Object attributes combine additively in visual search

R. T. Pramod
Centre for Neuroscience, Indian Institute of Science, Bangalore, India

S. P. Arun
Centre for Neuroscience, Indian Institute of Science, Bangalore, India

We perceive objects as containing a variety of attributes: local features, relations between features, internal details, and global properties. But we know little about how they combine. Here, we report a remarkably simple additive rule that governs how these diverse object attributes combine in vision. The perceived dissimilarity between two objects was accurately explained as a sum of (a) spatially tuned local contour-matching processes modulated by part decomposition; (b) differences in internal details, such as texture; (c) differences in emergent attributes, such as symmetry; and (d) differences in global properties, such as orientation or overall configuration of parts. Our results elucidate an enduring question in object vision by showing that the whole object is not a sum of its parts but a sum of its many attributes.

Introduction

Understanding object vision has been difficult because we perceive objects as containing a variety of attributes: local shape, texture, parts, part relations, and global attributes (Palmer, 1999). Studying these features has proved difficult primarily because they can potentially interact in a number of important ways. First, visual objects contain a vast number of potential local features that require rigorous and extensive testing to be validated with the result that we seldom understand how features combine (Attneave, 1950; R. Shepard, 1964; Tversky, 1977; Wolfe & Horowitz, 2004). Second, there may be emergent features formed by the combination of local features that may in turn interact with others (Kimchi & Bloch, 1998; Pomerantz & Portillo, 2011; Pomerantz & Pristach, 1989; Pomerantz, Sager, & Stoever, 1977). Third, features may be processed in a coarse-to-fine sequence with global factors taking precedence (Kimchi, 1992; Navon, 1977; Sripati & Olson, 2009; Wagemans et al., 2012b), but it is not clear how global and local factors combine in perception. Thus, understanding feature integration remains a fundamental issue in object vision. Our current understanding of feature integration comes from a number of experimental approaches, which we review below.

Studies of perceived dissimilarity

The first approach involves measuring the perceived dissimilarity between objects generated under some form of parametric control (Attneave, 1950; R. Shepard, 1964). A classic finding, for example, is that the perceived dissimilarity between objects differing in two features (e.g., rectangles that differ in both size and brightness) is a sum of the dissimilarity along each dimension (Attneave, 1950). Although some features sum linearly (Dunn, 1983; Wiener-Ehrlich, 1978), other features sum nonlinearly (Hyman & Well, 1967; Krantz & Tversky, 1975; Wender, 1971). A concordant pattern of findings were found using classification tasks in which subjects had to classify objects along one feature dimension while ignoring variations of a second feature (Garner & Felfody, 1970; Gottwald & Garner, 1972). Some features were separable, in that classification along one feature was unaffected by manipulations of the other, whereas other features were termed "integral" because classification was affected by the irrelevant variations. These studies do not explain why some features are separable and sum linearly whereas others are integral and sum nonlinearly. The resolution might lie in the tacit assumption that the manipulated parameters are the features relevant for perception. For instance, nonlinear interactions between two features might indicate the presence of an additional feature. We have recently shown this to be the case for rectangle length and width, a well-known pair of integral features (Pramod & Arun, 2014). Here, including aspect ratio as an additional feature makes length and width sum linearly. Thus, it is critical to study feature integration

Citation: Pramod, R. T., & Arun, S. P. (2016). Object attributes combine additively in visual search. *Journal of Vision*, *16*(5):8, 1–29, doi:10.1167/16.5.8.

doi: 10.1167/16.5.8 Received September 3, 2015; published March 10, 2016 ISSN 1534-7362

without making assumptions about the underlying features.

A broader, theoretical framework for understanding conceptual similarity was proposed by Amos Tversky (1977). In this framework, a concept is a set of features, and the net dissimilarity is a sum of shared and distinctive features of the two concepts. This model explains a variety of phenomena, such as asymmetric judgments and contextual effects in judgments of similarity. However, this framework also requires explicitly defined features, which again makes any inference about feature integration problematic.

Emergent features

The second approach to understanding feature integration involves investigating how adding a fixed context to an existing display changes perception. The classic finding is that searching for a "(" among ")" is hard, but adding "(" to every shape in the display to make it a search for "((" among "()," results in an easy search. In other words, detecting the same change when embedded in a fixed configuration is easier than detecting the change by itself (Kimchi & Bloch, 1998; Pomerantz & Portillo, 2011; Pomerantz & Pristach, 1989; Pomerantz et al., 1977). These configural superiority effects have been frequently attributed to Gestalt properties, such as closure, parallelism, symmetry, etc. (Pomerantz & Portillo, 2011). These emergent properties place important constraints on models of object vision. Despite these insights, this approach does not explain how features, emergent or not, combine in object vision.

Overview of the current study

As detailed above, feature integration has to be explored with the caveat that many features can potentially interact and result in emergent features. We therefore asked: Is it at all possible to understand feature integration without knowing or making assumptions about the underlying features? We started with the observation that many natural objects appear distinct despite containing many similar features (Figure 1A). However, natural objects may contain too many distinct features, making them difficult to manipulate systematically.

Our key insight was to create a large number of objects from a small number of parts (Figure 1B) and measure all possible pairwise dissimilarities between these objects. We then hypothesized that the net dissimilarity between two objects would depend on the dissimilarities between their parts. Importantly, because a particular part pair will occur in many pairs of objects, its contribution to object dissimilarity (assuming it is the same across object pairs) should be recoverable from many dissimilarity measurements. This amounts to solving a set of equations in which the observations are the large number of measured dissimilarities between objects and the unknowns represent the contribution of each part–part relation to the total dissimilarity.

Throughout this study, we have used the term "parts" to denote distinct fragments that are attached to predefined locations on a stem to form a whole object. These fragments do not always correspond to the perceived part structure of the whole object; for instance, some of these fragments are themselves perceived as containing further subparts. In general, objects are thought of as containing perceived part structure based on either decomposing them into shape primitives (Biederman, 1987; Marr & Nishihara, 1978) or using a set of boundary-based rules such as local curvature extrema (Hoffman & Singh, 1997; Palmer, 1999). We have explored this issue in greater detail in Experiment 7 by asking whether dissimilarities between a fixed set of objects can be explained better using fragments consistent or inconsistent with their perceived part decomposition.

Throughout this study, we have shown that a variety of object attributes sum linearly, implying that distances in perceptual space combine according to a linear rule. However this does not imply full linearity because a linear function must show additivity (i.e., linear summation of attributes) as well as scaling (i.e., scaling the input proportionally scales the output). Although scaling is impossible to test without knowing the underlying features, additivity or linear summation (which we use interchangeably throughout) can be evaluated despite this lack of knowledge (see General discussion).

In Experiment 1, for example, we created 49 objects by combining seven parts on either side of a stem and measured all 1,176 possible pairs of dissimilarities between these objects (1,176 is the number of ways of choosing two objects out of 49, hereafter denoted as ${}^{49}C_2$). The number of part–part relations, however, is only 21 (${}^{7}C_2$). This enabled us to quantitatively address an enduring question in object vision: Can distances between objects be understood in terms of their parts?

