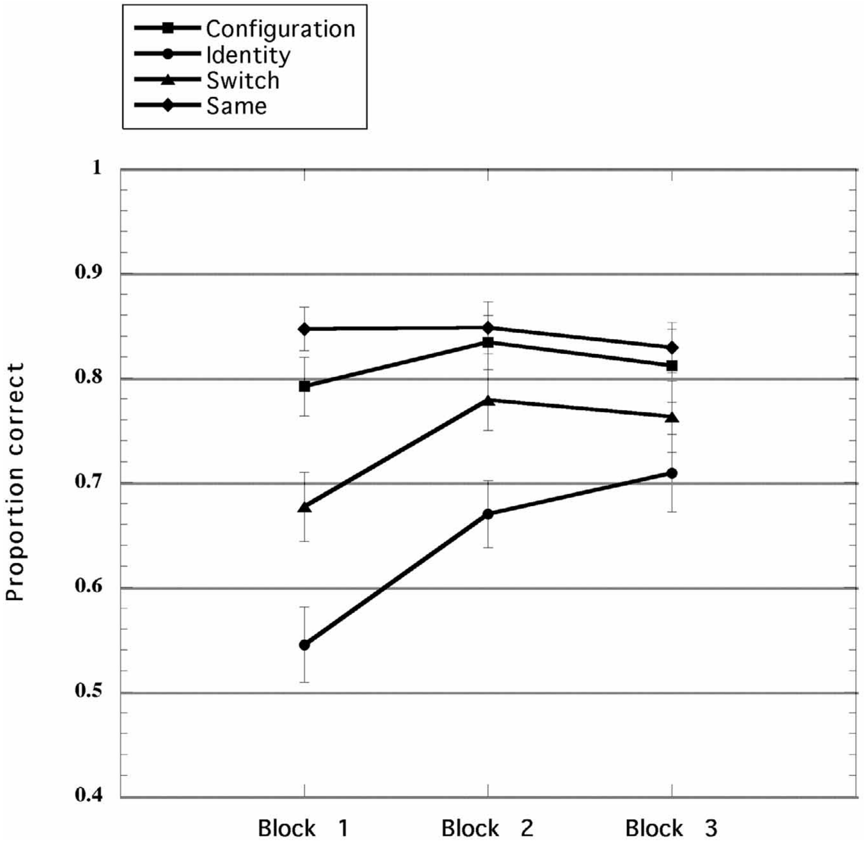

Figure 11. Mean proportion correct in each condition across blocks of Experiment 3.

pattern across conditions. The average number of pixels changing for each condition in Experiment 3 is 6438 (configuration), 3487 (identity), and 5577 (switch). Despite these differences in pixel change, the level of accuracy in detecting identity and switch changes are practically identical compared to Experiment 1 (proportion correct for identity change is .64 for both Experiments 1 and 3, and for switch change is .73 and .74 in Experiments 1 and 3, respectively). Further, the difference in the number of pixels changing between the switch and configuration conditions for both Experiments 1 and 3 are reasonably small, yet in both experiments configuration changes are detected significantly quicker and more accurately than switches. These results suggest that a measure of overall change (such as the pixel change count) is not the crucial variable.

EXPERIMENT 4

In Experiment 3 we attempted to address the problem of equalizing the magnitude of the different change types by increasing the size of the parts and decreasing the body size. This manipulation of the objects is biased against the spatial configuration changes and toward a part identity change. In Experiment 4 we explored configuration changes more closely. It could be argued that the global object shape changes produced by the configuration changes are more distinctive than the local feature or part identity changes.[1] To account for this possibility we used two different kinds of configuration changes, categorical and metric changes. Both changes are created by an equal-sized move along the body of the object (in terms of pixels moved), and presumably an equal-sized change in the overall global shape of the object. The only difference between the conditions is that the configuration of parts changes either categorically or metrically.

