Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 45 (2005) 2065–2080

Vision Research

www.elsevier.com/locate/visres

The interaction of shape- and location-based priming in object categorisation: Evidence for a hybrid "what + where" representation stage

Fiona N. Newell [a,*], Dianne M. Sheppard [b], Shimon Edelman [c], Kimron L. Shapiro [d]

[a] *Department of Psychology and Institute of Neuroscience, University of Dublin, Trinity College, Dublin 2, Ireland*
[b] *Department of Psychology, School of Psychology, Psychiatry and Psychological Medicine, Monash University, Clayton Campus, Victoria 3800, Australia*
[c] *Department of Psychology, 232 Uris Hall, Cornell University, Ithaca, NY 14853-7601, USA*
[d] *School of Psychology, University of Wales, Bangor, Gwynedd, LL57 2AS Wales, UK*

Received 4 April 2003; received in revised form 12 October 2004

Abstract

The relationship between part shape and location is not well elucidated in current theories of object recognition. Here we investigated the role of shape and location of object parts on recognition, using a classification priming paradigm with novel 3D objects. In Experiment 1, the relative displacement of two parts comprising the prime gradually reduced the priming effect. In Experiment 2, presenting single-part primes in locations progressively different from those in the composite target had no effect on priming. In Experiment 3, manipulating the relative position of composite prime and target strongly affected priming. Finally, in Experiment 4 the relative displacement of single-part primes and composite targets did influence response time. Together, these findings are best interpreted in terms of a hybrid theory, according to which conjunctions of shape and location are explicitly represented at some stage of visual object processing.

Keywords: Form; Object categorisation; Priming; "What" and "Where"

1. Introduction

Much of the current research in high-level vision focuses on object recognition, a task in which human observers excel, and which is commonly considered to be the epitome of the challenges that computer vision systems have yet to meet. In cognitive psychology, the last several years saw three special issues of journals devoted to object recognition (Vision Research 38(15, 16), 1998; Cognition 67(1, 2), 1998; Acta Psychologica 102(2, 3), 1999). Likewise, in computational vision, a number of recently published books have dealt with object recognition (Edelman, 1999; Ullman, 1996).

There are, however, other high-level visual tasks that relate to object shape, yet are not subsumed under the rubric of recognition, even if the latter is construed widely to include old/new identification, forced-choice classification, and categorisation. These are the tasks that require the observer to deal with object or scene *structure*, usually explicitly ("does this chair have armrests?"—locate the armrests), but sometimes implicitly ("will my cat be able to climb that ladder?"—locate the rungs and estimate their spacing in units of cat length). To understand the computational (and, eventually, the neural) basis of human performance in such tasks, one needs to examine theoretical approaches to

* Corresponding author. Tel.: +353 1 608 3914; fax: +353 1 671 2006.
E-mail address: fiona.newell@tcd.ie (F.N. Newell).

0042-6989/$ - see front matter
doi:10.1016/j.visres.2005.02.021

structure representation and processing, and test their predictions concerning the effects of structure manipulation in controlled experiments. We describe the results of four such experiments, and discuss their implications for two approaches to object representation found in the current literature: structural and image-based.

1.1. Structural description models

The best-known structural theory, Biederman's RBC (Recognition By Components), postulates an explicit treatment of structure in recognition and categorisation, pointing out that the latter task can be made especially easy by the availability of a "structural description" of the object in terms of its complement of generic parts and the prevailing spatial relations (Biederman, 1987). The RBC theory posits a small set of generic primitive shapes ("geons"), which are assumed to be easily detected in images due to their non-accidental properties. The latter are 3D features that are almost always (that is, barring an accident of viewpoint) preserved by the imaging (projection) process (Lowe & Binford, 1985).

To be able to recognise novel objects, a model based on structural descriptions must form the representation of the whole in terms of its parts dynamically (i.e., "on the fly"), for each shape it encounters. The JIM ("John and Irv's Model") implementation of the RBC theory described by Hummel and Biederman (1992) is an example of such a model. It is important to note that this implementation includes special relational units dedicated to the binding operation, over and above the shape units dedicated to each of the geons, as explained in Hummel and Biederman (1990, p. 619).

The current version of the Hummel/Biederman model, JIM.3 (Hummel, 2001), contains two binding mechanisms: a dynamic one and a static/retinotopic one, working side by side. The assumption is that the dynamic mechanism, which produces standard structural descriptions, is the preferred path, although it requires attention and is thus more time consuming. In recent experiments, JIM.3 was trained on single views of 20 objects, then tested on translated, scaled, reflected and rotated (in the image plane) versions of the same images (all line drawings). The model exhibited a pattern of results consistent with a range of psychophysical data obtained from human participants (Hummel, 2001; Stankiewicz, Hummel, & Cooper, 1998): its categorisation performance was invariant with respect to translation and scaling, and was reduced by rotation. When time is short, or when attention is scarce, JIM.3 falls back onto the use of static binding, producing a representation that is not as invariant as the dynamically bound one under various image transformations (notably, translation).

1.2. Image-based models

The holistic image-based approach suggests that objects are represented as collections of entire viewpoint-specific 'snapshots' (Tarr, 1995; Tarr & Bülthoff, 1998). The greatest challenge to the holistic image-based models lies in capturing the compositional aspects (Bienenstock & Geman, 1995) of object representation in human vision. If the structure of parts comprising an object is not made explicit, the model will lack certain features of the human competence in the domain of object perception, such as judging the similarity of composition, as opposed to the similarity of the global shape (Hummel, 2000). The need to treat object structure explicitly requires relaxing the holistic outlook of image-based models.

A recently proposed image-based model, the Chorus of Fragments (CoF), addresses this issue by using "parts" that are spatially anchored (i.e., are actually localised image fragments) rather than either floating or holistic (see Edelman & Intrator, 2003 for details). Instead of temporal binding, CoF uses binding by retinotopy (Edelman, 1999; Edelman & Intrator, 2000; Edelman & Intrator, 2001). In this approach, structure is represented explicitly, but in an image-based rather than object-centred manner (as in the static stream of Hummel's JIM.3). Indeed, the representational substrate in the CoF model is best conceptualised as an ensemble of "what + where" units, each of which is selective both to shape ("what") *and* location ("where") of the stimulus;[1] multiple units with similar shape selectivities are assumed to exist in various image loci.

We decided to investigate the roles of shape and location information in object recognition by manipulating the relative position of parts of a priming object, or the location of a complete prime, with respect to the target object, and measure the resulting priming effect.[2] We can now formulate the predictions of structural and image based models with respect to the kind of priming one should expect.

In the context of structural models, priming by two kinds of stimulus characteristics is expected. First, the shape units should respond to their preferred stimuli (geons) irrespective of their location in the image, leading to shape-based priming that is insensitive to the location of the shape. Second, the relational units should give rise to relation-based priming in which the relative

[1] Neurons with such response selectivity are common in the inferotemporal and the prefrontal areas of the monkey cortex (Op de Beeck & Vogels, 2000; Rao, Rainer, & Miller, 1997).

[2] Priming is defined as a modification of performance that (i) stems from exposure to a stimulus, and (ii) persists over time and manifests itself when the participant subsequently encounters similar stimuli (Ochsner, Chui, & Schacter, 1994; Tulving & Schacter, 1990).

position of object parts has a categorical (all-or-none) effect.[3] For example, displacing one object part that is to the left of another will result in a categorical change in the relation 'to left of' and will reduce the priming.

The image-based CoF model, on the other hand, proposes that the representations of shape and of retinal location are inextricably interwoven, so that a spatial predicate such as "above" is not represented explicitly as in structural models, but is represented only as the disjunction over the activities of all object-specific modules that "look" at the upper visual field. Consequently, translation-invariant priming is not expected for spatial relations. Moreover, the mutual priming between two shapes is expected to be the stronger, the closer their two retinal locations. These predictions can be contrasted with those of structural models, which predict priming for spatially "floating" geons.

1.3. Previous related work

A few studies have already manipulated part structure of the stimuli to characterise the effects of structural variables on recognition (e.g., Fiser & Biederman, 1995; Fiser & Biederman, 2001). For example, Biederman and Cooper (1991) used line drawings of familiar objects and reported that deletions of object components rather than deletions of object features caused a reduction in long-term priming effects. This result suggested that priming was activated by object components and their specified relations. Cave and Kosslyn (1993), who examined the effect of various kinds of object decomposition on time to name line drawings of familiar objects, found that the spatial arrangement of component parts of an object was important for recognition. However, they also reported that the manner in which an object is divided into parts has minimal effect on the time it takes to recognise it (Cave & Kosslyn, 1993). These data speak against the hallmark prediction of structural theories of object recognition: that object identification results from the decomposition of the object into predetermined parts or geons.