To measure perceived distances between objects, we used a visual search paradigm. For each pair of objects, we created search arrays in which one object was an oddball target and many copies of another object were used as distracter items. We used the reciprocal of the average time taken by subjects to find the oddball target as an index of the perceived distance between the objects. The reciprocal of reaction time (RT) can be interpreted as the underlying salience signal that drives search (Arun, 2012) and yields more accurate accounts of visual search compared to accounts based on RTs (Pramod & Arun,

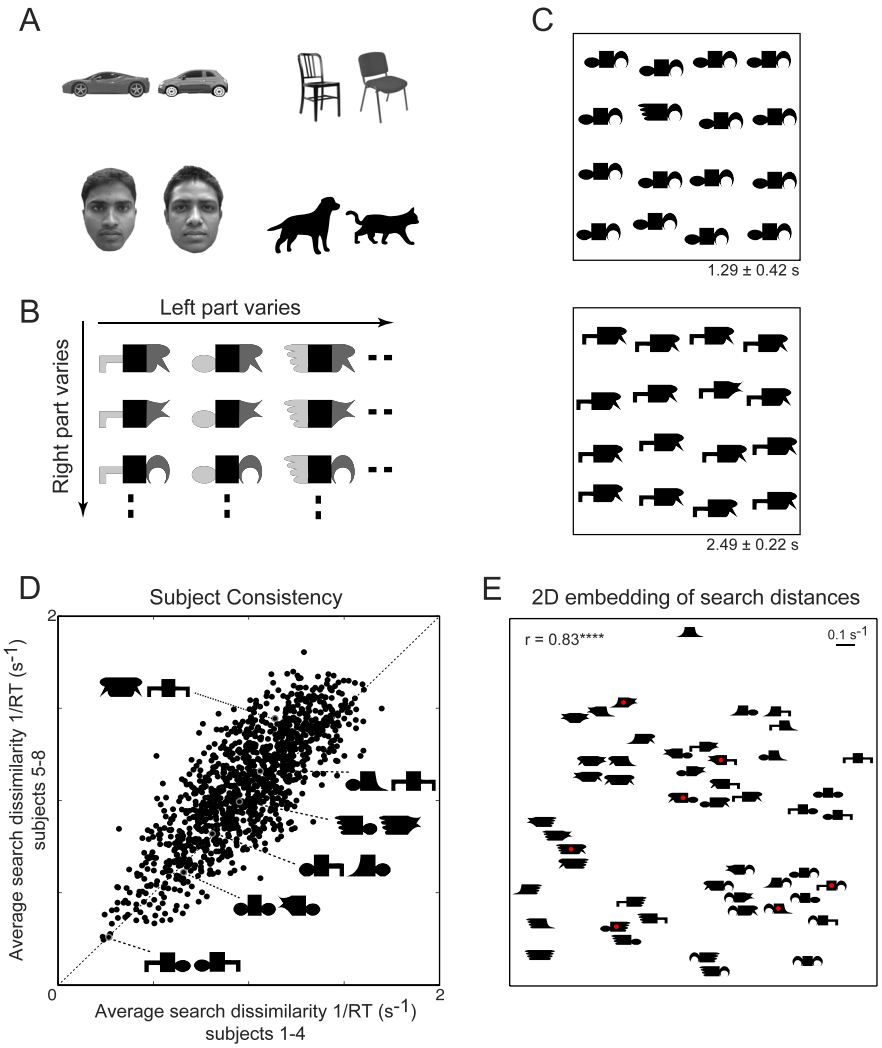

Figure 1. Objects and parts (Experiment 1). (A) Many natural objects appear distinct because of a collection of differences between their parts. How do part relations combine to make objects similar or dissimilar? (B) Schematic of object creation. We created a total of 49 objects by combining seven possible parts at either end of a stem and measured perceived distances between all 1,176 pairs of these objects. Left parts (light gray) and right parts (dark gray) are colored differently for illustration. In the experiment, each object was entirely white presented against a black background. (C) Example search arrays used to measure perceived distances (top panel: easy search; bottom panel: hard search). Each array contained an oddball target among identical distracters. In both panels, the target differs in one part from the distracters. RT (mean $\pm$ *SD*) is shown at the bottom right of each panel. (D) Observed dissimilarities from one group of four subjects were plotted against dissimilarities from a second group of four subjects. Here and in all figures, asterisks represent statistical significance (*$p < 0.05$, **$p < 0.005$, etc.). (E) Visualization of perceptual space. We performed multidimensional scaling to embed all 1,176 observed search distances between 49 objects into a two-dimensional space. Nearby images represent hard searches. The correlation coefficient represents the correspondence between the observed distances and the distances in the 2-D plot. Objects with red dots are the ones used in Experiment 9.

2014; Vighneshvel & Arun, 2013). Even in this study, models based on search distances produced much more accurate accounts of the data compared to models based on search times (Experiment 1).

We performed a total of 12 experiments. In Experiments 1–10, we investigated how local features combine in vision by asking whether part relations explain object relations. In Experiments 11 and 12, we asked how global and local features combine by asking how changes in a global attribute modify the dissimilarities between objects differing in their local parts. These experiments are detailed below.

Local feature integration (Experiments 1–10)

In Experiments 1–5, we used objects with distinctive parts joined together by a stem. The stem ensured that parts were spatially separated to minimize emergent

features, but we tested this assumption in later experiments. In Experiment 1, we found that the net dissimilarity between two-part objects is explained by the sum of local part relations. Importantly, symmetric objects were systematically more distinct by a fixed offset from the model predictions. Thus, symmetry adds to the local part dissimilarities and does not interact with them. In Experiment 2, we confirmed that our results from Experiment 1 hold when subjective dissimilarity ratings are used instead of visual search distances. In Experiment 3, we confirmed that model performance and the effect of symmetry do not depend on object orientation. In Experiment 4, we show that objects with repeated parts are also systematically more distinct than predicted by the part summation model. In Experiment 5, we measured distances between three-part objects to ask how part relations changed with distance. We found that parts that are further away interact weakly, suggesting that the part comparison process is spatially tuned.

In Experiments 6–9, we tested whether our results would hold when parts were no longer connected, unique, or salient. In short, they did. In Experiment 6, we asked how part relations change when parts are no longer connected. We found that part relations within connected objects and within disconnected objects are fundamentally different. In Experiment 7, we asked whether object distances could be explained better when they were broken into their natural parts compared to their unnatural parts. This was indeed the case. In Experiment 8, we asked whether part relations could explain distances between "holistic" objects that are not perceived as containing any obvious parts. Here too, object distances were accurately predicted as a sum of part relations. This somewhat surprising finding might reflect the subjects' tendency to parse even whole objects into parts when they are viewed along with other objects that share some contours with the whole object. To address this issue, in Experiment 9, we measured dissimilarities between holistic objects in a separate group of subjects that were never exposed to the shared parts. Even these dissimilarities were explained by the part-based model, suggesting that perceived dissimilarities might be driven by a contour-matching process.

