How the forest interacts with the trees: Multiscale shape integration explains global and local processing

Georgin Jacob

Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore

S. P. Arun

Centre for Neuroscience, Indian Institute of Science, Bangalore

Hierarchical stimuli have been widely used to study global and local processing. Two classic phenomena have been observed using these stimuli: the global advantage effect (we identify the global shape faster) and an interference effect (we identify shape slower when the global and local shapes are different). Because these phenomena have been observed during shape categorization tasks, it is unclear whether they reflect the categorical judgment or the underlying shape representation. Understanding the underlying shape representation is also critical because both global and local processing are modulated by stimulus properties.

We performed two experiments to investigate these issues. In Experiment 1, we show that these phenomena can be observed in a same-different task, and that participants show systematic variation in response times across image pairs. We show that the response times to any pair of images can be accurately predicted using two factors: their dissimilarity and their distinctiveness relative to other images. In Experiment 2, we show that these phenomena can also be observed in a visual search task where participant did not have to make any categorical shape judgments. Here too, participants showed highly systematic variations in response time that could be explained as a linear sum of shape comparisons across global and local scales. Finally, the dissimilarity and distinctiveness factors estimated from the same-different task were systematically related to the search dissimilarities observed during visual search.

In sum, our results show that global and local processing phenomena are properties of a systematic shape representation governed by simple rules.

Introduction

Visual objects contain features at multiple spatial scales (Oliva & Schyns, 1997; Morrison & Schyns, 2001; Ullman, Vidal-Naquet, & Sali, 2002). Our perception of global and local shape has been extensively investigated using hierarchical stimuli, which contain local elements arranged to form a global shape (Figure 1). Two classic phenomena have been observed using these stimuli (Navon, 1977; Kimchi, 1992). First, the global shape can be detected faster than the local shape; this is known as the global advantage effect. Second, the global or local shape can be detected faster in a congruent shape (e.g. circle made of circles) than in an incongruent shape (e.g. circle made of diamonds). Although this interference effect was initially reported as stronger when reporting the local shape (Navon, 1977; Kimchi, 1992), suggesting stronger global to local interference, subsequent studies have reported equal interference in both directions (Navon & Norman, 1983; Kimchi, 1992; Poirel, Pineau, & Mellet, 2008; Sripati & Olson, 2009). Moreover, these effects depend on the size, position, spacing, and arrangement of the local shapes (Lamb & Robertson, 1990; Kimchi, 1992; Malinowski, Hübner, Keil, & Gruber, 2002; Miller & Navon, 2002).

These global/local processing phenomena have since been extensively investigated for their neural basis as well as their application to a variety of disorders. Global and local processing are thought to be localized to the right and left hemispheres respectively (Fink, Halligan, Marshall, Frith, Frackowiak, & Dolan, 1996; Han, Weaver, Murray, Kang, Yund, & Woods, 2002, Han, Jiang, & Gu, 2004), and are mediated by brain oscillations at different frequencies (Romei, Driver, Schyns, & Thut, 2011; Liu & Luo, 2019). These phenomena have now been observed in a variety of other animals, especially during tasks that require speeded responses (Tanaka & Fujita, 2000; Cavoto & Cook, 2001; Pitteri, Mongillo, Carnier, & Marinelli, 2014; Avarguès-Weber, Dyer, Ferrah, & Giurfa, 2015). Global/local processing is impaired in

Citation: Jacob, G., & Arun, S. P. (2020). How the forest interacts with the trees: Multiscale shape integration explains global and local processing. *Journal of Vision*, *20*(10):20, 1–21, https://doi.org/10.1167/jov.20.10.20.

https://doi.org/10.1167/jov.20.10.20 Received July 18, 2019; published October 27, 2020 ISSN 1534-7362

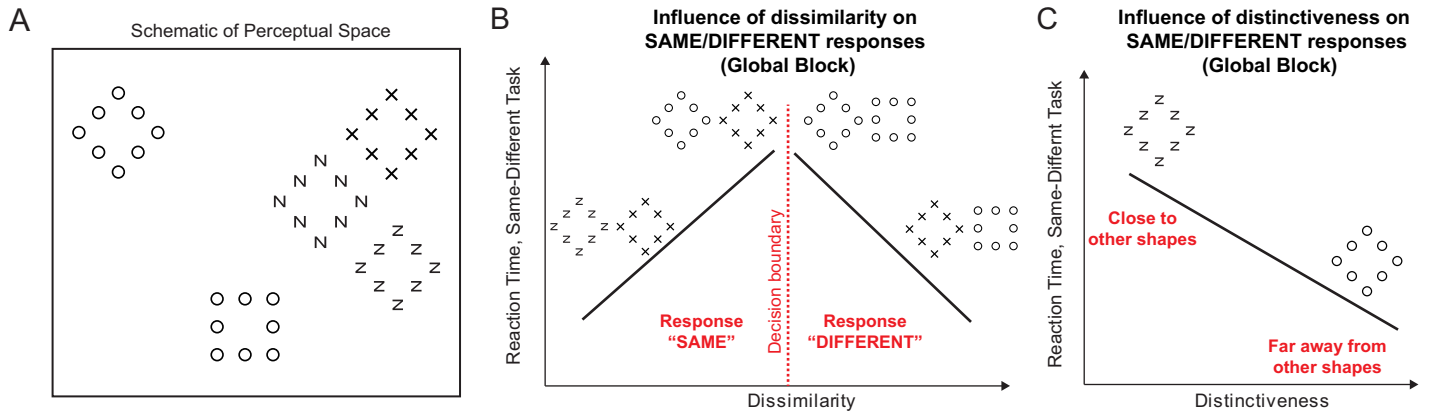

Figure 1. Quantitative models for response times in the same-different task. (**A**) To elucidate how same-different responses are related to the underlying perceptual space, consider a hypothetical perceptual space consisting of many hierarchical stimuli. In this space, nearby stimuli are perceptually similar. (**B**) We hypothesized that participants make "SAME" or "DIFFERENT" responses to an image pair based on the dissimilarity between the two images. In the global block, when two images have the same global shape, we predict that response times are longer when the two images are more dissimilar. Thus, two diamonds made using Xs and Zs evoke a faster response than two diamonds made of circles or Xs, because the latter pair is more dissimilar than the former. By contrast, when two images differ in global shape, responses are faster when they are more dissimilar. Thus, dissimilarity can either speed up or slow down responses. (**C**) We also hypothesized that shapes that are more distinct (i.e. far away from other shapes) will elicit faster responses because there are no surrounding distractors. Thus, the diamond made of circles, which is far away from all other stimuli in the schematic space of panel **A**, will elicit a faster response than a diamond made of Zs.

a variety of clinical disorders (Bihrle, Bellugi, Delis, & Marks, 1989; Robertson & Lamb, 1991; Slavin, Mattingley, Bradshaw, & Storey, 2002; Behrmann, Avidan, Leonard, Kimchi, Luna, Humphreys, & Minshew, 2006; Song & Hakoda, 2015), including those related to reading (Lachmann & Van Leeuwen, 2008; Franceschini, Bertoni, Gianesini, Gori, & Facoetti, 2017). Finally, individual differences in global/local processing predict other aspects of object perception (Gerlach & Poirel, 2018; Gerlach & Starrfelt, 2018).

Despite these insights, we lack a deeper understanding of these phenomena for several reasons. First, they have only been observed during shape detection tasks, which involve two complex steps: a categorical response made over a complex underlying representation (Freedman & Miller, 2008; Mohan & Arun, 2012). It is therefore possible that these phenomena reflect the priorities of the categorical decision. Alternatively, they may reflect some intrinsic property of the underlying shape representation.

Second, these shape detection tasks, by their design, set up a response conflict for incongruent but not congruent stimuli. This is because the incongruent stimulus contains two different shapes at the global and local levels, each associated with a different response during the global and local blocks. By contrast there is no such conflict for congruent stimuli where the global and local shapes are identical. Thus, the interference effect might reflect the response conflicts associated with making opposite responses in the global and local blocks (Miller & Navon, 2002). Alternatively, again, it might reflect some intrinsic property of the underlying shape representation, such as the congruence between the global and local shape. If shape congruence is indeed encoded in the underlying shape representation, it is not clear how it is encoded since we do not know how shapes combine across hierarchical levels.

Third, it has long been appreciated that these phenomena depend on stimulus properties, such as the size, position, spacing, and arrangement of the local elements (Lamb & Robertson, 1990; Kimchi, 1992; Malinowski et al., 2002; Miller & Navon, 2002). Surprisingly, hierarchical stimuli themselves have never been studied from the perspective of feature integration (i.e. how the global and local shapes combine). A deeper understanding of how hierarchical stimuli are organized in perception can elucidate how these stimulus properties affect global/local processing.

In summary, understanding the global advantage and incongruence effects will require reproducing them in simpler tasks, as well as understanding how global and local shape combine in the perception of hierarchical stimuli. This is not only a fundamental question but has clinical significance because deficits in global/local processing have been reported in a variety of disorders.

Overview of this study

Here, we addressed the above limitations as follows. In Experiment 1, we devised a simpler shape detection task, which involves participants indicating whether two shapes are the same or different at either the global or local level. This avoids any effects due to specific

shapes but still involves categorization, albeit a more general one. It also avoids response conflict because we can compare trials with either congruent or incongruent shapes, which elicited a SAME response. Although global advantage has been previously observed in a same-different task (Kimchi, 1988; Kimchi et al., 2005), these studies have not investigated interference effects. More generally, no previous study has attempted to explain responses in a same-different task at this fine-grained image-by-image level.

In Experiment 2, we devised a visual search task in which participants had to report the location of an oddball target. This task avoids any categorical judgment and the accompanying response conflicts. It also does not involve any explicit manipulation of global versus local attention unlike the global/local processing tasks. If global advantage and interference are present in visual search, it would imply that they reflect properties of the underlying shape representation of hierarchical stimuli. If not, they must arise from the categorization process. Although previous studies have shown a global advantage in that searching for a target differing in global shape is easier than searching for a target differing in local shape (Kimchi, 1998; Kimchi, Hadad, Behrmann, & Palmer, 2005), they have not investigated with a large set of shapes using a common task. More generally, previous studies have not investigated how global and local shapes combine in visual search.

For both Experiments 1 and 2, we devised quantitative models to explain systematic variation in the response times, as detailed below.

Quantitative model for the same-different task (Experiment 1)

We hypothesized that response times in the same-different task might be driven by two possible factors: dissimilarity and distinctiveness, as illustrated in Figure 1. Consider shapes in perceptual space as depicted in Figure 1A – nearby items in this space indicate perceptually similar items. In the global block of the same-different task, participants have to indicate whether two shapes are the same or different at the global level. We hypothesized that two potential factors that could influence response times: dissimilarity and distinctiveness. First, we reasoned that participants will find it easier to make a SAME response if the two shapes are similar, whereas they will find it harder to make a DIFFERENT response if the two shapes are similar (Figure 1B). Thus, the dissimilarity between the two images will have opposite effects on the response. Second, a shape that stands distinct from other shapes will experience less interference from other shapes and therefore elicit faster responses. In the schematic shown in Figure 1A, the diamond made of circles is more distinctive compared with the diamond made of Zs – and therefore is shown as eliciting faster responses (Figure 1C). We denote this factor as distinctiveness. We show that response times across image pairs can be accurately predicted using these two factors.