With regard to image-based models, a recent study addressed the role of part structure in object recognition by examining the effects of translation on object discrimination (Dill & Edelman, 2001). Dill and Edelman found complete translation invariance when the (same/different) task involved a *local* (image-based) discrimination (stimuli were composed of different parts, but matched in terms of spatial configuration of parts). The invariance was lost, however, when participants were asked to perform a *structural* discrimination (stimuli were composed of the same parts in different spatial configurations). As suggested by the authors, these results call for a model that would treat local and global/structural shape information differently, so that local features, but not specific arrangements thereof, would be processed in a translation-invariant manner. This kind of behaviour is compatible with the predictions of the CoF model of object recognition (Edelman, 1998, 1999; Edelman & Intrator, 2003).

Our present study aimed to investigate further the mechanisms behind the representation of object structure, focusing on two issues: the relationship between shape and location (represented independently or not), and spatial relations (categorical or graded). Following the logic of Biederman and Cooper (1992), we used a priming paradigm (with object classification time as the dependent variable) in an attempt to probe specifically those representations that are normally used for object recognition. We chose short-term priming, which, we felt, was better suited to the examination of the effects of object position on recognition.

The standard structural model suggests that a structural description is categorical and object-centred, and is encoded separately from the categorical information concerning the shapes of the parts. Accordingly, it predicts that the magnitude of priming should be reduced when the shapes of the parts or their structural arrangement change from prime to target. The image-based approach, that is, the CoF model, instead predicts graded position-dependent, shape-related priming, and no priming specific for spatial relations.

The study reported here consisted of four experiments involving forced-choice classification of novel objects. Experiment 1 specifically addressed the representation of structural relations of two-part novel objects. Priming effects were measured when the object shape of the prime was kept constant but its structural relations were altered. Experiment 2 examined the priming effects of *single part* (or geon) primes in various positions relative to the position of parts in the target objects. As the single-part priming effects turned out to be largely invariant to stimulus translation, Experiment 3 examined the effects of *two-part* primes, again in various positions relative to the target objects. Experiment 4 was conducted to follow up the seemingly incongruent results of Experiments 2 and 3: the priming effects obtained in Experiment 3 (two-part primes) were dependent on prime position, whereas those of Experiment 2 (single-part primes) were not. Thus, Experiment 4 used the same single-part primes as in Experiment 2, but adopted the paradigm of Experiment 3.

[3] The standard structural model can be modified to yield graded rather than all-or-none behaviour with respect to stimulus manipulations mentioned here (e.g., by assuming that its states are probabilistic). We decided not to consider here any such modifications, which would result in a qualitatively different model, rendering the standard structural description theory of representation effectively unstable. For a discussion of the testability of structural models, see Sanocki (1999).

2. General methods

2.1. Participants

Fifteen undergraduate students (mean age = 20.5 years, SD = 3.7 years) from the University of Wales, Bangor, participated in the Experiments 1, 2 and 3 either for a small payment or for course credit. Three of the participants were male. All participants had normal, or corrected-to-normal, vision. The order of the experiments was counter-balanced across participants.

2.2. Apparatus

An IBM computer with a 266 MHz Pentium II processor and a 800 × 600 Mitsubishi Diamond Pro 87 TXM monitor was used along with E-Prime™ software to program and run the experiment. A standard, English-language keyboard configuration was used for responding.

2.3. Target stimuli

The stimuli were created using 'Extreme 3D for Macintosh' software, and then saved as bitmap files for use with E-prime. We designed our stimuli so that each object afforded a unique geon structural description, as per Biederman and Gerhardstein (1993). Each target stimulus consisted of two unique geons (see Fig. 1(a)). Thus, each target constituted a unique category of object. The configuration of two of the four targets was such that one part appeared to the left and the other to the right of fixation. The other two targets were in an above/below fixation configuration. The component parts were standardised for size as much as possible (each part was approximately 22.5 mm^2 or 2.3° × 2.3° front view), and the point at which the parts were joined overlapped fixation. Thus, the maximum extent of each target, prime (including displacement) or mask display could be contained in a circle whose radius subtended a visual angle of approximately 2.3°. The target and prime objects used throughout the study were rendered with a metallic bronze finish with a shadowing effect to enhance the 3D appearance. The mask used for the target and prime objects consisted of a randomised mosaic of parts from each of the four targets.

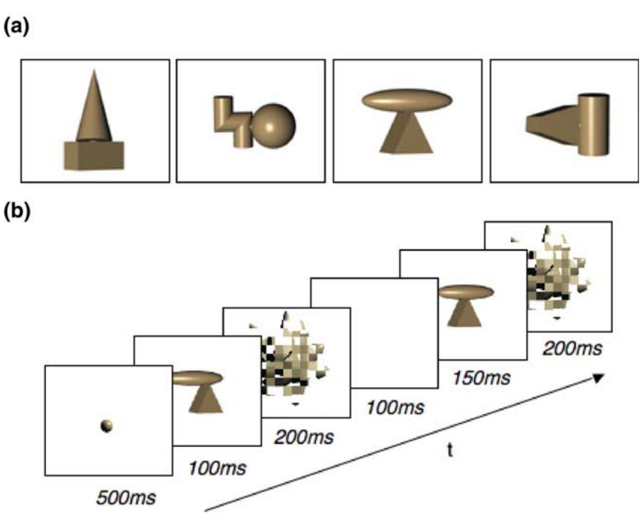

Fig. 1. (a) The four targets (A–D reading left to right), each with unique parts, used in Experiments 1–4. The parts of Targets B and D were positioned to the left and right of fixation, and for A and C above and below fixation. Each part was approximately 22.5 mm^2 (front view), and the point at which the parts were joined overlapped fixation. (b) An illustration of the priming paradigm used in our experiments. The illustration shows the structure of a typical trial involving, in sequence, the following events: a fixation, prime, mask, blank, target and mask. In this example the prime stimulus is from Experiment 1.

2.4. Prime stimuli

The relatively novel 'incremental priming technique' (Jacobs, Grainger, & Ferrand, 1995) was used in this study so that the magnitude of the priming effect could be assessed according to two different baselines (a within condition and a between conditions measure). The prime images were presented at three incremental levels of intensity (low, moderate and maximum intensity), in a pseudo-random repeated measures block design. The idea here was that with each incremental increase in prime intensity, the prime would become increasingly available to the shape processing system and any increase or decrease in response time (RT) due to the prime should increase in magnitude respectively.

The intensity levels were produced by manipulating the *luminance contrast* of the prime images (relative to the targets and backward masks) through added levels of lightness. The luminance of the screen background was 51.4 cd/m^2, and the mean luminance of the targets and masks was 5.4 cd/m^2. The first prime intensity level was the high luminance contrast or *low intensity level* (with 90% lightness applied), with a mean luminance contrast of 75.6% (calculated using the Mitchelson fraction). The second level was *moderate intensity level* (with 45% lightness applied), with a mean luminance contrast of 50.8%. Finally, the third level was the *maximum intensity level* (no lightness applied, thus no luminance contrast). Thus, the three incremental levels of prime intensity used throughout the experiments were low, moderate and maximum intensity.