Categorical changes involve a part moving, for example, from below the other appendages of the object to above, whereas a metric change involves a part moving, for example, from below to further below other appendages (see Figure 12). If the size of the configuration changes are kept equal in terms of distance moved along the body, then ability to detect these changes should not be different if configuration judgements are made on distance or overall shape change alone. However, if discrimination judgements are based on the kinds of configuration changes being made (categorical versus metric) then there should be differences in the ability to detect these changes (Hummel & Stankiewicz, 1996).

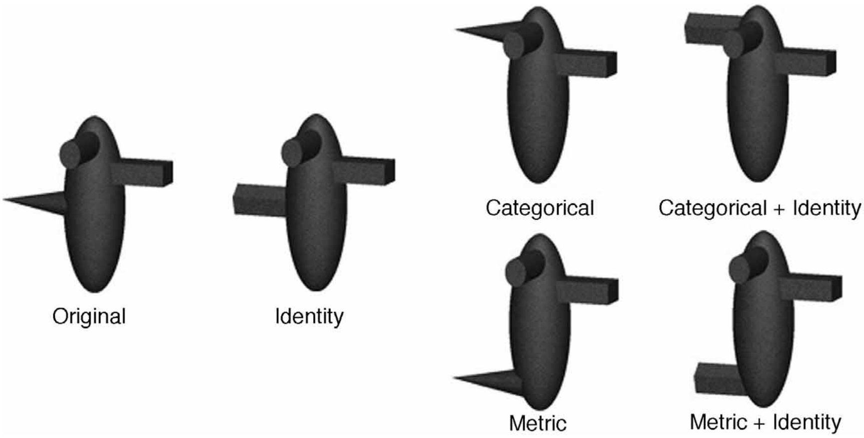

Figure 12. Examples of objects and types of changes used in Experiment 4.

[1] Thanks to an anonymous reviewer for this suggestion.

Method

Subjects. A total of 33 undergraduate students participated and were tested individually. Subjects received course credit for participating. None had participated in any of the three previous experiments.

Materials. The stimuli were similar to those used in Experiments 1 and 2. Stimuli were rendered images of three-dimensional novel objects. Each object was composed of a main body with three appendage parts. The appendages or ‘‘limbs’’ attached to the body at three of nine possible positions (see Figure 12 for examples). There were 6 ‘‘base’’ objects, each having configuration alone (categorical or metric/coordinate), identity alone, and both configuration and identity changes made to them, giving a total of 72 different object exemplars used in the current experiment. The distance an appendage part moved in configuration changes was identical for both categorical and metric changes (100 pixels, measured from the centre of the appendage). For any particular object the same part was involved in all the experimental conditions. The categorical change involved a part moving along the body such that it went from above another appendage to below (or vice versa), whereas the metric change involved that part moving along the body such that it was further above or below the two other appendages (see Figure 12 for examples). The two other appendage parts of the object were also involved in configuration and identity changes, however, these were included so that subjects would not focus attention solely on the part involved in the categorical or metric change (which would perhaps bias results). All objects were photorealistically rendered with the same colour and texture. Objects were shown at the same orientation and magnification. The entire background screen was white.

Procedure. The procedure for Experiment 4 was similar to the previous experiments; however, the timings used were shorter and there were no repeated blocks. The experiment consisted of 258 randomly ordered trials in which subjects viewed objects on a computer monitor. Each object was randomly placed at a position 25 pixels in any direction from the centre of the screen. Each trial began with a fixation cross appearing for 500 ms at the centre of the screen, followed by the first object for 2 s, immediately followed by a mask appearing on the screen for 1500 ms, and finally another object which remained on the screen until the subject responded or until the trial timed out at 5000 ms after the second stimulus was shown. The next trial began 1000 ms after the previous trial ended. The second object was either identical to the first or different in terms of part identity or configuration or both (as outlined above in the Materials section). Participants were asked to indicate whether the two objects presented to them were the ‘‘same’’ or ‘‘different’’ by pressing corresponding keys on a keyboard. Half of the trials were ‘‘same’’ trials and half ‘‘different’’.