Quantitative models for shape integration in visual search (Experiment 2)

In Experiment 2, we asked how search difficulty for a target differing in both global and local shape from the distractors can be understood in terms of global and local shape differences. Search reaction time (RT) is the natural measurement during any search task and it is proportional to the similarity between the target and distractor (Duncan & Humphreys, 1989; Wolfe, Cave, & Franzel, 1989; Vincent, 2011; Alexander & Zelinsky, 2012). Recently, we have shown that reciprocal of reaction time (1/RT) is the more useful measure for understanding visual search (Arun, 2012; Pramod & Arun, 2014). The reciprocal of search time can be thought of as the dissimilarity between the target and distractors in visual search, and has the intuitive interpretation as the underlying salience signal that accumulates to threshold (Arun, 2012). Models based on 1/RT consistently outperform models based directly on search time (Vighneshvel & Arun, 2013; Pramod & Arun, 2014; Pramod & Arun, 2016; Sunder & Arun, 2016). We therefore asked whether the net dissimilarity (1/RT) between a target differing in global and local shape from the distractors can be explained as a linear sum of multiscale comparisons between and within shapes. This approach has proved effective in our previous studies involving multiple object attributes (Pramod & Arun, 2014; Pramod & Arun, 2016). We show that visual search performance can be accurately predicted using this simple model.

Experiment 1: Same-different task

In Experiment 1, participants had to indicate (in separate blocks) whether a given pair of shapes are same or different at the global or local levels. Of particular interest to us were two questions: (1) Are the classic global advantage and interference effects present in this more general same-different task? (2) Do responses across image pairs vary systematically and can they be predicted using dissimilarity and distinctiveness?

Methods

Here and in all subsequent experiments, participants had normal or corrected-to-normal vision and gave written informed consent to an experimental protocol approved by the Institutional Human Ethics Committee

of the Indian Institute of Science, Bangalore, India. Participants were naive to the purpose of the experiment and received monetary compensation for their participation.

Participants. There were 16 human participants (11 men, aged 20–30 years) in this experiment. We chose this number of participants based on our previous studies of object categorization in which this sample size yielded consistent responses (Mohan & Arun, 2012).

Stimuli. We created hierarchical stimuli by placing eight local shapes uniformly along the perimeter of a global shape. All local shapes had the same area (0.77 squared degrees of visual angle), and all global shapes occupied an area that was 25 times larger. We used seven distinct shapes at the global and local levels to create 49 hierarchical stimuli (all stimuli can be seen in Supplementary Section S5). Stimuli were shown as white against a black background.

Procedure. Participants were seated approximately 60 cm from a computer monitor under the control of custom programs written in MATLAB with routines from PsychToolbox (Brainard, 1997). Participants performed two blocks of the same-different task, corresponding to global or local shape matching. In both blocks, a pair of hierarchical shapes were shown to the participant and the participant had to respond if the shapes contained the same or different shape at a particular global/local level (key "Z" for same, and "M" for different). Each block started with a practice block with eight trials involving hierarchical stimuli made of shapes that were not used in the main experiment. Participants were given feedback after each trial during the practice block.

In all blocks, each trial started with a red fixation cross (measuring 0.6 degrees by 0.6 degrees) presented at the center of the screen for 750 ms. This was followed by two hierarchical stimuli (with local elements measuring 0.6 degrees along the longer dimension and longest dimension of global shapes are 3.8 degrees) presented on either side of the fixation cross, separated by 8 degrees from center to center. The position of each stimulus was jittered by ± 0.8 degrees uniformly at random along the horizontal and vertical. These two stimuli were shown for 200 ms followed by a blank screen until the participant made a response, or until 5 seconds, whichever was sooner.

Stimulus pairs. To avoid any response bias, we selected stimulus pairs in each block such that the proportion of same- and different-responses were equal. Each block consisted of 588 stimulus pairs. These pairs were divided equally into four groups of 147 pairs (Figure 2A): (1) pairs with both global and local shape different (GDLD); (2) pairs with same global shape but different local shape (GSLD); (3) pairs with different global shape but same local shape (GDLS), and (4) pairs with same global and local shape (GSLS; i.e. identical shapes) Because there were different numbers of total possible pairs in each category we selected pairs as follows: for GSLS pairs, there are 49 unique stimuli and therefore 49 pairs, so we repeated each pair three times to obtain 147 pairs. For GSLD and GDLS pairs, there are 147 unique pairs, so each pair was used exactly once. For GDLD pairs, there are 882 possible pairs, so we selected 147 pairs that consisted of 21 congruent pairs (i.e. each stimulus containing identical global and local shapes), 21 incongruent pairs (in which global shape of one stimulus was the local shape of the other, and vice-versa), and 105 randomly chosen other pairs. The full set of 588 stimulus pairs were fixed across all participants. Each stimulus pair was shown twice. Thus, each block consisted of $588 \times 2 = 1176$ trials. Error trials were repeated after a random number of other trials.

Participants were highly accurate in the task overall, but slightly more so in the global block (mean and standard deviation [SD] of accuracy across participants: 91% ± 4% in the global block; 88% ± 7% in the local block, $Z = 2.13$; $p < 0.05$, sign-rank test on participant-wise accuracy in the two blocks).

We removed inordinately long or short response times for each image pair using an automatic outlier detection procedure (*isoutlier* function, MATLAB 2018). We pooled the reaction times across participants for each image pair, and all response times greater than three scaled median absolute deviations away from the median were removed. In practice this procedure removed approximately 8% of the total responses.

Estimating data reliability. To estimate an upper limit on the performance of any model, we reasoned that the performance of any model cannot exceed the reliability of the data itself. To estimate the reliability of the data, we first calculated the average correlation between two halves of the data. However, doing so underestimates the true reliability because the correlation is based on two halves of the data rather than the entire dataset. To estimate this true reliability, we applied a Spearman-Brown correction on the split-half correlation. This Spearman-Brown corrected correlation (rc) is given by $rc = 2r/(1+r)$ where r is the correlation between the two halves. This data reliability is denoted as rc throughout the text to distinguish it from the standard Pearson's correlation coefficient (denoted as r).

Results

Here, participants performed a same-different task in which they reported whether a pair of hierarchical stimuli contained the same/different shape at the global level or at the local level in separate blocks. We grouped the image pairs into four distinct types based on whether

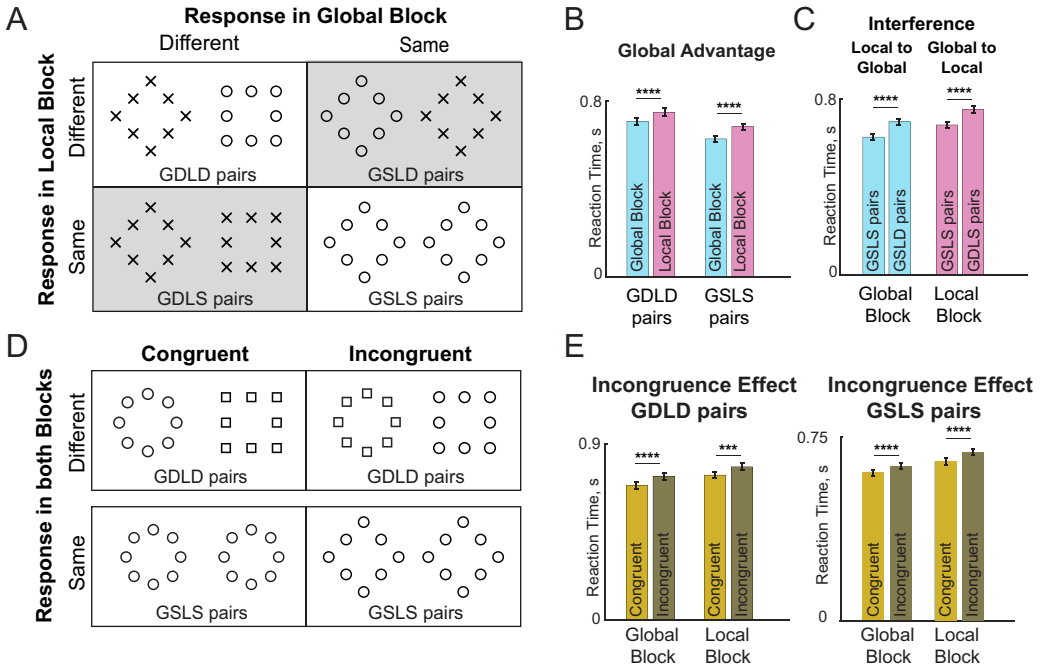

Figure 2. **Same-different task for global-local processing.** In the global block, participants had to indicate if a pair of images presented contain the same shape at the global level. Likewise, in the local block, they had to make same-different judgments about the shape at the local level. Block order was counterbalanced across participants. (**A**) Example image pairs from four image-pair types with its response in global and local block (GSLS = global same local same; GDLD = global different local different, etc.). Image pairs with identical response in both blocks are shown on a *white background* and pairs with opposite responses in the two blocks are shown on a *grey background*. (**B**) Average response times for GDLD and GSLS pairs in the global and local blocks. Error bars indicate SEM of average response times across participants. *Asterisks* indicate statistical significance of the main effect of interference in a linear mixed effects model on inverse response times (**** indicates $p < 0.00005$; see text for details). (**C**) Global-local Interference effects. *Left:* Average response times comparing GSLS ($n = 49$) and GSLD ($n = 147$) pairs in the global block, measuring how the presence of a local shape difference interferes with the SAME response. *Right:* Average response times comparing GSLS ($n = 49$) and GDLS ($n = 147$) pairs in the local block, measuring how the presence of an irrelevant global shape difference interferes with the SAME response. In both panels, *asterisks* indicate statistical significance (**** indicates $p < 0.00005$ for main effect of congruence in a linear mixed effects model on inverse response times; see text). (**D**) *Top Row:* Example congruent and incongruent image pairs which elicit the DIFFERENT response in both global and local blocks. *Bottom row:* Example congruent and incongruent pairs that elicit the SAME response in both blocks. (**E**) Average response times to congruent and incongruent stimuli in both global and local blocks for GDLD pairs (*left panel*) and GSLS pairs (*right panel*). Error bars indicate SEM across participants. *Asterisks* indicate statistical significance using linear mixed effects model on inverse reaction times (**** is $p < 0.00005$; see text).

the shapes were same/different at the global/local levels. The first type comprised pairs in which both global and local shapes were different, denoted by GDLD (see Figure 2A, top row left column). The second type comprised pairs with no difference at the global or local levels (i.e. identical images, denoted by GSLS; see Figure 2A, bottom row right column). These two types of pairs (GDLD and GSLS) elicited identical responses in the global and local blocks. The third type comprised pairs with the same global shape but different local shapes, denoted by GSLD (see Figure 2A, top row right column). The fourth type comprised pairs differing in global shape but with identical local shapes, denoted by GDLS (see Figure 2A, bottom row left column). These two types of pairs (GSLD and GDLS) elicited opposite responses in the global and local blocks. Because both global and local blocks consisted of identical image pairs from these two types, the responses in the two blocks are directly comparable and matched for image content as well as the eventual response type.