In Experiment 10, we investigated whether the results from the preceding experiments—based on manipulating shape—would extend to other object properties. To this end, we created objects that could differ in shape, texture, or both properties. Here too, distances between objects were explained by a linear sum of shape and texture dissimilarities.

Global properties (Experiments 11 and 12)

The results of Experiments 1–10 show that the net dissimilarity between objects is almost entirely explained as a linear sum of part distance relations. The objects in these experiments were designed with the same overall global structure to prevent these global factors from influencing search. In Experiments 11 and 12, we examined how a change in a global attribute modifies the dissimilarity between objects that differ in their local parts. By a global attribute, we mean a property that modifies the entire object as a whole rather than any single part. For instance, a change in the overall orientation of an object changes all features rather than only a subset. Likewise, a change in the configuration of parts induced by elongating the connecting stem or moving one part relative to another changes feature locations but leaves their local features unchanged. In Experiments 11 and 12, we investigated how the dissimilarity between two objects differing in their parts is modified by changes in a variety of global attributes. Strikingly, changes in global properties only introduced fixed increases in dissimilarities, suggesting that they sum linearly with local properties.

Experiment 1: Two-part objects

Here, our goal was to investigate whether the dissimilarity between objects (as measured using visual search) can be understood in terms of the dissimilarities between their parts. We created a total of 49 two-part objects by combining seven possible parts on either side of a stem (Figure 1B). We took advantage of the combinatorial nature of this set of objects by asking how a large number of object–object dissimilarities ($^{49}C_2 = 1{,}176$; where $^{49}C_2$ denotes the number of possible distinct pairs of 49 objects) can be explained using a relatively small number of part relations ($^{7}C_2 = 21$).

Method

Participants

Eight human subjects (five female, aged 20–30 years) participated in this experiment. In this and all following experiments, subjects 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.

Stimuli

Each stimulus was created using two of seven possible parts joined together by a stem (Figure 1B). The parts were designed such that the resulting objects ranged from very similar to very dissimilar. The set of

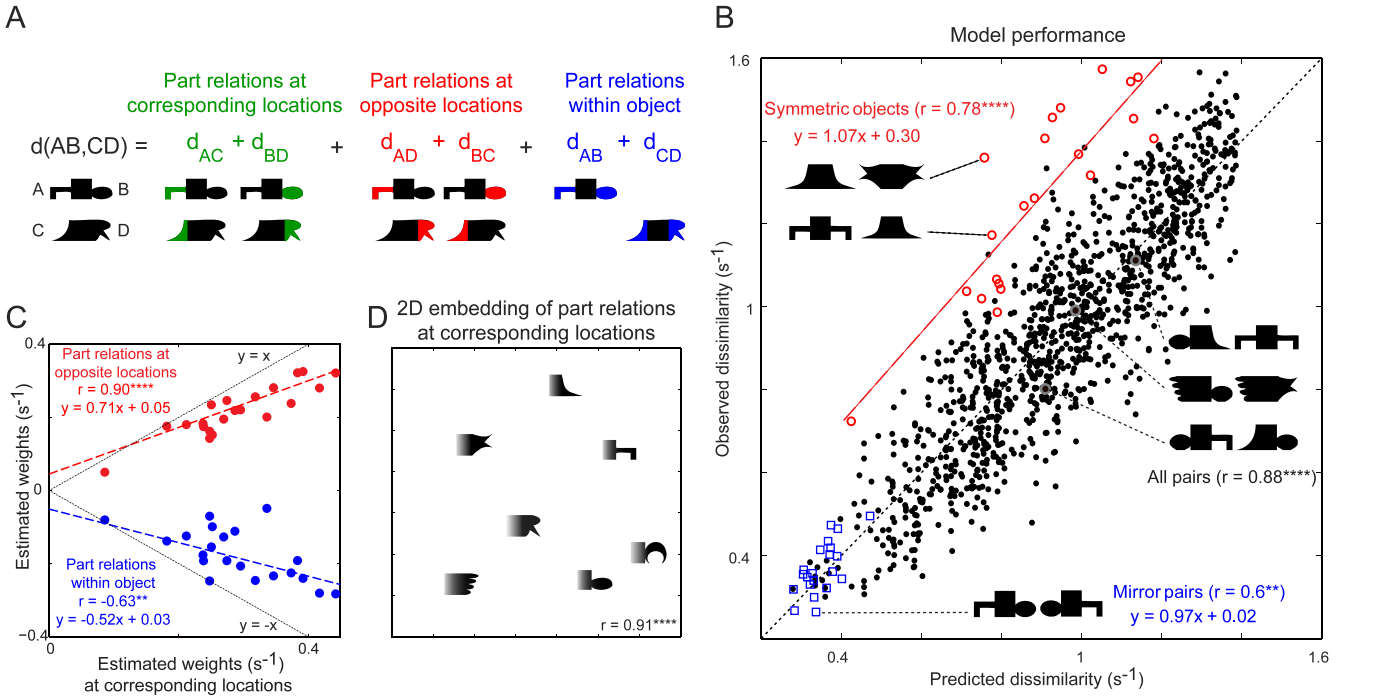

Figure 2. Perceived object relations are explained using part summation (Experiment 1). (A) Schematic of the part summation model. According to the model, the perceived distance between two objects AB and CD is a linear sum of distances between parts at corresponding locations (green), parts at opposite locations (red), and parts within each object *(blue)*. (B) Observed dissimilarity plotted against predicted dissimilarity for all 1,176 object pairs. Object pairs with global attributes are highlighted: mirror-related pairs (blue squares) and symmetric object pairs (red circles). The red dashed line is the best-fitting line for symmetric object pairs. (C) Part relations at opposite locations (red) and within-object locations (blue) plotted against part relations at corresponding locations. Dashed lines indicate the corresponding best-fitting lines. All part relations are significantly correlated but vary in magnitude, suggesting that a single set of part relations drives object dissimilarity. (D) Two-dimensional embedding of part relations at corresponding locations, showing differences between estimated part distances that ultimately drive object dissimilarity. The correlation coefficient represents the correlation between the estimated part relations and the 2-D distances in this plot.

seven parts used in this experiment is shown in Figure 2D. The entire set consisted of 49 objects containing all possible combinations of parts at either location (Figure 1E).

Procedure

Subjects were seated approximately 60 cm from a computer monitor that was under control of custom programs written using Psychtoolbox (Brainard, 1997) in Matlab. In all experiments, in each trial, a fixation cross was shown for 500 ms followed by a 4×4 search array (measuring $21° \times 21°$ with items measuring 3° along the longer dimension with 3° interitem spacing) containing one oddball item among multiple identical distracters with a red vertical line down the middle. Items were centered at the grid locations but were jittered about the center by $\pm 0.45°$ according to a uniform distribution to prevent alignment cues from guiding search. Subjects were asked to report using a key press (Z for left, M for right) the side on which the oddball target appeared as quickly and accurately as possible. They had to make a response within 10 s after which the trial was aborted. The reciprocal of the average time taken by subjects to make a key press was taken as an estimate of the perceptual distance between the oddball target and the distracter. In all, subjects performed two correct trials for each pair of objects (either item was target and could appear on the left or right). Error trials were repeated after a random number of other trials.