Results and discussion

A one-way repeated measures ANOVA including the within-subjects factor of change type (same/no, categorical configuration, metric configuration, identity changes, categorical-identity, and metric-identity) was conducted on accuracy rates and on RT. As shown in Figure 13, participants were least accurate at detecting an identity change, better at detecting a metric change, and best at detecting same/no, categorical, categorical-identity, and metric-identity changes. The ANOVA on accuracy rates for change type showed a significant variation between conditions, $F(5, 160) = 33.927$, $p < .01$. Post hoc Scheffé contrasts showed that accuracy for identity change is significantly worse than all other change types (all $p < .001$). Accuracy for metric changes was significantly worse than all other change types except identity change (all $p < .001$). There were no significant differences between any of the remaining conditions (all $p > .05$).

The ANOVA on RTs for change type showed significant variation between conditions, $F(5, 160) = 8.243$, $p < .01$. As shown in Figure 14, the pattern for RT strongly reflects that of the proportion correct data. Subjects are slowest to

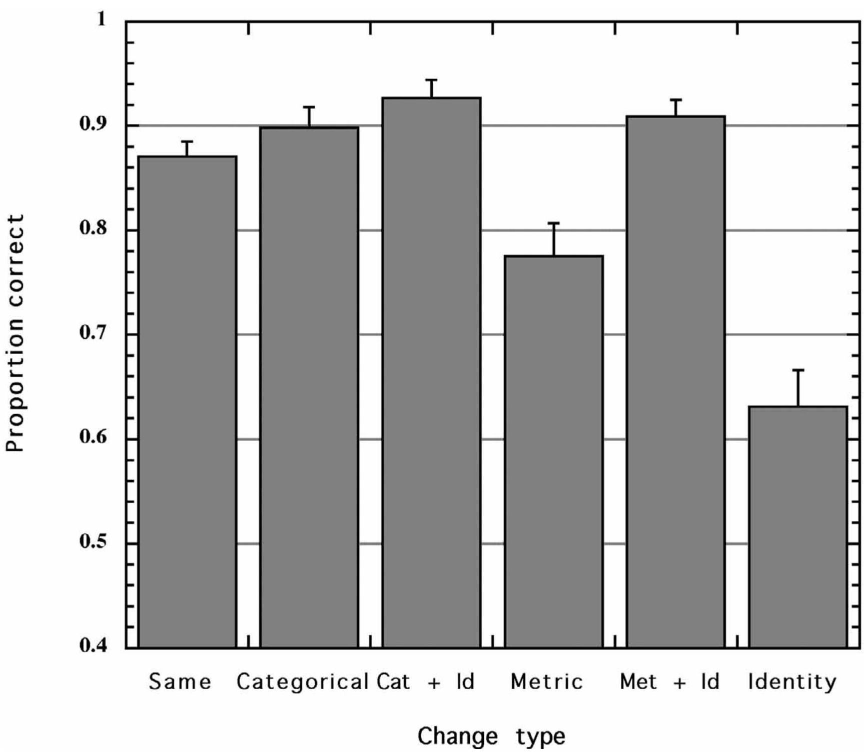

Figure 13. Mean proportion correct in each condition of Experiment 4.