Statistical comparisons using linear mixed effects models

We set out to investigate the statistical significance of the differences in RT across blocks and conditions. For comparing GDLD pairs, we had data from 16 participants who made two responses for each of 147

image pairs, and we are interested in knowing whether their response times are systematically different between the global and local blocks.

A naïve approach might be to compare the average response times (averaged across repeats and image pairs) for each subject in the two blocks using a paired *t*-test. However, such a test ignores the complexity of the data and might hide many confounding effects: for instance, a faster response in the global block might vary by subject or vary with image pair. A more appropriate statistical test would be an analysis of variance (ANOVA), but it is based on three assumptions: independence of observations, normality of errors, homogeneous variance across all conditions, and balanced data across all conditions. Violations of these assumptions leads to incorrect estimates of effect size and their statistical significance (Glass, Peckham, & Sanders, 1972; Lix, Keselman, & Keselman, 1996). In our case, the experimental design violates the assumption of independence because the same participants performed both global and local blocks - this means any systematic variations due to participants are not independent across blocks. Second, the residuals of an ANOVA performed on RT data are not normally distributed (Supplementary Figure S1). Third, due to removal of excessively long response times, the data can become unbalanced (i.e. have unequal numbers of observations in each experimental condition). This can make the model interpretation ambiguous and ANOVA inoperable (Shaw, Mitchell-olds, & Mitchell-olds, 1993). A potential solution is to use repeated measures ANOVA, but this is typically applied on the average response times, which ignores the trial-to-trial variability present in the data and also continues to assume normally distributed residuals.

Recent statistical approaches have overcome these limitations using linear mixed effect models (Baayen, Davidson, & Bates, 2008; Lo & Andrews, 2015). To investigate these issues, we extensively compared statistical results obtained using ANOVA, repeated measures ANOVA, as well as using linear mixed effects modeling. We found that fitting the inverse response times (1/RT) to a linear mixed effects model yielded residuals that were much closer to normality compared with the same model applied to the RT data itself. This often resulted in stronger effect sizes and statistical significance as well. Our results are summarized in Supplementary Section S1.

All statistical analyses described hereafter are reported using linear mixed models applied to inverse response times. We report the partial eta-squared (η_p^2) as a measure of effect size that can be compared across experiments or studies (Richardson, 2011; Lakens, 2013). For ease of exposition, we describe only the key statistical comparisons in the main text and provide detailed descriptions in Supplementary Section S1.

Global advantage in the same-different task

To investigate the presence of a global advantage we compared global and local task responses to the GDLD and GSLS pairs, where participants made identical responses in both task blocks (Figure 2B). Participants were faster to respond in the global block in both cases (mean ± SD of RT: 707 ± 69 ms and 749 ± 89 ms in the global and local blocks for GDLD pairs; 629 ± 40 ms and 684 ± 40 ms for GSLS pairs; see Figure 2B). These effects were highly significant as evidenced by a significant main effect of block in a linear mixed effects model applied to inverse response times ($F(1,8602) = 97.76$, $p < 0.00005$ for GDLD pairs; $F(1,8647) = 413.06$, $p < 0.00005$, $\eta_p^2 = 0.046$ for GSLS pairs; see Supplementary Section S1 for details). We conclude that participants show a robust global advantage effect in the same-different task.

Global-local interference in the same-different task

Next, we asked whether interference effects were present in the same-different task. To assess local-to-global interference, we compared GSLS and GSLD pairs in the global block, where participants had to make a "SAME" response in the absence or presence of interfering local information. In the global task, participants were faster on the 49 GSLS pairs compared with 147 GSLD pairs (mean ± SD of RT: 629 ± 40 ms for GSLS; 698 ± 73 ms for GSLD pairs; Figure 2C). This difference was statistically significant, as evidenced by a main effect of interference in a linear mixed model applied to inverse response times ($F(1,8772) = 433.18$, $p < 0.000005$, $\eta_p^2 = 0.047$; see Supplementary Section S1).

To measure interference in the opposite direction (i.e. global-to-local interference), we compared response times with GSLS and GDLS pairs in the local block, because participants had to make a "SAME" response to the local level in the absence or presence of interfering global information (see Figure 2C). Again, participants were significantly faster on the GSLS pairs (mean ± SD of RT: 684 ± 40 ms for GSLS pairs, 753 ± 77 for GDLS pairs; see Figure 2C). This difference was also statistically significant ($F(1,8564) = 351.16$, $p < 0.00005$, $\eta_p^2 = 0.039$; see Supplementary Section S1).

In sum, we conclude that there is robust global-local interference in the same-different task.

Shape congruence effect in the same-different task

Previous studies have shown that participants respond faster to congruent compared with incongruent

stimuli, but these comparisons are confounded by response conflict. In other words, for an incongruent stimulus, like a circle made of diamonds, participants had to make one response for the global circle in the global block and a different one for the local diamonds in the local block. By contrast, a congruent stimulus like a circle made of circles was always associated with a single response. Thus, the slower responses in the classic studies could have been due to the associated response differences rather than the visual features of the stimulus itself.

Interestingly, this confound is absent in the same-different task. Pairs of identical congruent stimuli as well as pairs of identical incongruent stimuli belong to GSLS pairs, which elicit the SAME response in both global and local blocks. Likewise, nonidentical pairs of congruent stimuli can be compared with nonidentical pairs of incongruent stimuli, because they elicit the DIFFERENT response in both global and local blocks. These comparisons thereby are devoid of any response conflict (see Figure 2A). Thus, if we observe a slower response to incongruent stimuli in the same-different task, it can be attributed solely to the incongruence in visual features.

To test this prediction, we compared the response times for congruent-congruent and incongruent-incongruent GDLD pairs. As predicted, participants were faster for congruent stimuli in both global and local blocks (mean ± SD of RT for congruent and incongruent stimuli: 683 ± 59 ms and 736 ± 92 ms in the global block; 732 ± 89 ms and 779 ± 99 ms in the local block; Figure 2E). This difference was statistically significant as evidenced by a main effect of congruence in a linear mixed model applied to inverse reaction times ($F(1,1212) = 36.33$, $p < 0.00005$, $\eta_p^2 = 0.029$ in the global block; $F(1,1206) = 31.95$, $p < 0.00005$, $\eta_p^2 = 0.026$ in the local block; see Supplementary Section S1).

We observed similar results for GSLS pairs. Participants were faster for congruent pairs (mean ± SD of RT for congruent and incongruent stimuli: 604 ± 34 ms and 630 ± 40 ms in the global block of GSLS pairs; 648 ± 19 ms and 685 ± 40 ms in the local block of GSLS pairs; see Figure 2E). This difference was also statistically significant as assessed by a main effect of congruence in a linear mixed model applied to inverse reaction times ($F(1,4362) = 24.4$, $p < 0.0005$, $\eta_p^2 = 0.005$ in the global block; $F(1,4269) = 38.85$, $p < 0.0005$, $\eta_p^2 = 0.009$ in the local block; see Supplementary Section S1).

We conclude that participants responded faster to congruent stimuli in both global and local blocks. Thus, the global-local interference observed in previous studies can be explained by stimulus incongruence rather than response conflict.

Quantitative modeling of same-different task response times

So far, we have shown that the global advantage and incongruence effects are present in a same-different task. We next wondered whether response times vary systematically within each block across image pairs, and if these variations can be explained using quantitative models.

To establish that response times are systematic in the global and local blocks, we simply asked whether the average response times from one half of the participants across image pairs are correlated with the other half of the participants. This revealed striking correlations (see Supplementary Section S2), suggesting that there is highly systematic variation across image pairs.

We next asked whether these systematic variations can be explained using the factors described in the Introduction, namely dissimilarity and distinctiveness (see Figure 1). How do we estimate distinctiveness? We reasoned that distinctiveness might be the only influence on response time to identical image pairs, because these pairs have no variation in dissimilarity. In this case, images that elicited fast responses must be more distinctive than those that elicit slow responses. We accordingly took the reciprocal of the average response time for each GSLS pair (across trials and participants) as a measure of distinctiveness for that image. The distinctiveness for each hierarchical stimulus in the global and local blocks is shown in Figure 3A,B. It can be seen that shapes with a global circle ("O") are more distinctive in the global block than shapes containing the global shape "A." In other words, participants responded faster when they saw these shapes. It can also be seen that global distinctiveness is unrelated to local distinctiveness ($r = 0.16$, $p = 0.26$), suggesting that they are qualitatively different. Interestingly, distinctiveness estimated from GSLS pairs is correlated with both SAME and DIFFERENT response times in both blocks, and also explained the faster responses to the congruent stimuli (Supplementary Section S3).

How do we estimate dissimilarity? Unlike distinctiveness, there is no direct subset of image pairs that can be used to measure the contribution of image dissimilarity to response times, because distinctiveness correlates with all other response times (Supplementary Section S3). We therefore devised a quantitative model for the response times to estimate the underlying image dissimilarities and elucidate the contribution of dissimilarity and distinctiveness. Because high dissimilarity can increase response times for "SAME" responses and decrease response times for "DIFFERENT" responses, we devised two separate models for these two types of responses, as detailed below.