2.5. Design

As mentioned above, the targets throughout the experiment were randomly selected from one of four two-part objects (see Fig. 2(a)). The order of the first

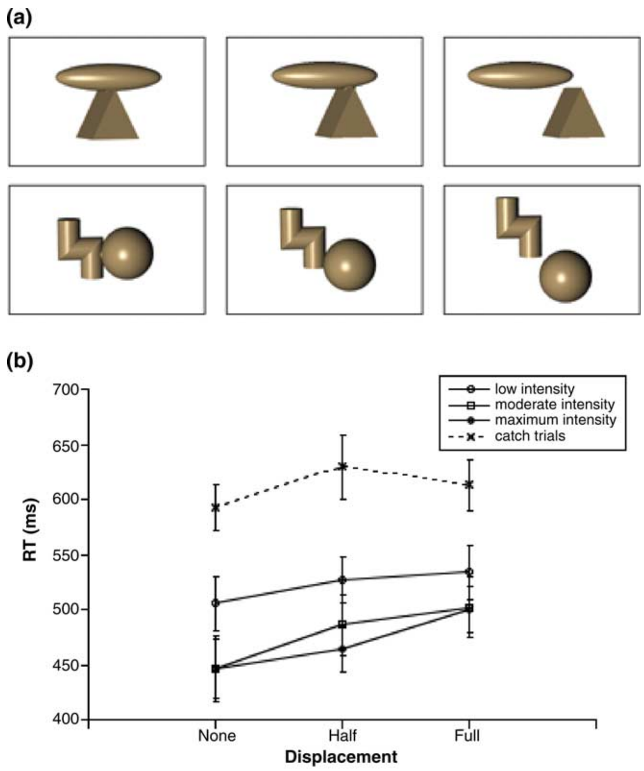

Fig. 2. (a) Examples of the three priming conditions used in Experiment 1. Note that the above prime stimuli use Target C (up/down parts) and Target B (left/right parts) for illustrative purposes only—the prime object could be any one of the four target objects in one of the three 'displacement' conditions. (b) Mean RT (ms) for increasing levels of displacement of the prime's parts for each prime intensity level in Experiment 1. Intensity levels included low (circles), moderate (squares) and maximum intensity (diamonds). The mean RT (collapsed across intensity) for the Catch (different object) Trials is also shown. The error bars indicate the standard error of the mean.

three experiments was counterbalanced across participants. There were three between-subject orders used in the experiment (1, 3, 2; 2, 1, 3; and 3, 2, 1), and each of these was used five times across the sample of 15 participants. Regardless of experiment order, before each next experiment there was an initial set of 24 practice trials (of moderate prime intensity) to familiarise the participant with the new priming procedure. Within each experiment there were three blocks of trials (see below for the number of trials per block for each experiment respectively), each of which presented primes at one of three intensity levels. The trials were blocked by intensity to avoid confusing the participant and to facilitate performance in the low and moderate intensity trial blocks. The order of the prime intensity level blocks was pseudo-randomly varied between participants.

2.6. Procedure

The task for the participant was to classify the novel target object (see Fig. 1(a)) using a 4-alternative forced choice design. The participant responded to each target object by pressing one of four keys on the computer keyboard ('g', 'h', 'j' or ',') using their dominant hand only. The experiments began with a training session and participants were allowed to proceed to the test once criterion performance (measured in speed and accuracy) was reached with training. Feedback was given after each trial during the training session. Each experiment took approximately 20 min to complete.

2.6.1. Training block

Participants were first required to complete a training phase that was essentially a 4-alternative forced-choice classification task. A trial consisted of a fixation dot for 400 ms, followed by one of the four targets chosen at random, presented for 150 ms. The target was immediately followed by the mask (200 ms), which was replaced by a blank screen until a response was made. During training the participants were given feedback regarding the accuracy and timing of their responses directly after each trial.

Each participant was required to reach an accuracy criterion of 80% correct and a mean RT criterion of 900 ms or faster before moving on to the first experiment. If after the first block of 36 training trials the participant failed to reach either criterion, the same training block (in a different random order of presentation) was repeated until both were attained. For Experiment 1, 2 participants reached criterion after only 1 repetition of the training block, 8 after 2 repetitions, 3 after 3 repetitions, 1 after 4 repetitions and the remaining participant required 5 repetitions.

2.6.2. Priming block

The parameters for the *priming conditions* remained the same across all experiments, despite changes in the type of prime object and spatial locations used. Fig. 1(b) illustrates a typical trial structure used in our experiments. The start of each trial (i.e., immediately before the onset of fixation) was signalled by a short 300 ms sound. A different short 300 ms sound, presented immediately after a response was made, signalled the end of each trial. These sounds were for the purpose of monitoring eye movements (see Experiment 3). Following fixation (500 ms), a prime was presented for 100 ms (too brief to make a saccade), followed by a mask for 200 ms. The target (identity unpredictable) was then presented for 150 ms following a blank interval of 100 ms. Finally, a mask was presented for 200 ms, and a blank screen followed until a response was made. No immediate feedback was given, however, at the end of each block of prime trials participants received summary feedback regarding their average RT and accuracy performance. This feedback also warned them (if necessary) when their mean accuracy and/or response times fell below criterion.

Any specific methodology details are mentioned under each Experiment.

3. Experiment 1

In this experiment, we manipulated the similarity of the *within-object spatial structure* between the prime and target stimuli by altering the relative position of parts of the prime objects. Only the structural arrangement of the prime object's parts was allowed to change. There were three levels of displacement of the two parts comprising the *prime*: 'no displacement', 'half part displacement' and 'full part displacement'; the two-part *target* object was always intact. Recall that the structural model predicts that the target will be maximally primed in the 'no displacement' condition (same structural description and same parts). Once the structural description of the prime is altered relative to the target, the structural model predicts no 'part-relation priming' at all (only part-based priming). This effect would be twofold: a relative reduction in priming for the two part displacement levels, and no difference in the magnitude of priming between the half and full displacement levels themselves. In contrast, the Chorus of Fragments (CoF) model, which holds that structure is represented explicitly in a coarse-coded image-based fashion, predicts a more gradual, monotonic decrease in priming as the relative displacement between parts increases. In both cases, some residual priming is predicted for the two part displacement levels due to the presence of identical geons in the target and prime displays.

3.1. Method

In this experiment, the primes and the targets were constructed from identical parts. However, the *relative position of parts* in each prime was manipulated by displacing one relative to the other. There were three different levels of displacement of prime parts: none, half part displacement (maximum shift of 0.6°), or full part displacement (maximum shift of 1.2°) (see Fig. 2(a) for examples). Thus, the *spatial structure* of the prime was manipulated to determine the effect on the amount of priming.

Catch trials were introduced to ensure that the target was not fully predictable given the prime. Therefore, 25% of targets were preceded by 'different object' primes (also in one of the three displacement configurations). These trials were not used in the analyses but allowed the examination of the overall extent of perceptual priming in the 'same object' condition. Each of three blocks in this experiment consisted of 48 trials (12 'different object' primes, and 12 trials per 'same object prime' displacement condition), which resulted in a total of 144 trials.

3.2. Results and discussion

Catch trials (different object primes) and incorrect trials were excluded from all RT analyses. In addition, RT outliers (±2.5 SDs from mean) were removed from each participant's data. This resulted in the removal of an average of only 1.65% of trials per participant. As reflected in statistics as well as Fig. 2(b), RT increased with displacement regardless of the intensity level of the prime. As the mean RT for the catch trials or different-object prime trials was much slower than the RTs for the same-object primes (see Fig. 2(b)), the increase in RT with an increase in prime part displacement is probably better described as a *decrease in facilitation*. The RT facilitation also decreased with the intensity level of the prime, which rendered the prime less effective.

A 2-way ANOVA with displacement (none, half, and full) and intensity (low, moderate, max) as factors showed significant main effects of both displacement ($F(2,28) = 12.04$, $p < .001$) and intensity ($F(2,28) = 6.79$, $p < .005$), and a non-significant interaction, $F(4,56) < 1$. Post hoc Newman–Keuls tests revealed that RTs to the 'no displacement' condition were significantly faster than those to the 'half displacement' condition ($p < 0.01$) and the 'full displacement' condition ($p < 0.001$). Furthermore, RTs to the 'half displacement' condition were also significantly faster than to the 'full displacement' condition ($p < 0.05$). This pattern of *gradually* increasing facilitation with decreasing part displacement fits the image-based, CoF model hypothesis. Structural models, on the other hand, predicted 'all-or-none' structural description priming (i.e., no structural description priming expected *at all* for the two 'part displacement' conditions, only shape-related priming, which should not differ).

An additional 2-way ANOVA, again with displacement (none, half and full) and intensity (low, moderate, max) as factors, was conducted on the percentage error data. The main effects of displacement ($F(2,28) < 1$) and intensity ($F(2,28) = 2.55$, $p = 0.10$) failed to reach significance, as did the interaction ($F(4,56) < 1$).

In sum, the findings show that the relative position of parts of objects affected priming in a graded manner. In addition, the prime's intensity had the expected effect on RT: as the intensity increased, the same primes produced more RT facilitation. Although the size of the displacement effect was not significantly affected by prime intensity (non-significant interaction), the moderate and maximum intensity levels did produce a numerically larger displacement effect than the low intensity level (within condition baseline) in accordance with our predictions.