Model fitting

In Experiment 1, we estimated perceived distances in visual search between 49 objects created by combining seven parts on either side of a stem. This resulted in ${}^{49}C_2 = 1{,}176$ perceived distances measured using visual search. The part summation model consisted of ${}^{7}C_2$, i.e., 21 part relations at corresponding and opposite locations and an equal number of within-object part relations. Together with a constant term, the model had 64 free parameters. We used similar models for smaller numbers of parts. In addition to the linear fit, we also

transformed model predictions using a sigmoid function (integral of Gaussian with three parameters: mean, variance, and peak level) to account for the fact that the measured dissimilarity (1/RT) cannot increase indefinitely due to motor constraints on RT. In general the sigmoid nonlinearity yielded only subtle improvements in the quality of fit ($r = 0.86$ for linear model; $r = 0.88$ for the sigmoid model in Experiment 1).

Model cross-validation

The model performance reported in each experiment was based on fitting all the data, so it is possible that the striking fits of the model are due to the model overfitting the data. To assess this possibility, we repeated all our analysis by fitting the model on 80% of the data and testing it on the remaining 20% of the observed dissimilarities that the model had never been trained on before. We obtained qualitatively similar fits using cross-validation. We observed a strong correlation between observed dissimilarities and model predictions on test data sets obtained through 10 random 80/20 splits of the data (mean ± standard deviation of correlation across the 10 splits: $r = 0.85 \pm 0.02$, $p < 0.00005$). Thus, the striking agreement between the model and data did not arise from overfitting.

Analysis of search asymmetries

The dissimilarity between a pair of objects was calculated by taking the reciprocal of the search time averaged across all trials and across all subjects. However doing so ignores any search asymmetries that may be present in the data. It is well known that searches can be sometimes asymmetric; for instance, it is easier to search for a Q among Os than for an O among Qs (Wolfe, 2001). A detailed analysis of search asymmetries is beyond the scope of the present study because we have only two search times per object pair from each subject with either object as target. To assess whether there are any consistent search asymmetries at all, we separated the subjects into two groups and asked whether there are object pairs that show statistically significant asymmetries in both groups (using a t test between search times with one object as target vs. the other as target, criterion of $p = 0.05$). This yielded only five object pairs (of 1,176) with a significant asymmetry in both groups, which rules out any further systematic analysis. We nonetheless tried to assess, as far as the model is concerned, whether model performance is affected by search asymmetries. To this end, for each object pair, we identified the target with the smaller average search time (i.e., the easy target) and took the dissimilarity for that pair to be the reciprocal of the search time. We then asked whether model fits using the easier searches across object pairs were any different from model fits based on the hard searches. We obtained equally striking correlations in both cases (model-data correlations: $r = 0.83$ for the easy target; $r = 0.85$ for the hard target, $p < 0.00005$ in both cases). Thus, at least in our data, search asymmetries do not affect model fits.

Although asymmetries were, in general, harder to detect for individual object pairs, there is one group of objects for which search asymmetries do systematically occur. It is known that search for an asymmetric target among symmetric distracters is easier than the other way around (Olivers & van der Helm, 1998). To test if this was true in our data, we compared search RTs for object pairs involving symmetric and asymmetric objects. This included a total of 294 object pairs. For these pairs, search times for a symmetric target among asymmetric distracters were longer than search times for an asymmetric target among symmetric distracters (mean search times: symmetric targets: 1321 ms; asymmetric targets: 888 ms; z statistic = 12.34, $p < 0.00005$, ranksum test).

Results

To measure perceived dissimilarity between objects, we created visual search displays in which one object was embedded among multiple instances of another object (Figure 1C). We took the reciprocal of the average RT for each object pair as a measure of perceived dissimilarity between them (Arun, 2012). Visual search has the advantage of being an objective, naturalistic task requiring no explicit similarity judgments, yet dissimilarities in visual search are strongly correlated and less categorical than subjective dissimilarity ratings (see Experiment 2).

In Experiment 1, subjects performed visual search involving pairs of 49 objects created by placing two parts at either end of an elongated stem (Figure 1B). We measured all possible 1,176 ($^{49}C_2$) pairwise distances between the 49 objects. To assess whether search performance was consistent across subjects, we repeatedly split the subjects randomly into two groups and calculated the average "split-half" correlation. A large split-half correlation implies that subjects were highly consistent in their search performance. Its statistical significance (p value) represents the probability of observing a correlation at least as large as that observed given the null hypothesis that the two subject groups are uncorrelated. This analysis yielded a highly consistent correlation (mean ± SD: $r = 0.80 \pm 0.005$, $p < 0.00005$; Figure 1D). The high consistency in search RTs indicates a systematic and consistent perceptual space across subjects.

The variation in observed RTs were not due to a speed–accuracy trade-off as evidenced by a significant negative correlation between search RT and accuracy across all

searches ($r = -0.31$, $p < 0.00005$). In other words, when subjects were faster, they were also more accurate.

To visualize these objects in perceptual space, we embedded the observed distances for all 49 objects using multidimensional scaling into a two-dimensional plot. In the resulting plot, nearby objects represent hard searches (Figure 1E). This plot reveals several interesting patterns. For instance, objects that are vertical mirror images are close together, and objects that share parts tend to cluster together. We confirmed this quantitatively by comparing search times for object pairs involving vertical mirror images or shared parts with object pairs with all distinct parts. There were significant differences between these three groups (mean search times: 2630 ms for mirror pairs, 1402 ms for shared parts, and 971 ms for distinct parts; $p < 0.0005$ for the main effect of pair type in an ANOVA with subject and pair type as factors).

Overview of the part summation model

We set out to investigate whether the large number of measured dissimilarities between objects made from essentially seven parts can be explained using relations between parts. Because parts might be perceived differently in isolation and when embedded in an object, we indirectly estimated part relations rather than directly measuring them. Consider two objects AB and CD made from parts (A, B) and (C, D), respectively. We hypothesized that the perceived distance between these objects is a linear sum of all possible pairwise relations between these parts: i.e., d_{AC}, d_{BD}, d_{AD}, d_{BC}, d_{AB}, and d_{CD}. Further, perceived distance might be driven differently by part relations at corresponding locations (AC & BD), by part relations at opposite locations (AD & BC), and by part relations within the object (AB & CD) (Figure 2A). Thus, the net dissimilarity between objects AB and CD can be written as

$$d(\mathrm{AB},\mathrm{CD}) = d_{\mathrm{AC}} + d_{\mathrm{BD}} + x_{\mathrm{AD}} + x_{\mathrm{BC}} + w_{\mathrm{AB}} + w_{\mathrm{CD}} + \mathrm{constant}$$

where d_{AC} and d_{BD} represent the dissimilarity between parts AC and BD when they are at corresponding locations in the two objects, x_{AD} and x_{BC} are the dissimilarities between parts AD and BC when they are at opposite locations, and w_{AB} and w_{CD} likewise are the dissimilarities between these parts when they occur within the object. The working of the model becomes clearer on writing down the dissimilarity between another pair of objects AB and CE where only one part has changed.