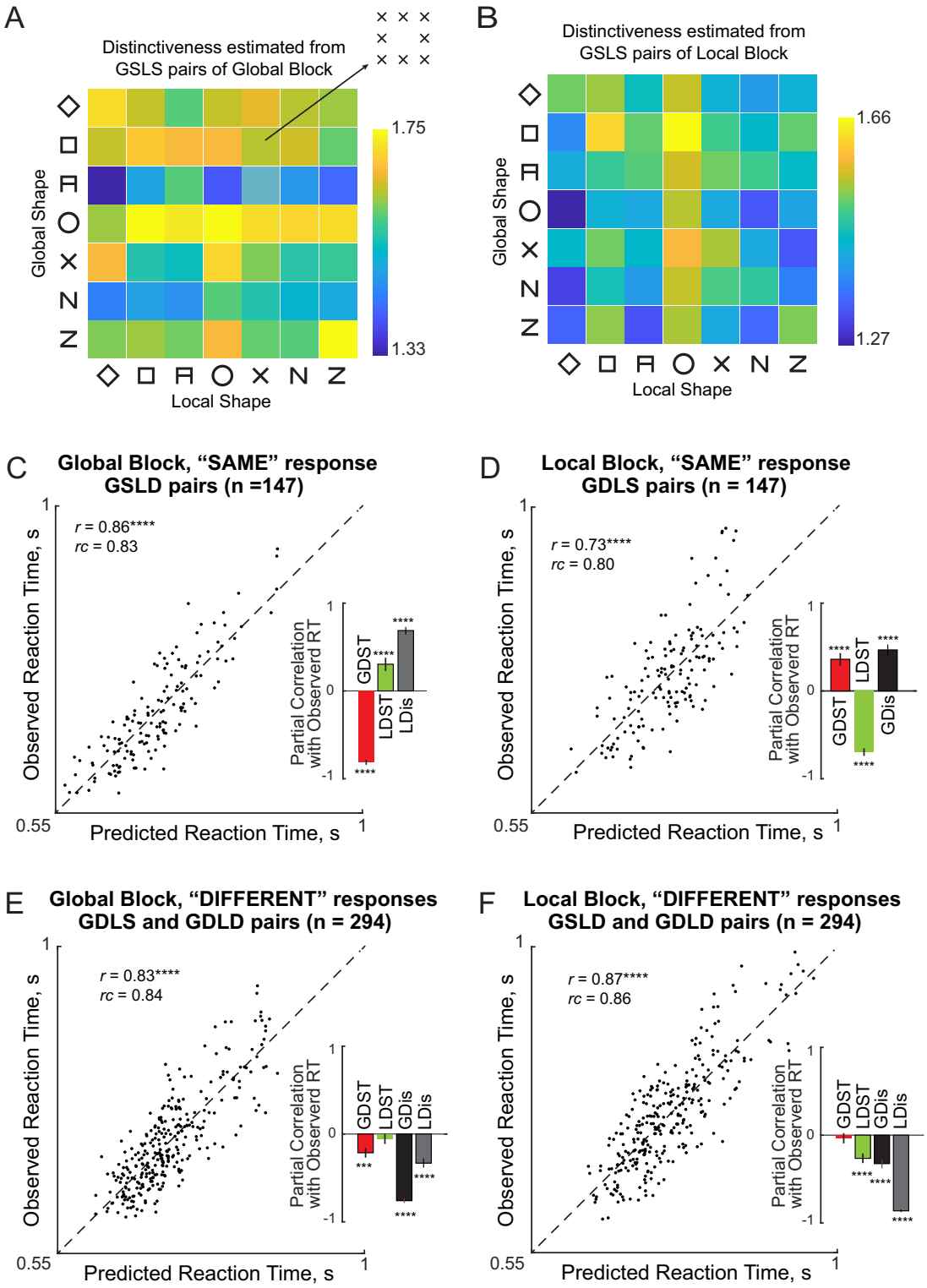

Figure 3. Quantitative model predictions for the same-different task. (**A**) Distinctiveness of each hierarchical stimulus estimated using the inverse response times for identical image pairs (GSLS) in the global block. (**B**) Distinctiveness of each hierarchical stimulus in the local block. (**C**) Observed versus predicted response times for "SAME" responses in the global block. *Inset*: Partial correlation between observed response times and each factor while regressing out all other factors (GDST and LDST = global and local distinctiveness; GDis and LDis: global and local dissimilarity). Error bars represents 68% confidence intervals, corresponding to ±1 standard deviation from the mean. (**D**) Same as **C** but for "SAME" responses in the local block. (**E**) Same as **C** but for "DIFFERENT" responses in the global block. (**F**) Same as **C** but for "DIFFERENT" responses in the local block.

Modeling "SAME" responses using distinctiveness and dissimilarity

Recall that "SAME" responses in the global block are made to image pairs in which the global shape is the same and local shape is different. Let AB denote a hierarchical stimulus made of shape A at the global level and B at the local level. We can denote any image pair eliciting a "SAME" response in global block as AB and AC, because the global shape will be identical. Then, according to our model, the response time (SRT) taken to respond to an image pair AB and AC is given by:

$$SRT\,(AB, AC) \;=\; k_G * GD \;+\; k_L * LD + \;L_{BC}$$

where GD is the sum of the global distinctiveness of AB and AC (estimated from GSLS pairs in the global block), LD is the sum of local distinctiveness of AB and AC (estimated from GSLS pairs in the local block), k_G and k_L are constants that specify the contribution of GD and LD toward the response time, and L_{BC} denotes the dissimilarity between local shapes B and C. Because there are seven possible local shapes there are only ${}^7C_2 = 21$ possible local shape terms. When this equation is written down for each GSLD pair, we get a system of linear equations of the form y = Xb where y is a 147 × 1 vector containing the GSLD response times, X is a 147 × 23 matrix containing the net global distinctiveness and net local distinctiveness as the first two columns, and 0/1 in the other columns corresponding to whether a given local shape pair is present in that image pair or not, and b is a 23 × 1 vector of unknowns containing the weights k_G, k_L, and the 21 estimated local dissimilarities. Because there are 147 equations and only 22 unknowns, we can estimate the unknown vector b using linear regression.

The performance of this model is summarized in Figure 3. The model-predicted response times were strongly correlated with the observed response times for the GSLD pairs in the global block ($r = 0.86$, $n = 147$, and $p < 0.00005$; Figure 3C). These model fits were close to the reliability of the data (rc = 0.83 ± 0.02; see Methods), suggesting that the model explained nearly all the explainable variance in the data. However, the model fits do not elucidate which factor contributes more toward response times. To do so, we performed a partial correlation analysis in which we calculated the correlation between observed response times and each factor after regressing out the contributions of the other two factors. For example, to estimate the contribution of global distinctiveness, we calculated the correlation between observed response times and global distinctiveness after regressing out the contribution of local distinctiveness and the estimated local dissimilarity values corresponding to each image pair. This revealed a significant negative correlation ($r = -0.81$, $n = 147$, and $p < 0.00005$; see Figure 3C, inset). Likewise, we obtained a significant positive partial correlation between local dissimilarities and observed response times after regressing out the other factors ($r = 0.69$, $n = 147$, and $p < 0.00005$; see Figure 3C, inset). However, local distinctiveness showed positive partial correlation ($r = 0.30$, $n = 147$, and $p < 0.0005$) suggesting that locally distinctive shapes slow down responses in the global block. Thus, response times are faster for more globally distinctive image pairs, and slower for more dissimilar image pairs.

We obtained similar results for local "SAME" responses. As before, the response time for "SAME" responses in the local block to an image pair (AB and CB) was written as:

$$SRT\,(AB, CB) = \;k_G * GD \;+\; k_L * LD + \;G_{AC}$$

where SRT is the response time, GD and LD are the net global and net local distinctiveness of the images AB and CB, respectively, k_G and k_L are unknown constants that specify the contribution of the net global and local distinctiveness, and G_{AC} is the dissimilarity between the global shapes A and C. As before, this model is applicable to all the GDLS pairs ($n = 147$), has 23 free parameters and can be solved using straightforward linear regression.

The model fits for local "SAME" responses are depicted in Figure 3D. We obtained a striking correlation between predicted and observed response times ($r = 0.72$, $n = 147$, and $p < 0.00005$; see Figure 3D). This correlation was close to the reliability of the data itself ($rc = 0.80 \pm 0.03$), suggesting that the model explains nearly all the explainable variance in the response times. To estimate the unique contribution of distinctiveness and dissimilarity, we performed a partial correlation analysis as before. We obtained a significant partial negative correlation between observed response times and local distinctiveness after regressing out global distinctiveness and global dissimilarity ($r = -0.70$, $n = 147$, and $p < 0.00005$; see Figure 3D, inset). We also obtained a significant positive partial correlation between observed response times and global dissimilarity after factoring out both distinctiveness terms ($r = 0.47$, $n = 147$, and $p < 0.00005$; see Figure 3D, inset). Finally, as before, global distinctiveness showed a positive correlation with local "SAME" responses after accounting for the other factors ($r = 0.36$, $n = 147$, and $p < 0.00005$; see Figure 3D inset).

Modeling "DIFFERENT" responses using distinctiveness and dissimilarity

We used a similar approach to predict "DIFFERENT" responses in the global and local blocks. Specifically, for any image pair AB and CD, the

	GSL	GDG	GDL	LSG	LDG	LDL
Global SAME model, L terms	1	0.54*	0.17	0.14	0.09	0.48*
Global DIFFERENT model, Global terms		1	0.24	0.34	0.30	0.47*
Global DIFFERENT model, Local terms			1	0.03	–0.08	0.14
Local SAME model, Global terms				1	0.11	–0.04
Local DIFFERENT model, Global terms					1	–0.31
Local DIFFERENT model, Local terms						1

Table 1. **Correlation between estimated dissimilarity terms within and across models.** Each entry represents the correlation coefficient between pairs of model terms. Asterisks represent statistical significance (* is $p < 0.05$). Column labels are identical to row labels but are abbreviated for ease of display.

response time according to the model is written as:

$$DRT\,(AB, CD) = k_G * GD + k_L * LD - G_{AC} - L_{BD}$$

where DRT is the response time for making a "DIFFERENT" response, GD and LD are the net global and net local distinctiveness of the images AB and CD, respectively, k_G and k_L are unknown constants that specify their contributions, G_{AC} is the dissimilarity between the global shapes A and C, and L_{BD} is the dissimilarity between the local shapes B and D. Note that, unlike the "SAME" response model, the sign of G_{AC} and L_{BD} is negative because large global or local dissimilarity should speed up "DIFFERENT" responses. The resulting model, which applies to both GDLS and GDLD pairs, consists of 44 free parameters, which are the two constants specifying the contribution of the global and local distinctiveness and 21 terms each for the pairwise dissimilarities at the global and local levels respectively. As before, this is a linear model whose free parameters can be estimated using straightforward linear regression.

The model fits for "DIFFERENT" responses in the global block are summarized in Figure 3E. We obtained a striking correlation between observed response times and predicted response times ($r = 0.82$, $n = 294$, and $p < 0.00005$; see Figure 3E). This correlation was close to the data reliability itself ($rc = 0.84 \pm 0.02$), implying that the model explained nearly all the explainable variance in the data. To estimate the unique contributions of each term, we performed a partial correlation analysis as before. We obtained a significant negative partial correlation between observed response times and global distinctiveness after regressing out all other factors ($r = –0.21$, $n = 294$, and $p < 0.0005$; see Figure 3E, inset). We also obtained a significant negative partial correlation between observed response times and both dissimilarity terms ($r = –0.76$, $n = 294$, and $p < 0.00005$ for global terms; and $r = –0.33$, $n = 294$, and $p < 0.00005$ for local terms; see Figure 3E, inset). However, we note that the contribution of global terms is larger than the contribution of local terms. As before, local distinctiveness did not contribute significantly to "DIFFERENT" responses in the global block ($r = –0.06$, $p = 0.34$, and $n = 294$; see Figure 3E, inset). We conclude that "DIFFERENT" responses in the global block are faster for globally distinctive image pairs, and for dissimilar image pairs.