The observation could be made that the target objects used in Experiment 1 consisted of two attached parts, whereas the parts were often separate in the prime stimuli (e.g., in the full displacement condition). In effect, 'separation' per se may be considered an additional, non-accidental relation. Thus, the relative differences in facilitation between the displacement conditions may be due to a confound, i.e., that this additional relation is present in the 'half' and 'full displacement' primes, but not in the 'no displacement' primes. We feel this is an unlikely account of our findings in Experiment 1, simply because an additional relational difference would decrease the likelihood of priming for both the half and full displacement, whereas we found clear evidence of priming in both these cases. Nonetheless, we repeated Experiment 1 with 15 new participants (mean age = 28.3 years, SD = 5.9 years), and introduced a small gap between the parts of the object primes. The resulting gap size between the two parts was no greater than the largest distance between two geons in the prime stimuli from the 'full displacement' condition, (i.e., the gap between the two parts was never greater than 2 mm). This replication led to an essentially equivalent data set.

4. Experiment 2

Of interest in this and the two subsequent experiments was the effect of the relative position of objects, or the translation of the *prime* stimulus within the visual field, on the magnitude of perceptual priming. In terms of theoretical predictions, the position held by structural theorists was made clear in a recent paper that stated "*Supra*liminal visual priming is thus likely to affect an area with RFs (receptive fields) large enough to fully accommodate the translation..." (Bar & Biederman, 1998, p. 468). Thus, the structural model predicts geon-related visual priming effects (i.e., RT facilitation relative to the different geon condition), regardless of the relative position of objects in the visual field.

The image-based CoF model, in comparison, postulates that *conjunctions of spatial location and shape* of the object are explicitly represented. If that is the case, then both target *position and shape* should be amenable to priming. The greatest RT facilitation was, therefore, predicted for the condition in which both the shape and the position of the single geon prime are identical to those in the target, and deviations in either shape or position were expected to result in less facilitation. Moreover, the magnitude of geon-based visual priming should be dependent on the position of the prime—it should decrease as relative displacement of objects increases.

This experiment presented single-part or geon primes at the same or a somewhat different position relative to the corresponding part in the target object. In addition, the primes were either part of the subsequent target object (same geon condition) or part of a different target object (different geon condition).

4.1. Method

4.1.1. Prime stimuli

In this experiment, the prime display consisted of a single component part or geon that was either the same as (50% of trials) or different from (50% of trials) one of the geons in the following target. The single-geon prime was the same size as its corresponding part in the target, and occurred in one of three positions relative to its position in the target—the 'same position', or one of two different positions ('position 1' and 'position 2') (see Fig. 3(a)). Targets of the above–below fixation part-configuration (Targets A and C) were primed either by a geon occurring in the 'same position', 'position 1' (a geon slightly to the left or right of its position in the target) or 'position 2' (a geon in the opposite field, again slightly to the left or right) (see Fig. 3(a)). Similarly, targets of the left–right of fixation geon configuration (Targets B and D) were primed either by a geon occurring in the 'same position', 'position 1' (a geon slightly above or below its position in the target) or 'position 2' (a geon in the opposite field, again slightly above or below). The actual displacement of the single geon primes from fixation was minimised to avoid the need for saccades. The geons were vertically or horizontally displaced by a maximum of 1° visual angle from the normal part-position, so that the centre of the geon was in line with the 45° diagonal relative to fixation. Thus, the geons were displaced by a maximum of 1° visual angle from their normal target location in the 'position 1' priming condition, and by a maximum of 2° visual angle from their normal target location in the 'position 2' priming condition.

Each of the three blocks of Experiment 2 consisted of 72 randomly presented trials (12 trials per condition, i.e., 'same geon, same position'; 'same geon, position 1'; 'same geon, position 2'; 'different geon, same position'; or 'different geon, position 1'; 'different geon, position 2').

4.2. Results and discussion

Incorrect trials were again excluded from the RT analyses and outliers (±2.5 SDs from mean) were removed from each participant's individual data. This resulted in the removal of an average of only 2.21% of trials per participant.

A 3-way ANOVA with geon (same, different), position (same, position 1, position 2) and intensity (low, moderate, max) as factors was conducted on the RT data. Neither the main effect of position, $F < 1$, nor any interactions involving position reached significance. Fig. 3(b) and (c) plots the mean RT for each prime intensity level for each prime position condition for both the same and different geon primes respectively. The significant main effect of geon (same, different),

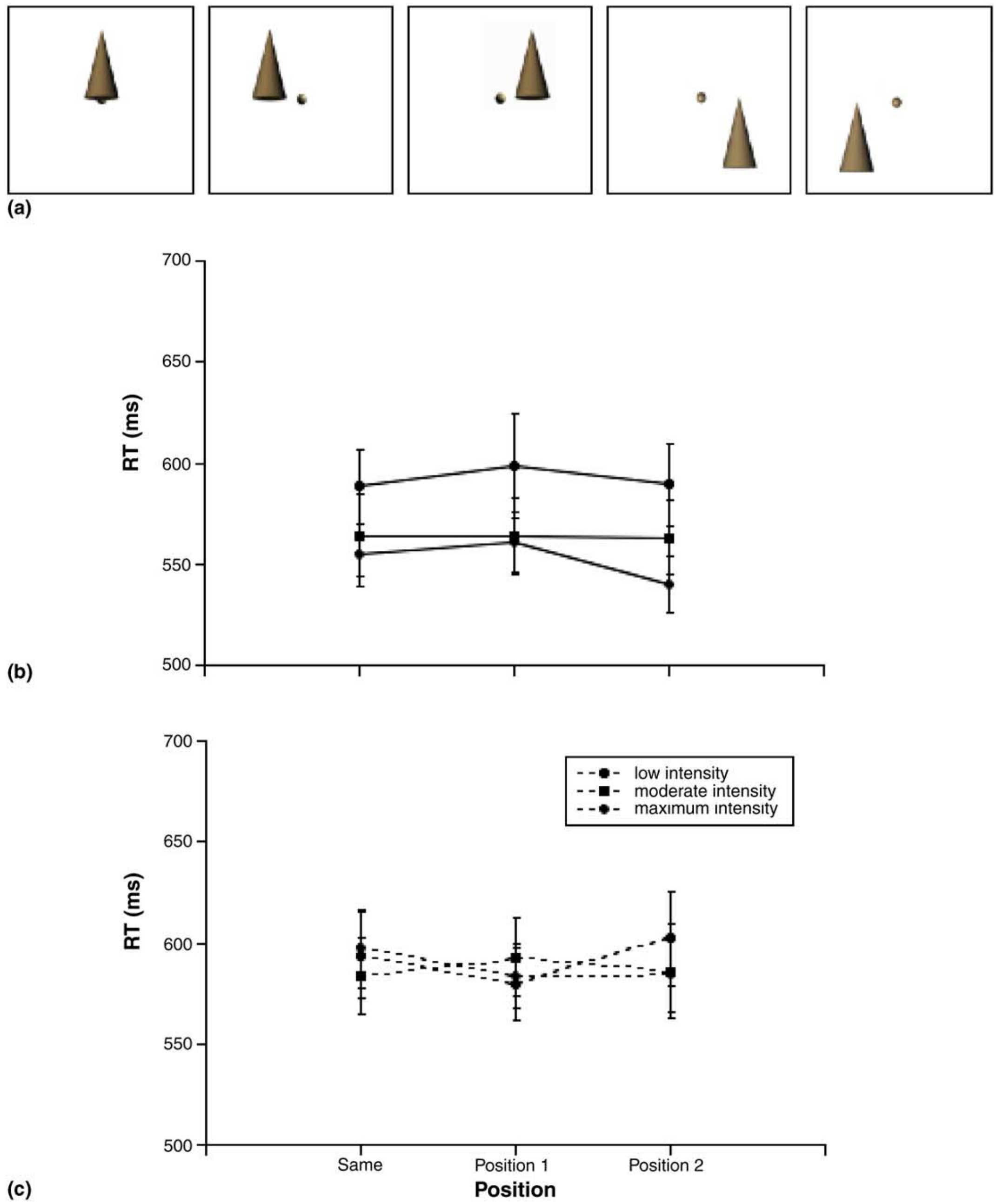

Fig. 3. (a) An illustration (using the cone of Target A as an example) of the single geon prime position conditions (same position, position 1 left, position 1 right and position 2 left and right) of Experiment 2. The fixation dot only serves to illustrate the relative location of the single geon primes and was not present during the prime displays. The plots show the mean RT (ms) for the (b) same and (c) different geon prime conditions of Experiment 2. The data are plotted for each prime intensity level across each prime position condition (same, position 1, and position 2). Error bars are standard error of the mean.