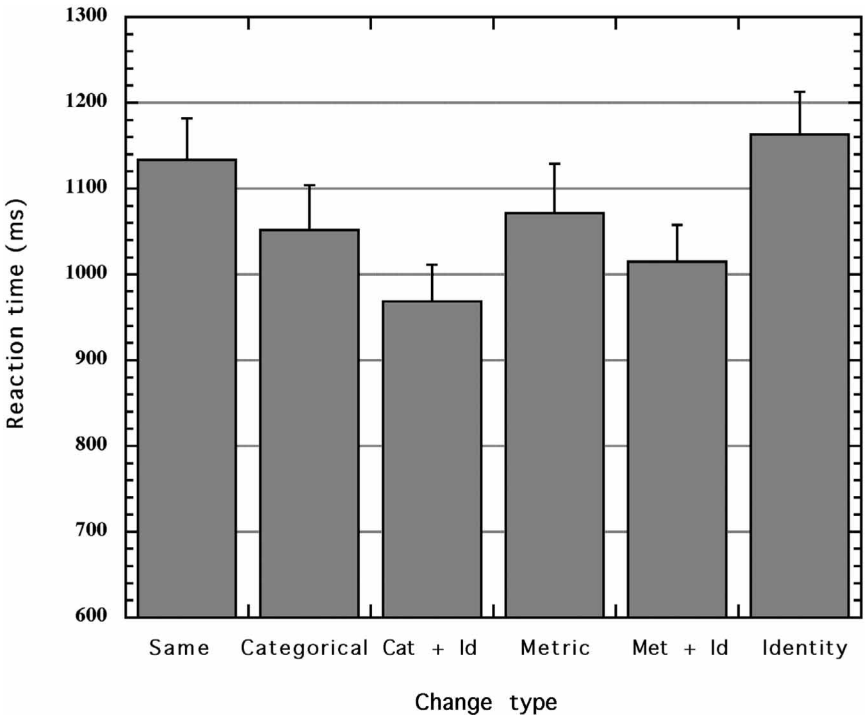

Figure 14. Mean reaction time in each condition of Experiment 4.

detect identity changes and quicker to detect categorical and identity and metric and identity changes. There was one difference between the patterns of means for accuracy and RT. Whereas subjects were more accurate at detecting categorical configuration changes than metric changes, they were equally fast at responding to these two conditions. Given the higher accuracy with equal speed for configural changes, we argue that this is evidence for better change detection performance with categorical than metric configural changes.

Categorical configural changes are detected more accurately than metric changes suggesting that it isn't the *size* of the change that is important, rather, it's the *kind* of change. Judgements seem to be made on the basis of the configuration of the object parts. Configuration is important and in terms of object perception we are more sensitive to categorical changes than metric changes. Importantly, both types of configuration changes are detected more accurately than identity changes. The addition of identity information does not appear to improve detection of categorical changes, but it does for metric changes, although this may be due to a ceiling effect. It again appears that the configural properties of an object dominates change detection performance.

GENERAL DISCUSSION

Despite the apparent simplicity of the experimental task, detecting whether a single object on a plain white background had changed after disappearing for about 4 s, participants are notably poor at detecting some types of changes in an object. In particular, the results of the present experiments show changes to the spatial configuration of object parts are quickly and accurately detected, whereas changes to the identity and the arrangement of object parts are more difficult to detect. This pattern replicates Simons' (1996) results with similar changes to arrays of two-dimensional novel and common objects.

Further, analyses of individual object changes in Experiments 1 and 2 showed that the variance due to the number of pixels changing did not account for the differences observed between conditions in either experiment. This implies that the present results are not entirely attributable to quantitative changes, rather at least some of the differences in performance between conditions are based on the qualitative nature of the changes. The results of Experiments 3 and 4 also support this idea that it is the qualitative nature of the changes being made that increases detectability rather than the size of the change.

Compared to Experiment 1, the results of Experiment 2 showed a decrease in accuracy but also shorter reaction times. However, the only methodological difference between Experiments 1 and 2 was the inclusion of an articulatory suppression task. The pattern of results suggests that subjects' response criterion was shifted by this manipulation. Although articulatory suppression was used to prevent verbal rehearsal, it appears that the task may have also incurred a cognitive load, thus affecting the speed/accuracy tradeoff. It is possible that as a result of this extra task, performance in Experiment 2 is based more on automatic processing than performance in Experiment 1, thus, subjects respond more quickly (because they do not engage in additional conscious processing) but are less accurate (for the same reason). Further, the fact that all conditions are affected equally suggests that differences between conditions are not due to verbal processing.

Experiments 3 and 4 used different methods to try to account for the possibility that the configuration changes were a result of larger physical changes. Experiment 3 used objects with increased salience of parts to bias against configuration changes. Experiment 4 used equal-sized distances in two types of configuration changes. Both experiments showed that despite these attempts to account for physical differences, configuration changes are easier and quicker to detect than part identity changes. Further, the results of Experiment 4 demonstrated that not all configural changes are equal. The detection of categorical changes is significantly easier than metric changes (both of which are detected more accurately than part identity changes).