We obtained similar results for "DIFFERENT" responses in the local block for GSLD and GDLD pairs. Model predictions were strongly correlated with observed response times ($r = 0.87$, $n = 294$, and $p < 0.00005$; see Figure 3F). This correlation was close to the data reliability ($rc = 0.85 \pm 0.01$) suggesting that the model explained nearly all the variance in the response times. A partial correlation analysis revealed a significant negative partial correlation for all terms except global distinctiveness (correlation between observed RT and each factor after accounting for all others: $r = –0.26$, $n = 294$, and $p < 0.00005$ for local distinctiveness; $r = –0.04$, $n = 294$, and $p = 0.55$ for global distinctiveness; $r = –0.32$, $n = 294$, and $p < 0.00005$ for global terms; and $r = –0.86$, $n = 294$, and $p < 0.00005$ for local terms; see Figure 3F). In contrast to the global block, the contribution of global terms was smaller than that of the local terms. We conclude that "DIFFERENT" responses in the local block are faster for locally distinctive image pairs and for dissimilar image pairs.

Relation between "SAME" and "DIFFERENT" model parameters

Next, we asked whether the dissimilarity terms estimated from "SAME" and "DIFFERENT" responses were related. In the global block, we obtained a significant positive correlation between the local dissimilarity terms (Table 1). Likewise, the global and local terms estimated from "DIFFERENT" responses were significantly correlated (see Table 1). In general, only 3 of 15 (20%) of all possible pairs were negatively correlated, and the median pairwise correlation across all model term pairs was significantly above zero (median correlation: 0.14, $p < 0.01$, sign-rank test). Taken together, these positive correlations imply that the dissimilarities driving the "SAME" and

"DIFFERENT" responses at both global and local levels are driven by a common underlying shape representation.

Experiment 2: Visual search

There are two main findings from Experiment 1. First, participants show a robust global advantage and an interference and incongruence effect in the same-different task. These effects could arise from the underlying categorization process or the underlying visual representation. To distinguish between these possibilities would require a task devoid of categorical judgments. To this end, we devised a visual search task in which participants have to locate an oddball target among multiple identical distractors, rather than making a categorical judgment about shape. Second, responses in the same-different task were explained using two factors: distinctiveness and dissimilarity, but it is not clear how these factors relate to the underlying visual representation.

We sought to answer four questions. First, are the global advantage and incongruence effects present in visual search? Second, can performance in the same-different task be explained in terms of the responses in the visual search task? Third, can we understand how global and local features combine in visual search? Finally, can the dissimilarity and distinctiveness terms in the same-different model of Experiment 1 be related to some aspect of the visual representations observed during visual search?

Methods

Participants. Eight right-handed participants (6 men, aged 23–30 years) participated in the study. We selected this number of participants here and in subsequent experiments based on the fact that similar sample sizes have yielded extremely consistent visual search data in our previous studies (Mohan & Arun, 2012; Vighneshvel & Arun, 2013; Pramod & Arun, 2016).

Stimuli. We used the same set of 49 stimuli as in Experiment 1, which were created by combining seven possible shapes at the global level with seven possible shapes at the local level in all possible combinations. The full stimulus set can be seen in Supplementary Section S5.

Procedure. Participants were seated approximately 60 cm from a computer. Each participant performed a baseline motor block, a practice block, and then the main visual search block. In the baseline block, on each trial, a white circle appeared on either side of the screen and participants had to indicate the side on which the circle appeared. We included this block so that participants would become familiar with the key press associated with each side of the screen, and to estimate a baseline motor response time for each participant. In the practice block, participants performed 20 correct trials of visual search involving unrelated objects to become familiarized with the main task.

Each trial of the main experiment started with a red fixation cross presented at the center of the screen for 500 ms. This was followed by a 4×4 search array measuring 24 degrees square with a spacing of 2.25 degrees between the centers of adjacent items. Images were slightly larger in size (1.2 times) compared with Experiment 1 to ensure that the local elements were clearly visible. The search array consisted of 15 identical distractors and one oddball target placed at a randomly chosen location in the grid. Participants were asked to locate the oddball target and respond with a key press ("Z" for left and "M" for right) within 10 seconds, failing which the trial was aborted and repeated later. A red vertical line was presented at the center of the screen to facilitate left/right judgments.

Search displays corresponding to each possible image pair were presented two times, with either image in a pair as target (with target position on the left in one case and on the right in the other). Thus, there were $49C_2 = 1176$ unique searches and 2352 total trials. Trials in which the participant made an error or did not respond within 10 seconds were repeated randomly later. In practice, these repeated trials were very few in number, because participants accuracy was extremely high (mean and SD accuracy: $98.4\% \pm 0.7\%$ across participants).

Model fitting

We measured the perceived dissimilarity between every pair of images by taking the reciprocal of the average search time for that pair across participants and trials. We constructed a quantitative model for this perceived dissimilarity following the part summation model developed in our previous study (Pramod & Arun, 2016). Let each hierarchical stimulus be denoted as AB where A is the shape at the global level and B is the local shape. The net dissimilarity between two hierarchical stimuli AB and CD is given by:

$$d(AB, CD) = G_{AC} + L_{BD} + X_{AD} + X_{BC} + W_{AB} + W_{CD} + \text{constant}$$

where G_{AC} is the dissimilarity between the global shapes, L_{BD} is the dissimilarity between the local shapes, X_{AD} and X_{BC} are the across-object dissimilarities between the global shape of one stimulus and the local shape of the other, and W_{AB} and W_{CD} are the dissimilarities between global and local shape within each object. Thus, there are four sets of unknown

parameters in the model, corresponding to global terms, local term, across-object terms, and within-object terms. Each set contains pairwise dissimilarities among the seven shapes used to create the stimuli. Note that model terms repeat across image pairs: for instance, the term G_{AC} is present for every image pair in which A is a global shape of one and C is the global shape of the other. Writing this equation for each of the 1176 image pairs results in a total of 1176 equations corresponding to each image pair, but with only 21 shape pairs times four types (global, local, across, and within) + 1 = 85 free parameters. The advantage of this model is that it allows each set of model terms to behave independently, thereby allowing potentially different shape representations to emerge for each type through the course of model fitting.

This simultaneous set of equations can be written as y = Xb where y is a 1176 × 1 vector of observed pairwise dissimilarities between hierarchical stimuli, X is a 1176 × 85 matrix containing 0, 1, or 2 (indicating how many times a part pair of a given type occurred in that image pair) and b is a 85 × 1 vector of unknown part-part dissimilarities of each type (corresponding, across, and within). We solved this equation using standard linear regression (*regress* function, MATLAB).

The results described in the main text, for ease of exposition, are based on fitting the model to all pairwise dissimilarities, which could result in overfitting. To assess this possibility, we fitted the model each time on 80% of the data and calculated its predictions on the held-out 20%. This too yielded a strong positive correlation across many 80-20 splits ($r = 0.85 \pm 0.01$, $p < 0.00005$ in all cases), indicating that the model is not overfitting to the data.

Results

Participants performed searches corresponding to all possible pairs of hierarchical stimuli ($^{49}C2 = 1176$ pairs). Participants were highly accurate in the task (mean ± SD accuracy: 98.4% ± 0.7% across participants).

Note that each image pair in visual search has a one-to-one correspondence with an image pair used in the same-different task. Thus, we have GDLS, GSLD, and GDLD pairs in the visual search task. However, there are no GSLS pairs in visual search because these pairs correspond to identical images, and can have no oddball search.

Global advantage effect in visual search

We set out to investigate whether there is a global advantage effect in visual search. We compared searches with target differing only in global shape (i.e. GDLS pairs) with equivalent searches in which the target differed only in local shape (i.e. GSLD pairs). Two example searches are depicted in Figure 4A,B. It can be readily seen that finding a target differing in global shape (see Figure 4A) is much easier than finding the same shape difference in local shape (see Figure 4B).

The above observation held true across all GDLS/GSLD searches. Participants were equally accurate on GDLS searches and GSLD searches (accuracy, mean ± SD: 98% ± 1% for GDLS, 98% ± 1% for GSLD, $p = 0.48$, sign-rank test across participant-wise accuracy). However, they were faster on GDLS searches compared with GSLD searches (search times, mean ± SD: 1.90 ± 0.40 seconds across 147 GDLS pairs, 2.11 ± 0.56 seconds across 147 GSLD pairs; Figure 4C). This difference was statistically significant as evidenced by a main effect of scale of change in a linear mixed effects model analysis performed on inverse RT ($F(1,4696) = 163.24$, $p < 0.00005$, $\eta_p^2 = 0.034$; for details see Supplementary Section S4). We conclude that searching for a target differing in global shape is easier than searching for a target differing in local shape. Thus, there is a robust global advantage effect in visual search.

Incongruence effect in visual search

Next, we compared whether searches involving a pair of congruent stimuli were easier than those with incongruent stimuli. Example searches of each type are shown in Figure 4D,E. Searches for pairs of congruent stimuli were faster than searches for incongruent stimulus pairs (mean ± SD of RT: 1127 ± 409 ms and 1359 ± 398 ms for congruent and incongruent pairs respectively; Figure 4F). This difference was statistically significant as assessed using a linear mixed effects model applied to inverse response times ($F(1,664) = 35.87$, $p < 0.00005$, $\eta_p^2 = 0.051$ for main effect of congruence; Supplementary Section S4).

The above incongruence effect compares image pairs with congruent target and congruent distractor with image pairs for with incongruent target and incongruent distractor, so the difference could arise due to the sharing of features within a shape or due to the juxtaposition of shared shapes. To resolve this issue, we calculated the effect of target congruence and distractor congruence separately. We calculated the effect of target incongruence by calculating the search times for either a congruent or incongruent target formed by two shapes against all possible incongruent distractors chosen from the remaining five shapes ($n = 2 \times {}^7C_2 \times 2 \times {}^5P_2 = 1680$). Target congruent searches were slightly faster than target incongruent searches (mean ± SD of RT: 1095 ± 399 ms and 1117 ± 401 ms for congruent and incongruent targets; Figure 4G, *top panel*). However, this difference was not statistically insignificant, as