$F(1, 14) = 10.99$, $p < 0.006$, was indicative of a generalised RT facilitation when the prime's geon was the *same* as one of the target geons (mean = 20 ms facilitation). The only other effect to reach significance was the geon by intensity interaction, $F(2, 28) = 9.07$, $p < 0.002$. Post hoc, Newman–Keuls tests were conducted on the RTs across the geon and intensity factors. For the 'same geon' condition, RTs were significantly faster to the maximum intensity (mean = 552 ms) than the low intensity (mean = 592 ms, $p < 0.001$) and were faster to the moderate (mean = 564 ms) than the low intensity ($p < 0.001$), but there was no difference between the low and moderate intensity conditions ($p = 0.067$). There was no effect of intensity in the different geon condition.

An additional 3-way ANOVA with the same factors as above was conducted on the percentage error data. The only effect to even approach significance was the main effect of geon, $F(1,14) = 3.79$, $p = 0.07$. These data suggest that the RT effects (see above) were not due to a speed-accuracy trade-off.

The absence of any effects of prime position is in accordance with the structural description models of object recognition, which predicted that the priming effects should be translation invariant. It could be argued, however, that the manipulation of prime position (maximum of 2° visual angle) was not large enough to produce a noticeable effect on RT. Another explanation can be offered by analogy to the results of Dill and Edelman (2001), mentioned earlier. Assuming that detectors for local features in CoF are replicated across a number of locations in the visual field, translation invariance for such features can be acquired via interpolation (Edelman & Intrator, 2003), resulting in little or no effect of position. In comparison, the processing of *configurations* of local features (e.g., F1 *above* F2) will depend on position, because in the CoF model relations are derived (in contrast to locations, which are primitive): F1 *above* F2 would have to be represented as F1 *here* and F2 *there*—a representation that is inherently location-specific (see Edelman & Intrator, 2003).

A critical test of this explanation (and of the CoF model from which it can be derived) would be, therefore, to repeat this experiment with composite primes, which is what we did in the next experiment.

5. Experiment 3

As in the previous experiment, we were primarily interested in the effect of the relative position of objects, or prime translation, on the magnitude of visual priming. The hypotheses were as before: structural models predict shape-related visual priming effects (i.e., RT facilitation relative to the different object condition), regardless of the position in the visual field. In contrast, image-based models predict that as the relative displacement of objects increases, the magnitude of priming will decrease. In this experiment, in contrast to Experiment 2, two-part whole-object primes were used, and the relative position of the prime and target displays was again varied. In addition, the primes were either identical to the subsequent target object (same object condition) or one of the three different target objects (different object condition).

5.1. Method

5.1.1. Prime stimuli

In this experiment, unlike the previous two experiments, the 2-part prime and target stimuli were presented at slightly eccentric positions relative to fixation. All stimuli were presented at the same eccentricity with respect to fixation, while (as in Experiment 2) the relative position of the prime and target was varied. The targets appeared in a fixed, predictable location (in the lower left or upper right quadrant—8 participants with the former, 7 with the latter); thus the locus of covert attention was deployed to a predictable target object location, as in Experiments 1 and 2. The primes appeared at one of three possible positions (lower or upper left or upper right quadrants; see Fig. 4(a) for an illustration of the prime positions). Therefore, the primes and targets appeared in the *same* position, a *short distance* apart ('near position' = 2.6°), or a *longer distance* apart ('far position' = 3.7°) relative to each other.

In this experiment, all the stimuli were made slightly smaller (subtending a maximum visual angle of $1.5° \times 1.5°$), so that the overall eccentricity of stimuli displayed in a given trial would not much exceed that of the previous two experiments. Furthermore, both in this and the following experiment the size of the mask was large such that all possible positions of the prime were masked and therefore could not serve as a position cue. The prime objects were either the same as (50% of trials) or different from (50% of trials) the target object.

5.1.2. Procedure

It was imperative in this experiment for participants to maintain central fixation, as we were primarily interested in manipulating the relative *retinal* location of primes and targets. As neither prime nor target objects were presented in the centre of the screen, it took practice to be able to maintain central fixation. To facilitate this, a fixation spot remained visible before the onset of the trial and also throughout the entire trial. To ensure that participants were able to effectively maintain central fixation after practice, eye movements were visually monitored by the experimenter for the first of three blocks of experimental trials. Participants moved their eyes away from fixation on an average of only 0.21 trials (0.29%) in this first block. They were then instructed to continue with the task and to try to be extremely diligent at maintaining central fixation.

Each of the three blocks of trials of Experiment 3 consisted of 72 trials (12 trials per condition, i.e., same object, same position; same object, near position; same object, far position; different object, same position; or different object, near position; different object, far position), and again each block of trials presented the primes at one of three intensity levels.

5.2. Results and discussion

Again, incorrect trials were excluded from the RT analyses, and outliers (±2.5 SDs from mean) were removed from each participant's data. This resulted in the removal of an average of only 2.97% of trials per participant. Participant 6 was excluded from the final

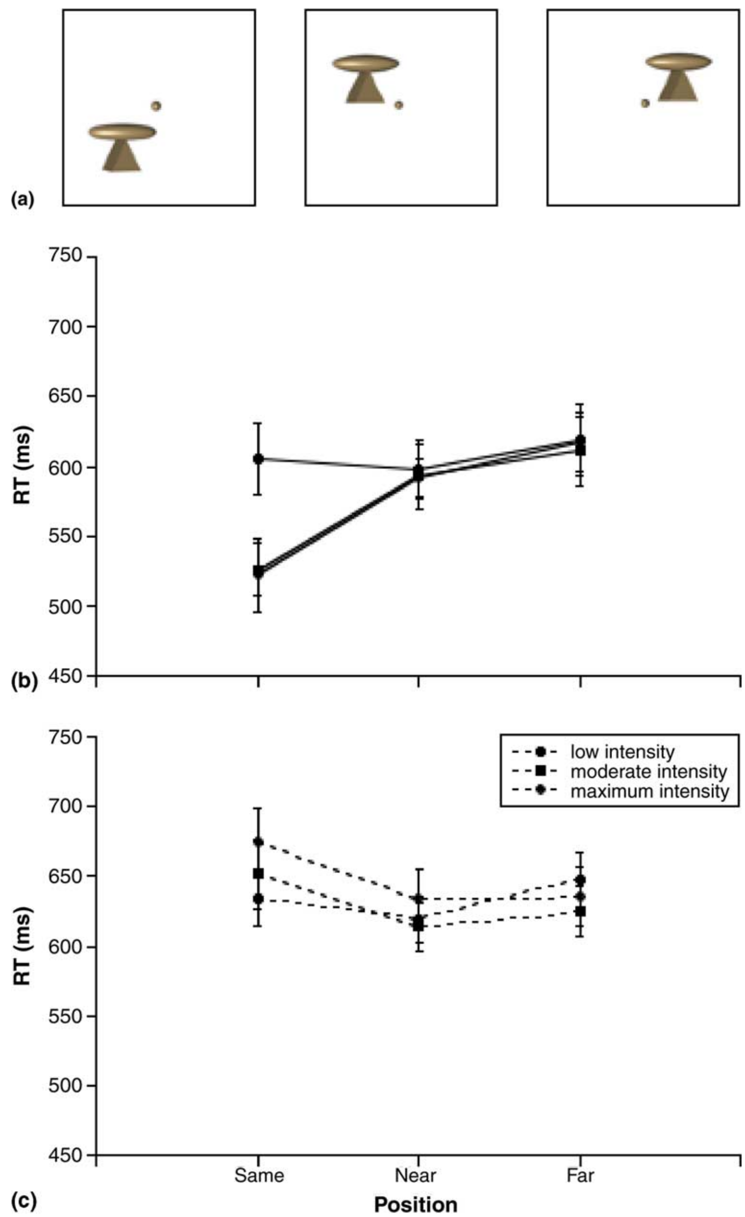

Fig. 4. (a) An illustration (using Target C as an example) of the three possible *prime positions* (relative to fixation) used in Experiments 3 and 4. The targets appeared in either the lower left or the upper right quadrant (counterbalanced between subjects) and the primes appeared at one of the three different locations. Therefore the primes and targets could appear in the same position, near or far positions relative to each other. The plots show the mean RT (ms) for the (b) same and (c) different object prime conditions of Experiment 3. The data are plotted for each prime intensity level across each prime position condition (same, near and far positions). Error bars are standard error of the mean.

sample as their overall accuracy was only 63.89%. This participant performed as if the prime object was fully predictive of the target's identity (i.e., mean accuracy for the 'different object' prime trials was only 19.44% compared to accuracy for the 'same object' trials, 93.52%). This left a final sample of 14 participants.