The finding that categorical changes to the configuration of object parts are detected more accurately than metric changes is an important element of these results. Besides clarifying that performance in change detection and object perception depends upon the type rather than the size of the changes made, a particular type of configural information has been found to be more useful. Specifically, a change that does not affect the qualitative nature of the components results in less accurate change detection than a configuration change that does alter the qualitative relations between components. Here we see a caveat on our (and Simons', 1996) claim that configural changes are more easily detected than other changes. Such changes will be easiest to detect when they result in a change to the categorical, or qualitative, nature of part relations in an object. Such a finding is consistent with structural description models (e.g., Hummel & Biederman, 1992; Hummel & Stankiewicz, 1996), which propose feature binding on the basis of qualitative relations between object components.

A notable difference between the present results and those of Simons (1996) is that subjects performed better in the switch than the identity condition. One possible explanation is that a switch change made to three-dimensional objects is not the same as that made to two-dimensional arrays. A switch change for three-dimensional objects involves changes to the two-dimensional appearance of the object parts. Since a switch involves two parts and an identity change only one, perhaps this results in a greater change to the overall appearance of the object.

Spatial configuration changes to object parts are more readily detected than shape changes. These results may have important implications for theories of object recognition. In terms of GSD theory, it may be that the spatial relations between object parts are computed before the identity of the parts are processed (Cave & Kosslyn, 1993; Kimchi, 1992; Kimchi & Bloch, 1998; Palmer, 1978). Palmer, for example, has shown using a same–different verification task that the comparison of two line drawings was made on the basis of higher order structures of the figures rather than on their individual line segments. Kimchi and Bloch argue that when configural and component properties are present and can be used for the task at hand, configural properties dominate discrimination and classification performance. When configural properties are not of use for completing the task, then component properties are used, but at a cost. This may account for the results of the present experiment in that if the overall spatial configuration of an object is computed first, it is used quicker than and in preference to part identity information.

In an attempt to integrate some of the features of structural descriptions with image-based theories of object recognition, Tarr and Bülthoff (1998) suggest that associations between views may be formed on the basis of global information. One way in which this might occur is through the use of ''medial-axis'' representations (Kovacs & Julesz, 1994), which are derived from an object's silhouette. The representation takes the form of a skeleton describing the global structure of the object, thus, the information it provides is limited. It is important to note that medial-axis representations are postulated only as a supplement to object recognition, in that they may help constrain the search space during recognition, but, in and of themselves, they are not sufficient for recognition.

Medial-axis representations could be used to account, at least in part, for the finding that configural information is used quicker and more accurately than part identity information in Experiments 1, 2, and 3. A configuration change will result in a different medial axis representation whereas a part switch or identity change will not. Tarr and Bülthoff's (1998) idea of including structural information in the representations of objects corresponds with Simons and Levin's (1997) idea that gist and perhaps spatial layout are used to integrate information across views. However, medial-axes cannot account for the results of Experiment 4. Both categorical and metric changes will result in changes to the medial-axis. It appears that the kinds of configural relations that change is an important factor to consider in change detection for object parts.

A workable theory of object recognition can be constructed that could account for the present results. Such a theory might propose that the overall spatial configuration of an object is computed first (with priority given to categorical relations), then finer details such as object parts are analysed. When perceiving an object, if the spatial configuration information is not sufficient for recognition or detection of change, then the visual system goes on to access details such as the identity of the object parts (Kimchi & Bloch, 1998). These processes need not necessarily be completed serially. If they are carried out in parallel, it may be that the analysis of spatial configuration provides more useful information for the recognition of objects and is given preference in terms of accessibility. Alternatively, configural information may be processed parallel with, but faster than part identity information.