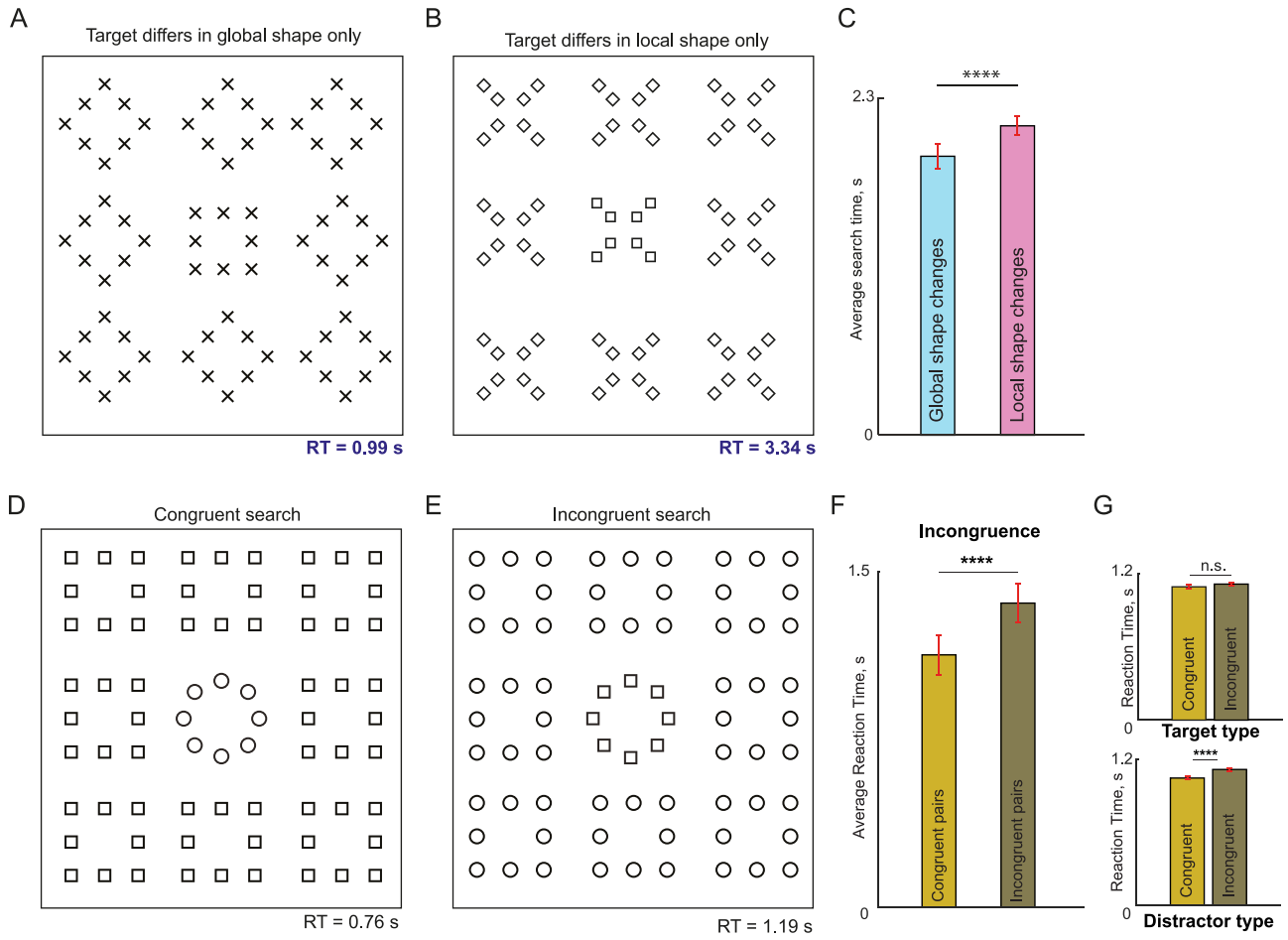

Figure 4. Global advantage and congruence in visual search (Experiment 2). (**A**) Example search array with an oddball target differing only in global shape from the distractors. The actual experiment used 4 × 4 search arrays with stimuli shown as *white* against a *black background*. (**B**) Example search array with an oddball target differing only in local shape from the distractors. (**C**) Average response times for GDLS and GSLD search pairs. Error bars represent SEM across image pairs. *Asterisks* indicate statistical significance (**** indicates $p < 0.00005$ for main effect of type of change (global/local) in a linear mixed effects model on inverse response times; see text). (**D**) Example search array for a congruent target among congruent distractor. (**E**) Example search array for an incongruent target among an incongruent distractor. (**F**) Average response times for congruent and incongruent searches ($n = 21$). Error bars represents SEM across image pairs. *Asterisks* indicate statistical significance (**** indicates $p < 0.00005$ for main effect of congruence in a linear mixed effects model on inverse response times; see text). (**G**) *Top:* Average response times for searches with congruent or incongruent targets among a fixed set of distractors whose global and local shapes are not shared with the target ($n = 21 \times 2 \times 20 = 840$). *Asterisks* indicate statistical significance (n.s. indicates the main effect of target congruence is not significant in a linear mixed effects model on inverse response times; see text). *Bottom:* Average response times for searches with congruent or incongruent distractors and a fixed set of targets whose global and local shapes are not shared with the distractors ($n = 21 \times 2 \times 20 = 840$). *Asterisks* indicates statistical significance (**** indicates $p < 0.00005$ for main effect of distractor congruence in a linear mixed effects model on inverse response times; see text).

evidenced by the lack of a main effect of congruence in a linear mixed effects model analysis performed on inverse response times ($F(1,328) = 3.73$, $p = 0.54$; see Supplementary Section S4).

Likewise, we calculated the effect of distractor incongruence by calculating the search times for searches with either a congruent or incongruent distractor and all possible incongruent targets chosen from the remaining five shapes ($n = {}^7C_2 \times 2 \times {}^5P_2 = 840$). Searches with congruent distractors were faster than incongruent distractors (mean ± SD of RT: 1047 ± 96 ms and 1117 ± 116 ms for congruent and incongruent distractors; see Figure 4G, *bottom panel*). This difference was statistically significant as evidenced by a main effect of congruence in a linear mixed effects model applied to inverse response times ($F(1,328) = 34.85$, $p < 0.00005$, $\eta_p^2 = 0.096$; see Supplementary Section S4).

Systematic variations in response times in the visual search task

Having established that participants showed a robust global advantage effect and incongruence effects, we wondered whether there were other systematic variations in their responses as well. Indeed, response times were highly systematic as evidenced by a strong correlation between two halves of the participants (split-half correlation between RT of odd- and even-numbered participants: $r = 0.83$, $n = 1176$, and $p < 0.00005$).

Previous studies have shown that the reciprocal of search time can be taken as a measure of dissimilarity between the target and distractors. We therefore took the reciprocal of the average search time across all participants (and trials) for each image pair as a measure of dissimilarity between the two stimuli. Because we performed all pairwise searches between the hierarchical stimuli, it becomes possible to visualize these stimuli in visual search space using multidimensional scaling (MDS). The multidimensional scaling plot obtained from the observed visual search data is shown in Supplementary Section S5. Two interesting patterns can be seen. First, stimuli with the same global shape clustered together, indicating that these are hard searches. Second, congruent stimuli (i.e. with the same shape at the global and local levels) were further apart compared with incongruent stimuli (with different shapes at the two levels), indicating that searches involving congruent stimuli are easier than incongruent stimuli. These observations concur with the global advantage and incongruence effect described above in visual search.

Quantitative modeling of global and local shape integration

So far, we have shown that the global advantage and incongruence effects in the same-different task also arise in the visual search task, suggesting that these effects are intrinsic to the underlying representation of these hierarchical stimuli. However, these findings do not provide any fundamental insight into the underlying representation or how it is organized. For instance, why are incongruent shapes more similar than congruent shapes? How do global and local shape combine?

To address these issues, we asked whether search for pairs of hierarchical stimuli can be explained in terms of shape differences and interactions at the global and local levels. To build a quantitative model, we drew upon our previous studies in which the dissimilarity between objects differing in multiple features was found to be accurately explained as a linear sum of part-part dissimilarities (Pramod & Arun, 2014; Pramod & Arun, 2016; Sunder & Arun, 2016). Consider a hierarchical stimulus AB, where A represents the global shape and B is the local shape. Then, according to the model (which we dub the multiscale part sum model), the dissimilarity between two hierarchical stimuli AB and CD can be written as a sum of all possible pairwise dissimilarities among the parts A, B, C, and D as follows (Figure 5A):

$$d(AB, CD) = G_{AC} + L_{BD} + X_{AD} + X_{BC} + W_{AB} + W_{CD} + \text{constant}$$

where G_{AC} is the dissimilarity between the global shapes, L_{BD} is the dissimilarity between the local shapes, X_{AD} and X_{BC} are the across-object dissimilarities between the global shape of one stimulus and the local shape of the other, and W_{AB} and W_{CD} are the dissimilarities between global and local shape within each object. Because there are seven possible global shapes, there are ${}^{7}C_{2} = 21$ pairwise global-global dissimilarities corresponding to G_{AB}, G_{AC}, G_{AD}, etc., and likewise for L, X, and W terms. Thus, in all, the model has 21 part-part relations times 4 types + 1 constant = 85 free parameters. Importantly, the multiscale part sum model allows for completely independent shape representations at the global level, local level, and even for comparisons across objects and within object. The model works because the same global part dissimilarity G_{AC} can occur in many shapes where the same pair of global shapes A and C are paired with various other local shapes.

Performance of the part sum model

To summarize, we used a multiscale part sum model that explains the dissimilarity between two hierarchical stimuli as a sum of pairwise shape comparisons across multiple scales. To evaluate model performance, we plotted the observed dissimilarities between hierarchical stimuli against the dissimilarities predicted by the part sum model (Figure 5B). This revealed a striking correlation ($r = 0.88$, $n = 1176$, and $p < 0.00005$; see Figure 5B). This high degree of fit matches the reliability of the data (mean ± SD reliability: $rc = 0.84 \pm 0.01$; see Methods).

This model also yielded several insights into the underlying representation. First, because each group of parameters in the part sum model represent pairwise part dissimilarities, we asked whether they all reflect a common underlying shape representation. To this end, we plotted the estimated part relations at the local level (L terms), the across-object global-local relations (X terms), and the within-object relations (W terms) against the global part relations (G terms). This revealed a significant correlation for all terms (correlation with global terms: $r = 0.60$, $p < 0.005$ for

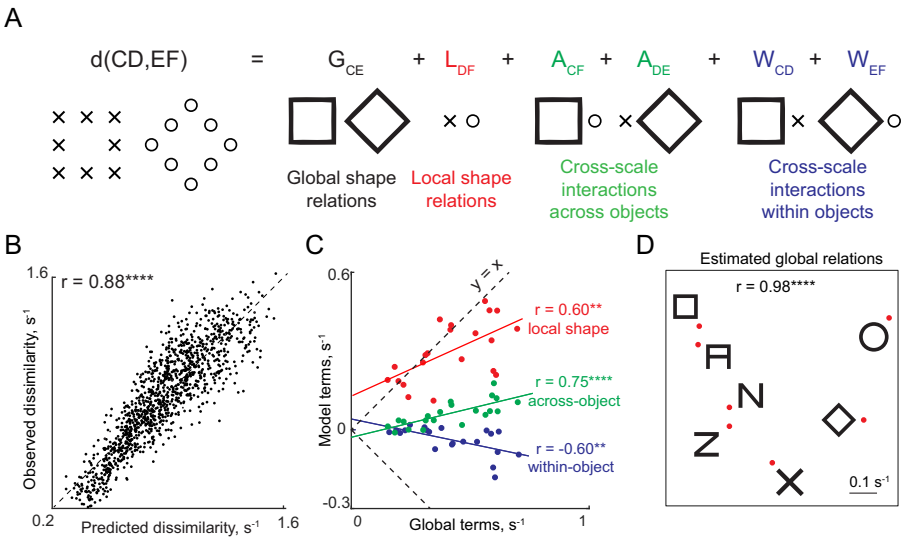

Figure 5. Global and local shape integration in hierarchical stimuli. (**A**) We investigated how global and local shape combine in visual search using the multiscale part sum model. According to the model, the dissimilarity between two hierarchical stimuli can be explained as a weighted sum of shape differences at the global level, local level, and cross-scale differences across and within objects (see text). (**B**) Observed dissimilarity plotted against predicted dissimilarity for all 1176 object pairs in the experiment. (**C**) Local and cross-scale model terms plotted against global terms. *Colored lines* indicate the corresponding best fitting line. *Asterisks* indicate statistical significance: *** is $p < 0.0005$, **** is $p < 0.00005$. (**D**) Visualization of global shape relations recovered by the multiscale model, as obtained using multidimensional scaling analysis.