A 3-way ANOVA with factors object (same, different), position (same, near position, far position) and intensity (low, moderate, max) was conducted on the RT data. See Fig. 4(b) and (c) for the RT data for the same and different object primes respectively, for each prime position condition and for each intensity level. Overall, the mean RTs for the different object prime condition ($M = 638$ ms) were slower than those for the same object prime condition ($M = 587$ ms), $F(1,13) = 53.89$, $p < 0.001$. We also found a significant main effect of position ($F(2,26) = 5.43$, $p < 0.02$). Post hoc, Newman–Keuls analyses revealed that RTs to the same position were faster than to the far position ($p < 0.02$), and RTs to the near position were also faster than to the far position ($p < 0.05$). The main effect of intensity was not significant, $F < 1$.

A significant object by position, $F(2,26) = 22.83$, $p < 0.001$ interaction was found. A post hoc, Newman–Keuls analysis was conducted on the object by position interaction. For the same object condition, RTs were significantly faster to the same position than both the near ($p < 0.001$) and far positions ($p < 0.001$). Similarly, RTs to the near position were faster than to the far position ($p < 0.05$) (see Fig. 4(b)). For the different object condition, RTs were significantly slower to the same position relative to the near position only ($p < 0.01$) (see Fig. 4(c)).

A further 3-way ANOVA with factors object (same, different), position (same, near, far) and intensity (low, moderate, max) was conducted on the percentage error data. The only significant effect was the object by position interaction, $F(2,26) = 4.26$, $p < 0.03$. Newman–Keuls post hoc analyses revealed that the number of errors to the same object condition was smaller than in the different object condition for the same position only ($p < 0.01$). There were no other differences found. These data are therefore congruent with the RT data, indicating that there was no speed-accuracy trade-off.

To summarise, our manipulation of the relative position of objects significantly affected the magnitude of priming, at least for the 'same object' condition. The RT facilitation was greatest when the primes were in the same position relative to the targets, and this effect decreased as the distance between the prime and target increased. This effect interacted with the intensity of the prime in a predicted fashion—for both moderate and maximum prime intensity conditions, a robust position-dependence was shown. In contrast, the same-object RT benefit was not apparent for the low intensity primes. Moreover, the position effect was not obtained at this intensity level.

These effects of position (at sufficient levels of prime intensity), which are in line with the findings of Dill and Edelman (2001), provide evidence for hybrid image-based representation models of object recognition, such as the CoF model, as opposed to structural models that predict translation invariant priming. Still, the discrepancy remains between position-invariant priming obtained in Experiment 2 with centrally presented single-geon primes, and position-dependent priming found in Experiment 3 with eccentrically presented two-part primes. Experiment 4 was designed to seek an explanation for this discrepancy by using single-geon primes (as in Experiment 2) in eccentric positions (as in Experiment 3).

6. Experiment 4

It is not possible at this juncture to unambiguously attribute the position-dependent priming effects to the use of whole-object primes instead of single-geon primes, because the paradigms used in Experiments 2 and 3 were slightly different (see above for a brief explanation). Experiment 4 addressed this issue by using the same paradigm as Experiment 3, but with single-geon primes instead of the two-part object primes.

6.1. Method

6.1.1. Participants

Fifteen undergraduate students (mean age = 26.6 years, SD = 10.1 years) from the University of Wales, Bangor participated in the experiment either for a small payment or course credit. Three of the participants were male. Again, all participants had normal or corrected-to-normal vision.

6.1.2. Prime stimuli

In this experiment, the primes were *single geons* that could be a part of the following target object (same geon condition) or from a different target object (different geon condition). As in Experiment 3, the relative position of the prime and target was varied. The targets always appeared in a fixed, predictable position (in the lower left or upper right quadrant—8 participants in the former and 7 in the latter) and the primes appeared at one of three different positions (lower or upper left, or upper right quadrants). Therefore, regardless of the geon condition (same or different), the primes and targets appeared in the same position, a short distance apart (near position) or a longer distance apart (far position) relative to each other. Again, eye movements were visually monitored by the experimenter for the first of three blocks of experimental trials. Participants moved their eyes on an average of only 2.43 trials (3.38%) in the first block of trials.

6.2. Results and discussion

Incorrect trials were excluded from the RT analyses and outliers (±2.5 SDs from mean) were removed from each participant's data. This resulted in the removal of an average of 2.38% of trials per participant. One participant was excluded for an unusually high error rate (only 70.3% correct overall, with 79.6% correct in the same object trials and 61.0% correct for the different object trials). This left a final sample of 14 participants.

A 3-way ANOVA with factors geon (same, different), position (same, near position, far position) and intensity (low, moderate, max) was conducted on the RT data. See Fig. 5(a) and (b) for the RT data for the same and different geon primes respectively, plotted for each different prime position and each intensity level. A significant main effect of geon, $F(1,13) = 30.46$, $p < 0.001$, was found. Overall the mean RT for the different geon prime condition (mean = 682 ms) was slower than the RT for the same geon prime condition (mean = 648 ms). The

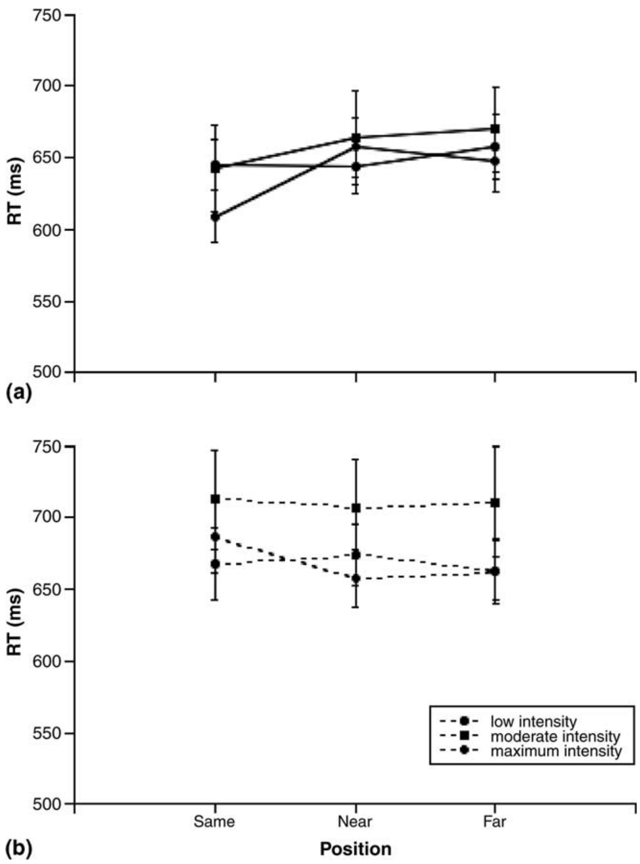

Fig. 5. Plots show the mean RT (ms) for the (a) same and (b) different geon prime conditions of Experiment 4. The data are plotted for each prime intensity level across each prime position condition (same, near and far positions). Error bars are standard error of the mean.

main effects of position, $F < 1$, and intensity, $F(2, 26) = 1.03$, $p = 0.37$, were not significant. We found a significant geon by position, $F(2, 26) = 4.20$, $p < 0.03$ interaction. Post hoc Newman–Keuls analyses found that in the same object condition, RTs to the same position were significantly faster than to both the near position ($p < 0.05$) and the far position ($p < 0.05$). There was no significant difference between the RTs to the near and far positions, though the difference itself was in the direction predicted by the image-based models. There was no advantage for any position in the different object condition.

A further 3-way ANOVA with factors geon (same, different), position (same, near position, far position) and intensity (low, moderate, max) was conducted on the percentage error data. The only effect even approaching significance was the geon by position interaction, $F(2, 26) = 2.96$, $p = 0.07$. This indicated perhaps that the same geon RT advantage was fractionally larger for the same position condition (mean difference = 3.86%), and the near position condition (mean difference = 2.31%), as compared to the far position condition (mean difference = −1.95%). These data are congruent with the RT data and indicate that there was no speed-accuracy trade-off.