Another possible explanation of the relatively good performance found for detecting configuration changes is that rather than configural information being preferentially accessed, there exists some special memory for layout that is stable or ''non-volatile'' (Chun & Nakayama, 2000; Rensink, 2000). Rensink suggests that a dynamic form of visual representation is used for scene perception. This architecture uses representations that are stable and representations that contain large amounts of visual detail, but never representations that are stable *and* contain large amounts of detail. Because representation of layout is stable and usually invariant, perhaps changes made to object configuration are more easily detected than those made to representations containing more detail, such as part shape or identity.

Finally, using isolated objects we have found results similar to those of scene perception. Simons (1996) found that memory for the spatial configuration of objects in a two-dimensional array was rapid and accurate, whereas identifying a change to the order of the objects or the identity of one of those objects was significantly slower and less accurate. In terms of parsimony, it is not unreasonable to assume that a similar mechanism for the perception of spatial layout or configuration applies to both objects and scenes.

The explanation of the present results is compatible with Simons' (1996) findings. Overall spatial configuration information appears to be used quickly and in preference to other types of available information. Further, the explanation given earlier could apply to the perception of natural scenes, with the overall layout of objects within the scene being computed before the identity of those objects. We are not implying that objects are viewed as scenes or vice versa. Rather, we suggest that the perception of complex visual information is executed in an hierarchical fashion, the configuration being computed and used before object or part properties.

REFERENCES

Aginsky, V., & Tarr, M.J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition*, *7*(1/2/3), 147–162.

Biederman, I. (1987). Recognition-by-components: A theory of human vision understanding. *Psychological Review*, *94*, 115–145.

Biederman, I., & Cooper, E.E. (1991). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 121–133.

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

Cave, C.B., & Kosslyn, S.M. (1993). The role of parts and spatial relations in object identification. *Perception*, *22*, 229–248.

Chun, M.N., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*, *7*(1/2/3), 65–81.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Vancouver Studies in Cognitive Science: Vol. 2. Perception* (pp. 89–110). New York: Oxford University Press.

Henderson, J.M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science*, *8*(1), 51–55.

Hoffman, D.D., & Richards, W.A. (1984). Parts of recognition. *Cognition*, *18*, 65–96.

Hoffman, D.D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *63*, 29–78.

Hollingworth, A., & Henderson, J.M. (1997). *Semantic informativeness mediates the detection of changes in natural scenes* (Tech. Rep. No. 6). Michigan, IL: Michigan State University Eye Movement Laboratory.

Hummel, J.E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.

Hummel, J.E., & Stankiewicz, B.J. (1996). Categorical relations in shape perception. *Spatial Vision*, *10*(3), 201–236.

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

Kimchi, R., & Bloch, B. (1998). Dominance of configural properties in visual form perception. *Psychonomic Bulletin and Review*, *5*(1), 135–139.

Kovacs, I., & Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, *370*, 644–646.

Levin, D.T., & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin and Review*, *4*(4), 501–506.

McConkie, G.W., & Currie, C.B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 563–581.

O'Regan, J.K. (1992). Solving the ''real'' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461–488.

Palmer, S.E. (1978). Structural aspects of visual similarity. *Memory and Cognition*, *6*, 91–97.

Rensink, R.A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1/2/3), 17–42.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

Saiki, J., & Hummel, J.E. (1998). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 227–251.

Simons, D.J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, *7*(5), 301–305.

Simons, D.J., & Levin, D. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*, 261–267.

Simons, D.J., & Levin, D. (1998). Failure to detect changes to real people during a real-world interaction. *Psychonomic Bulletin and Review*, *5*(4), 644–649.

Tarr, M.J., & Bülthoff, H.H. (1998). Image-based object recognition in man, monkey, and machine. *Cognition*, *67*, 1–20.

Williams, P., & Simons, D.J. (2000). Detecting changes in novel 3D objects: Effects of change magnitude, spatiotemporal continuity, and stimulus familiarity. *Visual Cognition*, *7*(1/2/3), 297–322.

Williams, P., & Tarr, M.J. (1998). RSVP: Experimental control software for MacOS. [Computer software]. Available: http://www.cog.brown.edu/~tarr/RSVP/

Manuscript received April 2001
Revised manuscript received February 2002