$$d(\mathrm{AB},\mathrm{CE}) = d_{\mathrm{AC}} + d_{\mathrm{BE}} + x_{\mathrm{AE}} + x_{\mathrm{BC}} + w_{\mathrm{AB}} + w_{\mathrm{CE}} + \mathrm{constant}.$$

It can be seen that a number of terms are common to both equations (d_{AC}, x_{BC}, w_{AB}) whereas other terms are present in one equation but not the other. For example, the contribution of the term d_{BE} is zero in the first equation but one in the second. One can then extrapolate this observation to the equations corresponding to the 1,176 dissimilarity measurements: Each term occurs frequently enough by itself for its contribution to be estimated independently of the others. Note that the number of part relations at corresponding locations (i.e., terms of type d_{AB}, d_{AC}, etc.) are a total of ${}^{7}C_2 = 21$ in number. In all, there are 21 part relations each for corresponding, opposite, and within-object locations, which together with the constant term amount to 64 unknown terms across 1,176 equations. The resulting set of linear equations can be written down as the matrix equation $\mathbf{y} = \mathbf{Xb}$ where $\mathbf{y}$ is a vector of 1,176 dissimilarities, $\mathbf{b}$ is a vector containing the 64 unknown part relations, and $\mathbf{X}$ is a $1{,}176 \times 64$ matrix whose rows contain 0s and 1s representing whether or not a particular part pair contributes to the object pair corresponding to that row. This equation is then solved using standard linear regression to estimate the unknown vector $\mathbf{b}$ given $\mathbf{y}$ and $\mathbf{X}$. Note that the model contains separate, independent terms for each type of part relation (corresponding, opposite, within) and therefore makes no assumption about how these terms may be related.

Performance of the part summation model

The part summation model produced striking fits to the observed data ($r = 0.88$, $F(63, 1113) = 49.23$, $p < 0.001$, $r^2 = 0.77$; Figure 2B) and outperformed both simpler models (e.g., with part relations of only one kind) as well as those based on RT alone (see below). The performance of this model is even better than the split-half correlation ($r = 0.80$) described above; this is because the split-half correlation estimates the consistency of half the data whereas the model is fit to the full data set, which is more consistent. To estimate the true consistency of the full data set, we applied a standard correction called the Spearman-Brown formula, which estimates the correlation between two full data sets based on the correlation obtained between n-way splits of the data. For a two-way split, i.e., the split-half correlation, the Spearman-Brown corrected correlation is $r_c = 2r/(r+1)$ where r is the split-half correlation. Applying this correction to the split-half correlation yields $r_c = 0.88$. Here and in all subsequent experiments, we have reported this corrected split-half correlation as a measure of data consistency. It can be seen here that the model data correlation ($r = 0.88$) is equal to the corrected split-half correlation ($r_c = 0.88$), implying that the part summation model explains search dissimilarities as well as can be expected given the consistency of the data. We conclude that perceived

distances between whole objects can be explained as a *linear sum* of part relations.

The estimated part relations revealed several interesting insights. First, estimated part relations at corresponding locations were significantly correlated with relations at opposite locations ($r = 0.9$, $p < 0.001$) and within objects ($r = -0.63$, $p = 0.0023$), suggesting that there is a common set of underlying part relations that are modulated by object-relative location (Figure 2C). Second, parts at corresponding locations exert a stronger influence compared to parts at opposite locations (Figure 2C). Third, part relations within an object have negative contribution, which means that objects with similar parts tend to become distinctive (Figure 2C). This negative weight is analogous to the finding that search becomes easy when distracters are similar (Duncan & Humphreys, 1989; Vighneshvel & Arun, 2013). To visualize the part relationships that drive the overall object dissimilarities, we performed multidimensional scaling on the estimated corresponding part dissimilarities. The resulting 2-D embedding of the part relationships is shown in Figure 2D. It can be seen that parts that are estimated as being dissimilar in Figure 2D result in objects containing these parts to also be dissimilar (Figure 1E).

Does the part summation model explain mirror confusion?

Because the part summation model is based on local part relations, its predictions can provide a useful baseline to evaluate global attributes. By global attributes, we mean object properties that cannot be inferred by the presence of a single part but only by considering the entire object. We examined two such global attributes. The first attribute was mirror confusion. There were 21 pairs of objects of the form AB and BA that were vertical mirror images of each other. For these pairs, the observed distances were generally smaller than the other object pairs, implying mirror confusion. But importantly, model predictions were significantly correlated with the data ($r = 0.60$, $p = 0.005$; Figure 2B) with no systematic deviation. The lower correlation of the model could be either due to the relatively fewer points or because subjects were themselves more variable in their responses for mirror pairs. We found the latter to be true (average split-half correlation of dissimilarities between mirror object pairs across two groups of subjects: $r = 0.71$, $p = 0.001$). What makes the model explain mirror confusion? Consider what happens for a mirror pair AB versus BA. The net dissimilarity can be written as $d(\mathrm{AB}, \mathrm{BA}) = d_{\mathrm{AB}} + d_{\mathrm{AB}} + x_{\mathrm{AA}} + x_{\mathrm{BB}} + w_{\mathrm{AB}} + w_{\mathrm{AB}}$. But the terms x_{AA} and x_{BB} are taken to be zero in the model. This reduces the net distance between objects, resulting in mirror confusion.

Does the part summation model explain symmetry?

The second global property we examined was symmetry. There were seven symmetric objects produced by adding the same part to both ends of the stem. Consequently, there were 21 ($^7C_2$) observed distances between symmetric objects. Model predictions were strongly correlated with the observed distances ($r = 0.78$, $p < 0.001$; Figure 2B). This correlation was close to the consistency between subjects for these symmetric object pairs (average split-half correlation: $r = 0.76$, $p = 0.0063$). Despite these strong correlations, the model systematically underestimated the observed distances by a constant offset (mean slope: 1.07 with 95% confidence interval [0.64 1.5]; intercept: 0.29 with 95% confidence interval [0.03 0.55]). The constant offset obtained for symmetric object pairs was present equally strongly in both horizontally and vertically oriented objects (Experiment 3). This constant offset suggests that symmetry exerts an additive influence on perceived distances independent of the part relations in the model.

The symmetric objects in this experiment consisted of the same part at both ends short of a vertical mirror reflection. Is the effect then due to mirror reflection or because of repetition of the parts? We tested this possibility in Experiment 4.