To summarise the findings of Experiment 4, it appears that the manipulation of the relative position of objects did have some effect on the strength of the 'geon effect' (i.e., same-geon benefit). Target RTs were clearly facilitated by the presence of the 'same geon' when it appeared in the *same* position. This effect was reduced for the near and far positions, but it did not decrease further for the far position relative to the near position. In addition, the effect did not interact with intensity for this experiment. Although the priming effects of the present experiment clearly show some degree of position-dependence as opposed to those seen in Experiment 2 (also with single geon primes), the position-dependence is not as robust as that seen for the two-part object primes of Experiment 3.

6.2.1. Further comparisons

As Experiment 3 and Experiment 4 differed only in the types of primes used, a formal statistical comparison allowed the examination of any differential effects of single versus two-part (whole object) primes. However, an obvious limitation of such a comparison is that the first three experiments were conducted *within participants* (in a counterbalanced order) and thus Experiment 3 was not always the first experiment completed after training as it was for Experiment 4. Therefore, we decided to run a further study to compare performance across these two experiments using a within subjects design with naive participants. Twelve undergraduate students from Trinity College Dublin (four female and eight male) took part in Experiments 3 and 4 for research credits. The average age of the participants was 25.5 years. The order of the experiments was counterbalanced across participants. In all other ways the methodology was identical to that reported in Experiments 3 and 4.

To compare the RT data for the two eccentric priming experiments, a 4-way within subjects ANOVA with factors Experiment (Experiment 3, Experiment 4), object (same, different), position (same, near position, far position) and intensity (low, moderate, maximum) was conducted. There were no main effects of Experiment ($F < 1$), of position ($F < 1$) or of intensity ($F < 1$). There was a main effect of object, $F(1, 11) = 12.63$, $p < 0.01$, with longer response times for the different- than the same-object trials.

The factor 'Experiment' did not interact with any of the other factors indicating that similar effects were found across both experiments. We found an interaction between object and position, $F(2, 22) = 7.42$, $p < 0.005$. This interaction suggests, and further post hoc comparisons confirmed this, that the object effect (i.e., the advantage when the prime was either identical to the target as in Experiment 3, or the prime contained one of the geons of the target as in Experiment 4) decreased with a larger displacement between the prime and the target.

This additional study therefore supports the conclusions drawn from the original Experiments 3 and 4. Both single-geon and whole-object primes can provide RT facilitation to two-part target objects, and this facilitation is position-dependent. In both experiments, target RTs were facilitated by the presence of a same object, or the same part of an object, when it appeared in the same position. This position-dependent priming effect supports the existence of a strong image-based component in the representation and processing of visual objects.

7. General discussion

Experiments 1–4 investigated the effects of the relative position of parts and the relative position of objects on the magnitude of (short-term) priming produced by 3D, single-part primes or composite-object primes. The four targets remained constant across the experiments, each consisting of two unique parts or geons.

Experiment 1 investigated the effect of changing the structural configuration of the prime objects, or relative position of parts, on the magnitude of priming. We found that as the level of part displacement increased, the magnitude of same object priming decreased.

Experiments 2–4 all involved the manipulation of the relative position of objects and were primarily concerned with the effect on priming of translation within the visual field. Experiment 2 showed that priming with single geons does occur, and that the magnitude of the priming effect, at least in the given situation, is independent of position in the visual field. Experiment 3, however, showed that when the prime is a whole (two-part) object (same as target configuration), the magnitude of priming *is* affected by the relative position of objects. The largest priming effects were seen when the prime was presented in the *same* position as the subsequent target, with the magnitude of the effect decreasing as the distance between the prime and target increased.

Finally, Experiment 4 showed that single geon primes have a *tendency* to produce a similar position-dependent pattern of results, provided that the same paradigm as in Experiment 3 is used. Although RTs were significantly facilitated by the occurrence of a 'part' (i.e., a single geon) of the subsequent target when the prime appeared in the *same* position relative to displaced positions, the RTs were no different for the 'near' and 'far' prime position conditions. However, when we directly compared performance between same-object (Experiment 3) and same-geon primes (Experiment 4), we found no significant differences. Furthermore, for the maximum intensity primes we found significant differences between the same, near and far positions indicating that facilitation occurred when the position of either a same-object or same-geon prime was the same as the target object. A discussion of the implications of these translation-dependent priming effects follows. In general, the results afford a distinction between the two classes of theories in their most recent incarnations: the *gradual* decrease in priming with increasing structural/image changes (relative to the target) agrees better with the predictions made by recent image-based accounts of object recognition as opposed to those made by structural theories.

The present study is not the first to show position-sensitive priming effects (Bar & Biederman, 1998; Cave et al., 1994; Dill & Edelman, 2001; Dill & Fahle, 1997; McAuliffe & Knowlton, 2000). The Bar and Biederman study, for example, used a long-term subliminal paradigm (primes presented too briefly to reach the level of recognition) and showed, in a similar fashion to the present work, that changing the position of the prime reduced visual priming (Bar & Biederman, 1998). Although their study involved recognisable familiar objects, the use of additional priming conditions (e.g., same name, different image primes) allowed the experimenters to isolate visual priming from semantic or categorical priming. Another long-term priming study (McAuliffe & Knowlton, 2000) used line drawings of familiar objects, and showed reflection-sensitive priming effects only when the target was presented at the same retinotopic location as the prime. Although the translation-dependent priming effects in both those studies could be taken as evidence for an image-based object recognition approach, an alternative hypothesis is that the effects instead reflect retinotopic priming of *low-level* visual representations (since early visual representations are mapped to specific retinal positions) (McAuliffe & Knowlton, 2000). This alternative explanation, however, is difficult to accept for priming effects over the *long-term*: if this were the case, then low-level representations would have to be durable enough to facilitate responses to images presented up to 10–15 min later during the probe phase, across intervening stimulus presentations.

Structural theories suggest that information regarding the structural description of an object is represented separately from shape information. Thus part-based priming effects and structural description priming effects should be dissociable. As the prime's parts were kept constant across the priming conditions in Experiments 1 and 3, structural description priming should have only been observed for the condition in which the structural description of the prime and target matched (i.e., the 'no displacement' condition). This 'all-or-none' prediction was not supported: the amount of priming for the half-part displacement and full displacement conditions was not equivalent. Traditional image-based theories, on the other hand, make no provision for priming from parts of objects since it is assumed that objects are represented as holistic image representations (see e.g., Newell & Findlay, 1997; Tarr & Bülthoff, 1998).

If we interpret position-dependent priming as an indication that high-level object representations include information about spatial location, we are faced with a theoretical dilemma. Taking the issue of view-dependence, there is ample evidence that object recognition can be viewpoint invariant (Biederman & Gerhardstein, 1993), yet it seems that viewpoint-dependence can no longer be explained away. Viewpoint invariance and dependence need not be mutually exclusive as traditional object recognition theorists once held. Likewise, neither purely structural nor image-based object recognition models alone can account for viewpoint invariance in some situations and not others. It is not surprising, therefore, that the two contemporary models of object representation that seem to offer the best account of our results (and other similar findings in the literature)—CoF and JIM3—are both "hybrid" in the sense that they combine structural and image-based elements.

The first of these hybrid models is the Chorus of Fragments (CoF), as described earlier and in Edelman (1999); Edelman and Intrator (2000, 2003). As the receptive fields of individual 'Chorus modules' are confined to fragments of the image, each with a 'retinal address', information regarding object structure is contained in the representation, albeit expressed in a location-anchored (image-based) form. The CoF model (similar to its predecessors; Edelman & Duvdevani-Bar, 1997; Poggio & Edelman, 1990) relies on interpolation among a few stored reference views in its dealing with novel views of familiar objects, consequently, the degree of viewpoint invariance that it offers decreases with both with the novelty of the view and with the novelty of the target object's shape. Indeed, this pattern of increasing viewpoint invariance with practice has been shown (paradoxically with familiar objects) (McKone & Grenfell, 1999). Similarly, translation invariance (over and above the limited range of locations corresponding to the receptive fields of the simulated V1 complex cells within the model; cf. Riesenhuber & Poggio, 1999) can only be obtained as a result of practice, e.g., multiple fixations of the object.