L terms; $r = 0.75$, $p < 0.00005$ for X terms; $r = -0.60$, $p < 0.005$ for W terms; Figure 5C). This is consistent with the finding that hierarchical stimuli and large/small stimuli are driven by a common representation at the neural level (Sripati & Olson, 2009).

Second, cross-scale within-object (W terms) were negative (average: -0.04, $p < 0.005$, sign-rank test on 21 within-object terms). The effect of within-object dissimilarity is akin to the effect of distracter heterogeneity in visual search. Just as similar distracters make search easier, similar shapes at the global and local level within a shape make the search easier. We have made a similar observation previously with two-part objects (Pramod & Arun, 2016).

Third, we visualized this common shape representation using multidimensional scaling on the pairwise global coefficients estimated by the model. The resulting plot (Figure 5D) reveals a systematic arrangement whereby similar global shapes are nearby. Ultimately, the multiscale part sum model uses this underlying part representation determines the overall dissimilarity between hierarchical stimuli.

Model explanation for global advantage and incongruence

Having established that the full multiscale part sum model yielded excellent quantitative fits, we asked whether it can explain the global advantage and incongruence effects.

First, the global advantage effect in visual search is the finding that shapes differing in global shape are more dissimilar than shapes differing in local shape. This is explained by the multiscale part sum model by the fact that global part relations are significantly larger in magnitude compared with local terms (average magnitude across 21 pairwise terms: 0.42 ± 0.17 s^{-1} for global, 0.30 ± 0.11 s^{-1} for local, $p < 0.005$, sign-rank test).

Second, how does the multiscale part sum model explain the incongruence effect? We first confirmed that the model shows the same pattern as the observed data (Figure 6A). To this end, we examined how each model term in the model works for congruent and incongruent shapes (Figure 6B). First, note that the terms corresponding to global and local shape relations are identical for both congruent and incongruent stimuli so these cannot explain the incongruence effect. However, congruent and incongruent stimuli differ in the cross-scale interactions both across and within stimuli. For a congruent pair, which have the same shape at the global and local level, the contribution of within-object terms is zero, and the contribution of across-object terms is non-zero, resulting in an overall larger dissimilarity (see Figure 6B). In contrast, for an incongruent pair, the within-object terms are negative and across-object terms are zero, leading to a smaller overall dissimilarity.

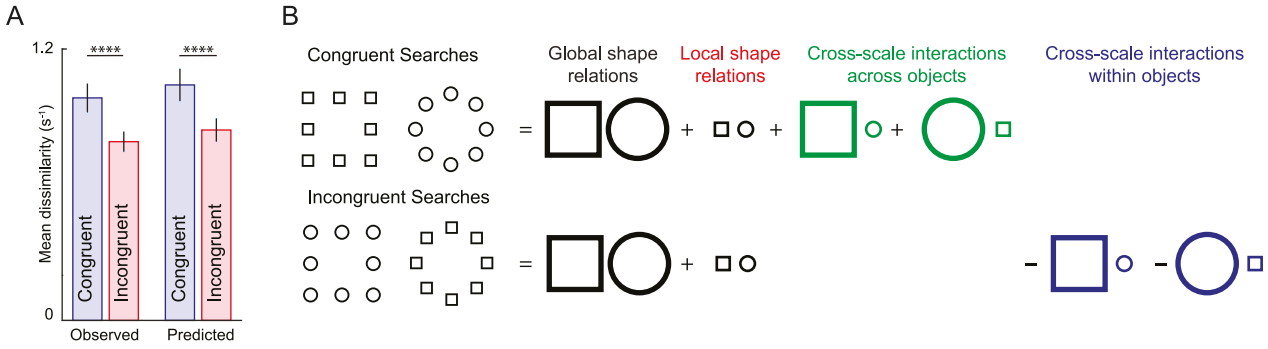

Figure 6. Incongruence effect in visual search. (**A**) Average dissimilarity for congruent and incongruent image pairs for observed dissimilarities (*left*) and dissimilarities predicted by the multiscale part sum model (*right*). Error bars indicate SD across image pairs. *Asterisks* indicate statistical significance, as calculated using an ANOVA, with conventions as before. (**B**) Schematic illustrating how the multiscale model predicts the incongruence effect. For both congruent and incongruent searches, the contribution of global and local terms in the model is identical. However, for congruent searches, the net dissimilarity is large because cross-scale across terms are non-zero and within-object terms are zero (because the same shape is present at both scales). In contrast, for incongruent searches, the net dissimilarity is small because across-object terms are zero (since the local shape of one is the global shape of the other) and within-object terms are non-zero and negative.

To investigate whether the incongruence effect observed in visual search is due to the relationship between the target and distractors, or due to the incongruency of the target itself, we fit two reduced models with either the cross-scale across terms (X) or the cross-scale within terms (W) terms removed. We reasoned that if the incongruence effect is due to the across terms, the model containing the across terms will perform better than the model containing the within terms in predicting the dissimilarities of the congruent and incongruent pairs – and likewise if it were due to the within terms. However, we obtained similar model performance with across or within terms removed (correlation across congruent and incongruent pairs, $n = 42$: $r = 0.87$ for model with across terms; $r = 0.88$ for model with within terms, $p < 0.00005$ in both cases). Thus, the incongruence effect arises from both factors.

To summarize, the multiscale model explains qualitative features of visual search, such as the global advantage and incongruence effects, and explains visual search for hierarchical stimuli using a linear sum of multiscale part differences. The excellent fits of the model indicate that shape information combines linearly across multiple scales.

Relation between same-different model and visual search

Recall that the responses in the same-different task were explained using two factors, distinctiveness and dissimilarity (see Figure 4). We wondered whether these factors are related to any aspect of the visual search representation.

We first asked whether the distinctiveness of each image as estimated from the GSLS pairs in the same-different task is related to the hierarchical stimulus representation in visual search. We accordingly calculated a measure of global distinctiveness in visual search as follows: for each image, we calculated its average dissimilarity (1/RT in visual search) to all other images with the same global shape. Likewise, we calculated local search distinctiveness as the average dissimilarity between a given image and all other images with the same local shape. We then asked how the global and local distinctiveness estimated from same-different task are related to the global and local search distinctiveness estimated from visual search.

We obtained a striking double-dissociation: global distinctiveness estimated in the same-different task was correlated only with global but not local search distinctiveness ($r = 0.55$, $p < 0.00005$ for global search distinctiveness; and $r = 0.036$, $p = 0.55$ for local search distinctiveness; Figure 7A). Likewise, local distinctiveness estimated in the same-different task was correlated only with local search distinctiveness but not global distinctiveness ($r = 0.35$, $p < 0.05$ for local search distinctiveness; and $r = 0.05$, $p = 0.76$ for global search distinctiveness; Figure 7B).

Next, we investigated whether the global and local shape dissimilarity terms estimated from the same-different task were related to the global and local terms in the part-sum model. Many of these correlations were positive and significant (see Table 2), suggesting that all dissimilarities are driven by a common shape representation.

Same-different model terms	Correlation with visual search global terms	Correlation with visual search local terms
Same-different task, Global block		
Same model Local Terms	0.47*	0.76****
Different model Global Terms	0.69****	0.82****
Different Model Local Terms	0.02	0
Same-Different task, Local Block		
Same model Local Terms	0.37	0.11
Different model Global Terms	0.38	0.21
Different Model Local terms	0.14	0.6**

Table 2. **Comparison of model parameters across tasks.** Each entry represents the correlation coefficient between model terms estimated from the same-different task and global and local terms from the visual search model. Asterisks represent statistical significance (* is $p < 0.05$, **** is $p < 0.00005$ etc.).

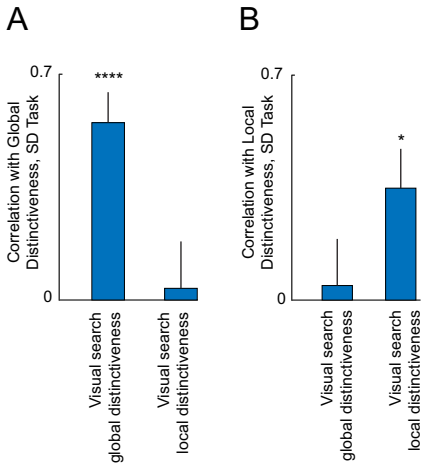

Figure 7. Relation between same-different model parameters and visual search. (**A**) Correlation between distinctiveness estimated from 49 GSLS trials in the global block of the same-different (SD) task with global and local search distinctiveness. Error bars represents 68% confidence intervals, corresponding to ±1 standard deviation from the mean. (**B**) Correlation between distinctiveness estimated from 49 GSLS trials in the local block of the same-different task with global and local search distinctiveness.

We conclude that both distinctiveness and dissimilarity terms in the same-different task are systematically related to the underlying representation in visual search.

Comparison of part-sum model with other models

The above results show that search for hierarchical stimuli is best explained using the reciprocal of search time (1/RT), or search dissimilarity. That models based on 1/RT provides a better account than RT-based models was based on our previous findings (Vighneshvel & Arun, 2013; Pramod & Arun, 2014; Pramod & Arun, 2016; Sunder & Arun, 2016). To reconfirm this finding, we fit RT and 1/RT based models to the data in this experiment. Indeed, 1/RT based models provided a better fit to the data (see Supplementary Section S6).

The above results are also based on a model in which the net dissimilarity is based on part differences at the global and local levels as well as cross-scale differences across and within object. This raises the question of whether simpler models based on a subset of these terms would provide an equivalent fit. However, this was not the case: the full model yielded the best fits despite having more free parameters (see Supplementary Section S6).

Simplifying hierarchical stimuli

One fundamental issue with hierarchical stimuli is that the global shape is formed using the local shapes, making them inextricably linked. We therefore wondered whether hierarchical stimuli can be systematically related to simpler stimuli in which the global and local shape are independent of each other. We devised a set of "interior-exterior" shapes whose representation in visual search can be systematically linked to that of the hierarchical stimuli, and thereby simplifying their underlying representation. Even here, we found that the dissimilarity between interior-exterior stimuli can be explained as a linear sum of shape relations across multiple scales (see Supplementary Section S7). Moreover, changing the position, size, and grouping status of the local elements leads to systematic changes in the model parameters (see Supplementary Sections S7–S9). These findings provide a deeper understanding of how shape information combines across multiple scales.