Comparison with models based on RT

Our results are based on using the reciprocal of visual search time (1/RT) as a measure of dissimilarity. Our choice of 1/RT over RT is motivated by our recent work showing that models that use 1/RT are able to predict search performance much better than models based directly on RT (Pramod & Arun, 2014; Vighneshvel & Arun, 2013). We reconfirmed this to be the case in this data as well: We simply substituted RT as the observed data in our model instead of 1/RT and asked whether observed RTs are explained using a part summation model. We assessed model performance using two criteria. The first was simply the correlation between model predictions and the observed data. This number can be misleading if the models to be compared differ in their number of free parameters because the model with more free parameters will tend to produce a better fit. We therefore also evaluated another criterion, namely the corrected Akaike information criterion—denoted as AICc—which is a measure of quality of fit for a model that penalizes it for complexity (Pramod & Arun, 2014). The AICc of a model is calculated as

$$\mathrm{AICc} = N\log\frac{\mathrm{SS}}{N} + 2K + \frac{2K(K+1)}{(N-K-1)}$$

where N is the number of observations, SS is the sum of squared error between the model and data, and K is the number of free parameters in the model. In general, the

Performance measure	RT model	1/RT model
Correlation with observed RT	0.77	0.88
Correlation with observed 1/RT	0.49	0.88
Quality of fit on RT (AICc, mean ± *SD*)	2623.8 ± 115	2847.6 ± 143
Quality of fit on 1/RT (AICc, mean ± *SD*)	1148 ± 436	4432.9 ± 58

Table 1. Comparison of RT and 1/RT models. *Notes*: Model performance was evaluated on the 1,176 object pairs used in Experiment 1. Standard deviations for the AICc were calculated by generating bootstrap resampled data, fitting the model each time, and calculating the AICc. All correlations were significant with $p < 0.00005$.

more negative the AICc, the better the model. But for ease of comparison, we considered the absolute value of the AICc so that a larger value of the AICc indicates better quality of fit. Our results are summarized in Table 1. Standard deviations for the AICc were calculated by generating bootstrap resampled data, fitting the model each time, and calculating the AICc. The AICc values for the 1/RT model were significantly larger than the AICc values for the RT model, both when explaining RT or 1/RT values ($p < 0.00005$, paired t test across 1,176 bootstrap-derived estimates of AICc). We conclude that 1/RT models outperform RT models in terms of explaining search data.

Comparison with reduced models

The full model described above consists of part relations at corresponding locations, part relations across locations, and part relations within the object. Here, we systematically evaluated how much each of these part relations contributes to the overall model performance. To this end, we fit a number of reduced models containing only subsets of these terms and assessed model performance using the two criteria described above. The results are summarized in Table 2. The full model yielded the best fit to the data not only in terms of the overall correlation but also after taking into account its increased number of free parameters (i.e., its AICc was substantially larger than the other models). The pattern of fits also indicates that part relations at corresponding locations are the largest contributor to the overall fit followed by part relations across locations and then by part relations within objects.

Because the reduced models were degenerate versions of the full model (obtained by setting specific parameters to zero), we were also able to compare the quality of fits using the partial F test. The null hypothesis in the partial F test is that the full and reduced models are equivalent. Here too we found that the full model was significantly better in terms of the quality of fit as assessed using the partial F test, $F(42, 1113) = 40.08$ for full model versus corresponding part terms only; $F(42, 1113) = 52.14$ for full versus opposite terms only; $F(42, 1113) = 72.2$ for full versus within terms only; all ps < 0.00005.

Part relation terms included	Model correlation	Quality of fit (AICc ± *SD*)
Corresponding	0.64	3,586.8 ± 52
Opposite	0.55	3,387.6 ± 55
Within	0.37	3,126.4 ± 47
Opposite + within	0.65	3,555.4 ± 51
Corresponding + within	0.73	3,792.3 ± 50
Corresponding + opposite	0.77	3,959.7 ± 74
Full model (corresponding + opposite + within)	0.88	4,432.9 ± 57

Table 2. Comparison of models with differing numbers of terms. *Notes*: In each case, the model was fit to the 1,176 distances measured in Experiment 1. All correlations were statistically significant at $p < 0.00005$.

Comparison with models with nonlinear interactions

The above sections show that the linear part summation model has a striking agreement with the observed dissimilarities. Would the model perform even better with nonlinear terms? Although this is unlikely given the striking fit of the linear model with the data, we tested this possibility by including nonlinear terms in the part summation model. The nonlinear terms consisted of products of part relations. For instance, the net distance between objects AB and CD in the nonlinear model is given by

$$d(\mathrm{AB},\mathrm{CD}) = d_{\mathrm{AC}} + d_{\mathrm{BD}} + x_{\mathrm{AD}} + x_{\mathrm{BC}} + w_{\mathrm{AB}} + w_{\mathrm{CD}} + (N_{\mathrm{AC,BD}} + N_{\mathrm{AD,BC}} + N_{\mathrm{AB,CD}}) + \mathrm{constant}$$

where $N_{\mathrm{AC,BD}}$ represents a term that is set to one when parts AC and BD are present and so on. The total number of such nonlinear terms is therefore ${}^{21}\mathrm{C}_2$, which is 210. The nonlinear model predicted the observed data only slightly better than did the linear model ($r = 0.86$ compared to $r = 0.85$ for the linear model, $p < 0.00005$) but at the cost of many additional parameters. To assess whether this improvement was significantly greater than expected given the additional degrees of freedom in the nonlinear model, we performed a partial F test taking the nonlinear model as the full model and the linear model as the reduced model. The null hypothesis in the partial F test is that the reduced model is equivalent to the full model. This test revealed no statistically significant difference

between the two models, $F(210, 900) = 0.0151$, $p = 1$, partial F test, implying that the nonlinear model did not perform significantly better than the linear model. We conclude that the part summation model is a linear sum of part relations.

Experiment 2: Dissimilarity ratings

The main findings in Experiment 1 are based on measuring dissimilarities between objects using visual search. Would our results hold for other types of dissimilarity data? In classic studies of dissimilarity, subjects are typically asked to produce a dissimilarity rating for each pair of objects (Attneave, 1950; Shepard, 1964; Tversky, 1977). How do search dissimilarities relate to subjective dissimilarity ratings? This is an interesting question in its own right because visual search has often been thought of as a preattentive process guided by a variety of features (Wolfe & Horowitz, 2004) whereas dissimilarity ratings involve explicit scrutiny and potentially different features (Torgerson, 1965). To investigate this issue, we conducted an experiment in which subjects had to indicate subjective dissimilarity rating for a pair of objects instead of doing visual search.

Method

Participants

A total of nine subjects (three female, aged 20–30 years) participated in this experiment.

Stimuli

We used a subset of 16 objects from Experiment 1 made from four parts. We presented all possible 120 pairs of objects in the rating task. Object size and interobject spacing was identical to that used in the visual search experiment.

Procedure

Participants were shown a pair of objects with a rectangular scale underneath it with the two ends marked "Very similar" and "Very different." They were instructed to make a mouse click at a point on the scale that best matched their subjective sense of dissimilarity between the two objects. The location chosen by the subject was converted into a continuous value between one and 10. Each subject performed one trial per pair of objects.

Results

Subjects were extremely consistent in their dissimilarity ratings (average corrected split-half correlation between two random groups of subjects [mean ± SD]: $r = 0.86 \pm 0.03$, $p < 0.00005$). Importantly, these ratings were strongly correlated with search dissimilarity ratings measured for the same pairs in Experiment 1 (Figure 3A; $r = 0.81$, $p < 0.00005$) even though the two data sets were collected from different groups of subjects. However, it can be seen that dissimilarity ratings tended to saturate at the extreme ends of the available range (i.e., for ratings below three and above eight). In other words, when objects became extremely similar or dissimilar, subjects tended to use a single rating that corresponded to the extreme ends of the range (Figure 3A). In contrast, there was no such clustering observed for search dissimilarities although they can and do saturate when the target–distracter dissimilarity is very large (Arun, 2012). Whereas the search dissimilarities were normally distributed ($p = 0.11$, Lilliefors test for normality), the dissimilarity ratings differed significantly from a unimodal distribution ($p < 0.00005$, dip = 0.1, Hartigan's dip test).