Adopting a hybrid view interpolation model such as CoF has implications for interpreting the results of the present study. In our experiments, the participants had no opportunity to develop eccentrically localised representations of target stimuli. During training the targets were presented at fixation, and subsequently, even when targets were not presented at fixation (Experiments 3 and 4) they appeared at a *single and completely predictable* position. Therefore, according to the CoF model, it follows that presenting the prime at the *same* position as the target would produce larger priming effects than presenting primes at one of the other two positions. This explanation accounts for the results of both Experiment 3 that used two-part whole object primes, and Experiment 4 that used single geon primes. In both cases the strongest 'same object' (or geon) facilitation effect was found for the primes that occurred at the *same position* as the target. In addition, this effect decreased as the relative distance between the prime and target increased (when Experiment 3 and 4 were directly compared consistent effects across the experiments were found).

It is curious that while the priming effects of Experiment 2 showed complete invariance to (single-geon) prime position (relative to the target), those of Experiment 4 showed a tendency to be modified by the position of identical primes. As neither the prime nor the target stimuli differed in these two experimental conditions, it follows that the different paradigms were responsible for producing these seemingly incongruent data: The distances between the prime positions in the paradigm used for Experiments 3 and 4 were slightly larger than those used in Experiment 2. Thus, the lack of translation effects in Experiment 2 may be due to the fact that the distance between the different prime positions was not large enough.

The CoF model requires cells coarsely tuned not only to shape but also to its location in the visual field. Neurons with these functional characteristics have also been described, in areas V4 and posterior IT by Kobatake and Tanaka (1994), and in the prefrontal cortex by Rainer, Asaad, and Miller (1998), who called them "what + where" cells. The most recent quantitative data on "what + where" receptive fields in the IT cortex in the monkey were reported by Op de Beeck and Vogels (2000), who give detailed maps of the spatial distribution of responses of the shape-selective neurons, and offer an interpretation of their findings in terms of the CoF model. A close correspondence between the predictions of the CoF model and the response patterns of some cells in IT is apparent also in the preliminary findings from optical recordings reported by Tsunoda, Yamane, Nishizaki, and Tanifuji (2001). They combined Tanaka's stimulus reduction technique (Tanaka, Saito, Fukada, & Moriya, 1991) with optical imaging of cortical activity (Wang, Tanaka, & Tanifuji, 1996), and found that clusters of neurons in IT respond to "moderately complex" geometrical features, and that their responses are spatially bound to form representations of structured objects.

We should also consider the possibility that our visual system is flexible (i.e., can use either viewpoint dependent or invariant mechanisms) in order to optimise recognition performance in various conditions (Newell, 1998; Tarr & Bülthoff, 1995). Hummel's recently developed hybrid model, JIM.3, is based on this notion. JIM.3 contains two parallel processing streams which deal with object structure in a somewhat complementary manner. The first involves dynamic binding (e.g., by synchronous firing of geon and spatial relation detectors) of part attributes and spatial relations, thus forming a view invariant structural description. This process is thought to require attention, as well as more processing time, than the second approach, which involves static binding of attributes to specific locations to form another (image-like) representation of the object. The static binding process is thought to be independent of attention, but has limitations including view-dependence. If we assume that our experimental conditions biased the system in favour of the static binding process, then this model could also adequately account for the pattern of performance exhibited in this study.[4] A comparative discussion of the two hybrid models, CoF and JIM.3, is beyond the scope of the present paper, but can be found in a recent review (Edelman & Intrator, 2003).

In conclusion, the four experiments described in this paper yielded an interesting pattern of results, all of which have implications for understanding the nature of structure and shape representation in the human visual system. Our findings appear to speak against the notions of exclusively categorical representation of spatial relations and of holistic image-based representations. In contrast, they are more compatible with the twin ideas of graded, coarse coding of spatial relations and image-based, location-anchored representation of shape components (fragments). These ideas are at present the focus of converging theoretical approaches, exemplified by the CoF and JIM.3 models. An experimental distinction between these models and a direct replication of our results in simulated experiments with the CoF model serving as the subject are the next items on our agenda.

Acknowledgments

This research was supported by an Economic and Social Research Council (ESRC) Grant (Ref. no. R000222913) to Fiona N. Newell, Shimon Edelman and Kimron L. Shapiro.

References

Bar, M., & Biederman, I. (1998). Subliminal visual priming. *Psychological Science, 9*(6), 464–469.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–117.

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Journal of Cognitive Psychology, 23*(3), 393–419.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception & Performance, 18*(1), 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 & Performance, 19*(6), 1162–1182.

Bienenstock, E., & Geman, S. (1995). Compositionality in neural systems. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks*. MIT Press.

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

Cave, K. R., Pinker, S., Giorgi, L., Thomas, C. E., Heller, L. M., Wolfe, J. M., et al. (1994). The representation of location in visual images. *Cognitive Psychology, 26*, 1–32.

Dill, M., & Edelman, S. Y. (2001). Imperfect invariance to object translation in the discrimination of complex shapes. *Perception, 30*, 707–724.

Dill, M., & Fahle, M. (1997). Limited translation invariance of human visual pattern recognition. *Perception and Psychophysics, 60*, 65–81.

Edelman, S. (1998). Representation is representation of similarities. *Behavioural & Brain Sciences, 21*(4), 449–498.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA, USA: MIT Press.

Edelman, S., & Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences, 352*(1358), 1191–1202.

Edelman, S., & Intrator, N. (2000). (Coarse coding of shape fragments) + (Retinotopy)^a Representation of structure. *Spatial Vision, 13*, 255–264.

Edelman, S., & Intrator, N. (2001). A productive, systematic framework for the representation of visual structure. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.). *Advances in neural information processing systems* (vol. 13, pp. 10–16). MIT Press.

Edelman, S., & Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science, 27*, 73–109.

Fiser, J., & Biederman, I. (1995). Size invariance in visual object priming of gray-scale images. *Perception, 24*(7), 741–748.

Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Research, 41*(2), 221–234.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in human shape perception. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 157–185). Mahwah, NJ: Lawrence Erlbaum Associates.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition, 8*, 489–517.

Hummel, J. E., & Biederman, I. (1990). Dynamic binding: A basis for the representation of shape by neural networks. Paper presented at the 12th Annual Conference of the Cognitive Science Society.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*(3), 480–517.

Jacobs, A. M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception and Psychophysics, 57*(8), 1101–1110.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71*, 856–867.

[4] Note that the interpretation of our results in terms of the CoF model does not depend on specific assumptions regarding the distribution of covert attention. The possible interactions between CoF-like representations and attentional processes, which certainly deserve a careful examination, are outside the scope of the present study.

Lowe, D. G., & Binford, T. O. (1985). The recovery of three-dimensional structure from image curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 7*(3), 320–326.

McAuliffe, S. P., & Knowlton, B. J. (2000). Long-term retinotopic priming in object identification. *Perception and Psychophysics, 62*(5), 953–959.

McKone, E., & Grenfell, T. (1999). Orientation invariance in naming rotated objects: Individual difference and repetition priming. *Perception & Psychophysics, 61*(8), 1590–1603.

Newell, F. N. (1998). Stimulus context and view dependence in object recognition. *Perception, 27*(1), 47–68.

Newell, F. N., & Findlay, J. M. (1997). The effect of depth rotation on object identification. *Perception, 26*, 1231–1257.

Ochsner, K. N., Chui, C.-Y. P., & Schacter, D. L. (1994). Varieties of priming. *Current Opinion in Neurobiology, 4*, 189–194.

Op de Beeck, H., & Vogels, R. (2000). Spatial sensitivity of (Macaque) inferior temporal neurons. *Journal of Comparative Neurology, 426*, 505–518.

Poggio, T., & Edelman, S. (1990). A network that learns to recognise three-dimensional objects. *Nature [letter], 343*, 263–266.

Rainer, G., Asaad, W., & Miller, E. K. (1998). Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Science, 95*, 15008–15013.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821–824.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.

Sanocki, T. (1999). Constructing structural descriptions. *Visual Cognition, 6*, 299–318.

Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology: Human Perception & Performance, 24*(3), 732–744.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66*, 170–189.

Tarr, M. J. (1995). Rotating objects to recognise them: A case study on the role of viewpoint dependency in the recognition of three dimensional objects. *Psychonomic Bulletin and Review, 2*(1), 55–82.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception & Performance, 21*(6), 1494–1505.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. In M. J. Tarr & H. H. Bülthoff (Eds.), *Object recognition in man, monkey and machine* (pp. 1–20). Amsterdam, The Netherlands: MIT/Elsevier Science.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by a combination of feature columns. *Nature Neuroscience, 4*, 832–838.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*, 301–306.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. MIT Press.

Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science, 272*, 1665–1667.