General discussion

Classic perceptual phenomena, such as the global advantage and interference effects, have been difficult to understand because they have been observed during shape detection tasks, where a complex category judgment is made on a complex feature representation. Here, we have shown that these phenomena are not a consequence of the categorization process but rather are explained by intrinsic properties of the underlying shape representation. Moreover, this underlying representation is governed by a simple rule whereby global and local features combine linearly.

Our findings in support of this conclusion are: (1) global advantage and interference effects are present in a same-different task as well as in a visual search task devoid of any shape categorization; (2) responses in the same-different task were accurately predicted using two factors: dissimilarity and distinctiveness; (3) dissimilarities in visual search were explained using a simple linear rule whereby the net dissimilarity is a sum of pairwise multiscale shape dissimilarities; and (4) shape parameters estimated in both tasks were correlated, indicative of a common underlying shape representation. Below, we discuss how these results relate to the existing literature.

Understanding same-different task performance

We have found that image-by-image variations in response times in the same-different task can be accurately explained using a quantitative model. To the best of our knowledge, there are no quantitative models for the same-different task. According to our model, responses in the same-different task are driven by two factors: dissimilarity and distinctiveness.

The first factor is the dissimilarity between two images in a pair. Notably, it has opposite effects on "SAME" and "DIFFERENT" responses. This makes intuitive sense because if images are more dissimilar, it should make "SAME" responses harder and "DIFFERENT" responses easier. It is also consistent with the common models of decision-making (Gold & Shadlen, 2002) and categorization (Ashby & Maddox, 1994; Mohan & Arun, 2012), where responses are triggered when a decision variable exceeds a criterion value. In this case, the decision variable is the dissimilarity.

The second factor is distinctiveness. Response times were faster for images that are more distinctive (i.e. far away from other stimuli). To the best of our knowledge, this has never been reported previously. However, it makes intuitive sense because nearby stimuli can act as distractors and slow down responses. Importantly, the distinctiveness of an image in the global block matched best with its average distance from all other stimuli with the same global shape (Figure 7A). Conversely, the distinctiveness in the local block matched best with its average distance from all other shapes with the same local shape (Figure 7B). Thus, distinctiveness is task-dependent and/or reflects attentional demands. This finding is concordant with norm-based accounts of object representations (Sigala, Gabbiani, & Logothetis, 2002; Leopold, Bondar, & Giese, 2006), wherein objects are represented relative to an underlying average. We speculate that this underlying average is biased by the level of attention, making stimuli distinctive at the local or global level depending on the block. Testing these intriguing possibilities will require recording neural responses during global and local processing.

Explaining global advantage and interference effects

We have shown that the global advantage and interference effects also occur in visual search, implying that they are intrinsic properties of the underlying shape representation. Further, we found that this representation is organized according to a simple linear rule whereby global and local features combine linearly (see Figure 5). This model provides a simple explanation of both effects. The global advantage occurs simply because global part relations are more salient than local relations (see Figure 5C). The interference effect occurs because congruent stimuli are more dissimilar (or equivalently, more distinctive) than incongruent stimuli, which in turn is because congruent stimuli have no within-object part differences (see Figure 6). These findings are consistent with previous studies showing that a variety of factors combine linearly in visual search (Pramod & Arun, 2016; Pramod & Arun, 2018; Sunder & Arun, 2016).

Finally, it has long been observed that the global advantage and interference effects vary considerably on the visual angle, eccentricity and shapes of the local elements (Navon, 1977; Navon & Norman, 1983; Kimchi, 1992; Poirel et al., 2008). Our results offer a systematic approach to understand these variations: the multiscale model parameters varied systematically with the position, size, and grouping status of the local elements (see Supplementary Sections S3–S5).

In sum, our results elucidate global and local processing phenomena by relating them to a systematic underlying shape representation governed by simple linear rules.

Keywords: global advantage, global-local interference, hierarchical stimuli

Acknowledgments

Supported by the DBT/Wellcome Trust India Alliance Senior Fellowship awarded to S.P.A. (Grant# IA/S/17/1/503081).

G.J. and S.P.A. designed the experiments. G.J. collected the data. G.J. and S.P.A. analyzed and interpreted the data and wrote the manuscript.

All data and code required to reproduce the findings is publicly available at https://osf.io/zt8k3/.

Commercial relationships: none.
Corresponding author: S. P. Arun.
Email: sparun@iisc.ac.in.
Address: Centre for Neuroscience, Indian Institute of Science, Bangalore 560012, India.

References

Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search: The Frankenbear experiment. *Vision Research, 54*, 20–30.

Arun, S. P. (2012). Turning visual search time on its head. *Vision Research, 74*, 86–92.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38*, 423–466.

Avarguès-Weber, A., Dyer, A. G., Ferrah, N., & Giurfa, M. (2015). The forest or the trees: preference for global over local image processing is reversed by prior experience in honeybees. *Proceedings Biological Sciences, 282*, 20142384.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., . . . Minshew, N. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia, 44*, 110–129.

Bihrle, A. M., Bellugi, U., Delis, D., & Marks, S. (1989). Seeing either the forest or the trees: dissociation in visuospatial processing. *Brain and Cognition, 11*, 37–49.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Cavoto, K. K., & Cook, R. G. (2001). Cognitive precedence for local information in hierarchical stimulus processing by pigeons. *Journal of Experimental Psychology Animal Behavior Processes, 27*, 3–16.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433–458.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature, 382*, 626–628.

Franceschini, S., Bertoni, S., Gianesini, T., Gori, S., & Facoetti, A. (2017). A different vision of dyslexia: Local precedence on global perception. *Scientific Reports, 7*, 17462.

Freedman, D. J., & Miller, E. K. (2008). Neural mechanisms of visual categorization: Insights from neurophysiology. *Neuroscience and Biobehavioral Reviews, 32*, 311–329.

Gerlach, C., & Poirel, N. (2018). Navon's classical paradigm concerning local and global processing relates systematically to visual object classification performance. *Scientific Reports, 8*, 324.

Gerlach, C., & Starrfelt, R. (2018). Global precedence effects account for individual differences in both face and object recognition performance. *Psychonomic Bulletin & Review, 25*, 1365–1372.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*, 237–288.

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron, 36*, 299–308.

Han, S., Jiang, Y., & Gu, H. (2004). Neural substrates differentiating global/local processing of bilateral visual inputs. *Human Brain Mapping, 22*, 321–328.

Han, S., Weaver, J. A., Murray, S. O., Kang, X., Yund, E. W., & Woods, D. L. (2002). Hemispheric asymmetry in global/local processing: effects of stimulus position and spatial frequency. *Neuroimage, 17*, 1290–1299.

Kimchi, R. (1988). Selective attention to global and local levels in the comparison of hierarchical patterns. *Perception & Psychophysics, 43*, 189–198.

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

Kimchi, R. (1998). Uniform connectedness and grouping in the perceptual organization of hierarchical patterns. *Journal of Experimental Psychology Human Perception and Performance, 24*, 1105–1118.

Kimchi, R., Hadad, B., Behrmann, M., & Palmer, S. E. (2005). Microgenesis and ontogenesis of perceptual organization: evidence from global and local processing of hierarchical patterns. *Psychological Science, 16*, 282–290.

Lachmann, T., & Van Leeuwen, C. (2008). Different letter-processing strategies in diagnostic subgroups of developmental dyslexia. *Cognitive Neuropsychology, 25*, 730–744.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.

Lamb, M. R., & Robertson, L. C. (1990). The effect of visual angle on global and local reaction times depends on the set of visual angles presented. *Perception & Psychophysics, 47*, 489–496.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature, 442*, 572–575.

Liu, L., & Luo, H. (2019). Behavioral oscillation in global/local processing: Global alpha oscillations mediate global precedence effect. *Journal of Vision, 19*, 12.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research, 66*, 579–619.

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171.

Malinowski, P., Hübner, R., Keil, A., & Gruber, T. (2002). The influence of response competition on cerebral asymmetries for processing hierarchical stimuli revealed by ERP recordings. *Experimental Brain Research, 144*, 136–139.

Miller, J., & Navon, D. (2002). Global precedence and response activation: evidence from LRPs. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 55*, 289–310.

Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision, 12*, 19.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin and Review, 8*, 454–469.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.

Navon, D., & Norman, J. (1983). Does global precedence really depend on visual angle? *Journal of Experimental Psychology Human Perception and Performance, 9*, 955–965.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*, 72–107.

Pitteri, E., Mongillo, P., Carnier, P., & Marinelli, L. (2014). Hierarchical stimulus processing by dogs (Canis familiaris). *Animal Cognition, 17*, 869–877.

Poirel, N., Pineau, A., & Mellet, E. (2008). What does the nature of the stimuli tell us about the Global Precedence Effect? *Acta Psychologica (Amsterday), 127*, 1–11.

Pramod, R. T., & Arun, S. P. (2014). Features in visual search combine linearly. *Journal of Vision, 14*, 1–20.

Pramod, R. T., & Arun, S. P. (2016). Object attributes combine additively in visual search. *Journal of Vision, 16*, 8.

Pramod, R. T., & Arun, S. P. (2018). Symmetric objects become special in perception because of generic computations in neurons. *Psychological Science, 29*, 95–109.

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*, 135–147.

Robertson, L. C., & Lamb, M. R. (1991). Neuropsychological contributions to theories of part/whole organization. *Cognitive Psychology, 23*, 299–330.

Romei, V., Driver, J., Schyns, P. G., & Thut, G. (2011). Rhythmic TMS over parietal cortex links distinct brain frequencies to global versus local visual processing. *Current Biology (CB), 21*, 334–337.

Shaw, R. G., Mitchell-olds, T., & Mitchell-olds, T. (1993). ANOVA for unbalanced data: An overview. *Ecology, 74*, 1638–1645.

Sigala, N., Gabbiani, F., & Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience, 14*, 187–198.

Slavin, M. J., Mattingley, J. B., Bradshaw, J. L., & Storey, E. (2002). Local-global processing in Alzheimer's disease: an examination of interference, inhibition and priming. *Neuropsychologia, 40*, 1173–1186.

Song, Y., & Hakoda, Y. (2015). Lack of global precedence and global-to-local interference without local processing deficit: A robust finding in children with attention-deficit/hyperactivity disorder under different visual angles of the Navon task. *Neuropsychology, 29*, 888–894.

Sripati, A. P., & Olson, C. R. (2009). Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. *Journal of Neuroscience, 29*, 7788–7796.

Sunder, S., & Arun, S. P. (2016). Look before you seek: Preview adds a fixed benefit to all searches. *Journal of Vision, 16*, 3.

Tanaka, H., & Fujita, I. (2000). Global and local processing of visual patterns in macaque monkeys. *Neuroreport, 11*, 2881–2884.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*, 682–687.

Vighneshvel, T., & Arun, S. P. (2013). Does linear separability really matter? Complex visual search is explained by simple search. *Journal of Vision, 13*, 1–24.

Vincent, B. T. (2011) Search asymmetries: Parallel processing of uncertain sensory information. *Vision Research, 51*, 1741–1750.

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