The high correlation between search and subjective dissimilarity might have come about because of highly similar or dissimilar object pairs. To assess this possibility, we calculated the correlation between search dissimilarity and subjective ratings for the middle range of dissimilarity ratings (between 3.0 and 8.0 ratings); this too revealed a positive and significant correlation ($r = 0.50$, $p < 0.00005$). Interestingly, for object pairs in this middle range, subjects were more consistent in the visual search task (correlation between search distances of two subject groups: $r = 0.63$, $p < 0.00005$) compared to the subjective rating task (correlation between ratings of two subject groups: $r = 0.45$, $p < 0.005$).

Next, we asked how well the part summation model explains the subjective dissimilarity data. Proceeding exactly as with the search data, we fit a model containing corresponding location, across-location, and within-object terms ($^4C_2$ = six parameters each) and asked whether the observed dissimilarity ratings could be explained using a weighted sum of the part relations. Model predictions were again strongly correlated with the data overall (Figure 3B); $r = 0.95$, $F(18, 102) = 29.84$, $p < 0.0005$, $r^2 = 0.9$. Model predictions were also strongly correlated in the middle range of dissimilarities ($r = 0.73$, $p < 0.00005$ for 75 object pairs with dissimilarity rated between three and eight). However we only saw a weak effect of symmetry in the dissimilarity data, but this could be because of the small number of symmetric objects in this experiment.

In sum, we conclude that both subjective dissimilarity ratings and visual search dissimilarities reflect a

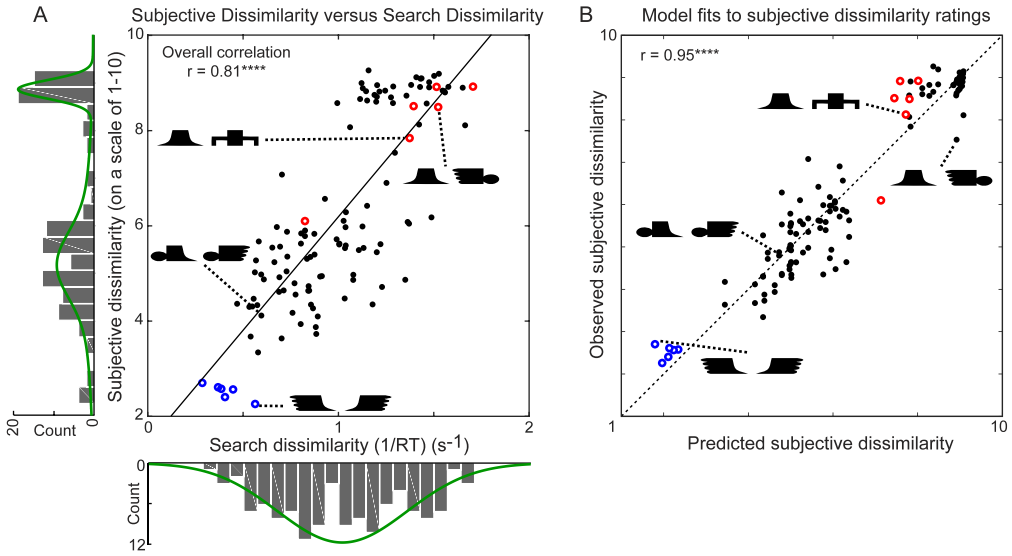

Figure 3. Visual search and subjective dissimilarity (Experiment 2). (A) Plot of subjective dissimilarity (rescaled to a range of 1–10) against search dissimilarity for the 120 object pairs created from 16 objects made by all possible combinations of four parts. Object pairs with global attributes are highlighted (symmetric object pairs: red, mirror-related pairs: blue). Marginal histograms along each axis are depicted. Green lines indicate the best-fitting mixture of Gaussian distribution of each case. (B) Performance of the part summation model on subjective dissimilarity data, showing observed dissimilarity plotted against predicted dissimilarity. (**** is $p < 0.00005$).

common underlying object representation that is explained well by a part summation model.

Experiment 3: Vertical versus horizontal objects

In Experiment 1, we showed that symmetric objects were systematically more distinct than predicted by the part summation model. These results were obtained using horizontally oriented objects. Objects in this case became symmetric when parts on either end of the stem were identical short of reflection about the vertical axis. This raises the intriguing possibility that symmetry may be related to mirror confusion. Based on the observation that mirror confusion is stronger about the vertical axis than the horizontal axis (Rollenhagen & Olson, 2000), we reasoned that symmetry should be weaker in vertically oriented objects (i.e., when it occurs due to parts reflected about the horizontal axis) than in horizontally oriented objects (i.e., when symmetry occurs due to parts reflected about the vertical axis). To address this issue, we conducted an experiment by using both horizontal and vertical objects.

Method

Participants

Eight subjects (four female, aged 20–30 years) participated in this experiment.

Stimuli

We used a subset of six parts from Experiment 1 and created two sets of 36 objects from these parts: one with objects in a horizontal orientation and the other with objects in a vertical orientation.

Procedure

We measured visual search dissimilarity for ${}^{36}C_2 = 630$ pairs of objects in each set with two repeats per condition. Trials containing horizontal objects were randomly interleaved with trials containing vertical objects. Subjects had to perform a total of 2,520 (2 sets $\times$ 630 conditions $\times$ 2 repeats) correct trials. All other details are the same as in Experiment 1.

Data analysis

As in the previous experiments, we fit a linear part summation model with 46 parameters (15 part dissimilarities each at corresponding, across, and within object locations and a constant term) to the observed data. To account for saturation in search RTs, we transformed the model predictions using a sigmoid function. In addition to the linear model, we also tested a model with extra nonlinear terms. For horizontal objects, this model did not perform significantly better than the linear model even though it had more free parameters: $r = 0.87$ for linear model, $r = 0.88$ for nonlinear model, $p = 0.12$, $F(105, 477) = 1.19$ for a partial F test comparing the two models.

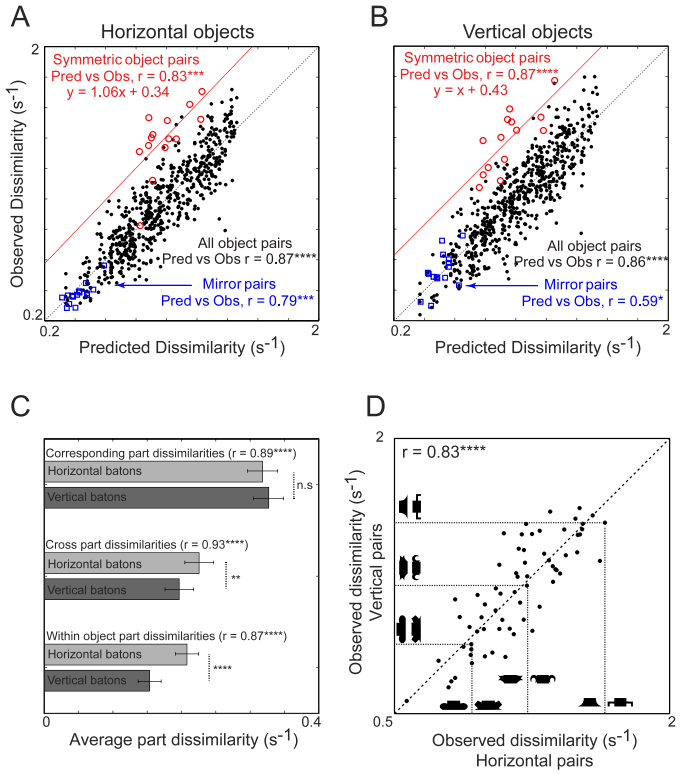

Figure 4. Horizontal and vertical objects (Experiment 3). (A) Observed dissimilarity versus predicted dissimilarity plot for pairs of horizontal objects. The best-fitting line for symmetric objects is shown as a red dashed line. (B) Similar plot as in (A) for pairs of vertical objects. (C) Average part dissimilarities for horizontal objects (light gray) and vertical objects (dark gray). Error bars indicate standard deviation. (D) Observed dissimilarities between vertical symmetric objects plotted against dissimilarities for horizontal symmetric objects. Example object pairs are shown along corresponding axes.

This was true for the vertical object searches as well: $r = 0.86$ for the linear model, $r = 0.87$ for the nonlinear model, $p = 0.063$, $F(105, 477) = 1.25$ for a partial F test comparing the two models. We also confirmed that the linear part summation model was not overfitting the data by performing cross-validation as detailed in Experiment 1 (cross-validated model correlation: $r = 0.84 \pm 0.02$ and $r = 0.84 \pm 0.03$ for horizontal and vertical objects, respectively). We did not analyze search asymmetries as only six out of 1260 object pairs showed significant effect of asymmetry (4/630 and 2/630 pairs for horizontal and vertical objects, respectively).

Results

Subjects were highly consistent in their search performance (corrected split-half correlations across search dissimilarities [mean $\pm$ SD]: $r = 0.84 \pm 0.01$ and $r = 0.83 \pm 0.01$). There was a strong correlation between search times across object pairs between the horizontal and vertical orientations, suggesting that object distances are fundamentally unaltered by overall orientation ($r = 0.80$, $p < 0.00005$). However, horizontally oriented pairs were slightly harder in visual search compared to vertically oriented pairs (median RTs: 1086 ms for horizontal, 992 ms for vertical, $p < 0.00005$, Wilcoxon signed rank test). Importantly, however, the part summation model produced excellent predictions for both horizontal and vertical objects: $r = 0.87$, with $F(45, 585) = 37.07$, $r^2 = 0.76$ for horizontal objects; $r = 0.86$, $F(45, 585) = 33.72$, $r^2 = 0.74$ for vertical objects; $p < 0.0005$ (Figure 4A, B). Distances between symmetric objects were systematically different from model predictions by a constant offset for both object orientations with no obvious difference in the amount of offset (offsets: 0.34 for horizontal, 0.43 for vertical; $p = 0.32$, Wilcoxon ranksum test on 15 bootstrap-derived offset estimates; Figure 4A, B). Mirror pairs were harder in horizontal orientation when compared to vertical orientation (Mean RT: 2.73 s for horizontal, 2.03 s for vertical), $t(28) = 3.8$, $p < 0.00005$, unpaired t test. This is in agreement with previous reports that mirror confusion is stronger about the vertical axis (Rollenhagen & Olson, 2000). Finally, we compared part relations for horizontal and vertical orientations to elucidate why distances were larger in the horizontal orientation. Part relations at corresponding locations did not differ in magnitude between horizontally and vertically oriented objects. However, part relations at opposite locations were slightly weaker for vertical objects, and part relations at within-object locations were substantially weaker for vertical objects (Figure 4C). According to the model, then, vertical objects are more distinct because within-part relations are weaker. We conclude that part matching is not isotropic and occurs preferentially along the horizontal direction compared to the vertical direction.

Although the above results indicate that the constant offset due to symmetry does not differ between horizontal and vertical objects, this finding is based on a small number of observations (${}^6C_2 = 15$ distances between six symmetric objects). To further confirm this negative result, we performed an additional experiment in which we took a larger number of symmetric objects ($n = 12$ objects) and measured all 66 pairs of distances between these objects in visual search for seven subjects (four females, aged 20–30 years). We reasoned that any difference in the strength of symmetry between horizontal and vertical objects should manifest as a systematic difference in the observed dissimilarity. We

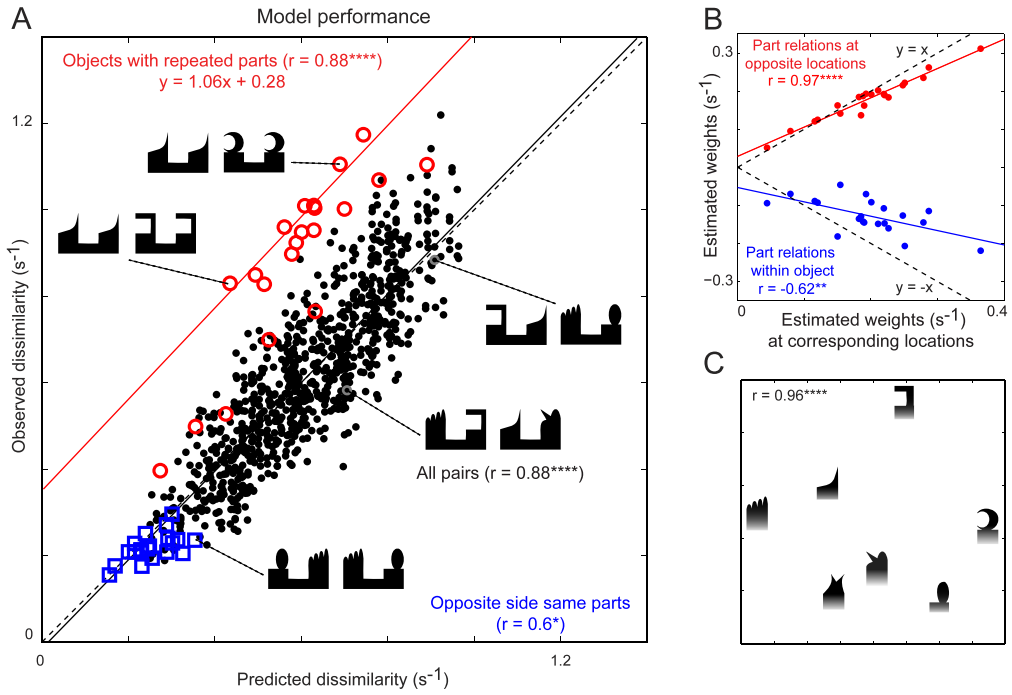

Figure 5. Symmetry versus repetition of parts (Experiment 4). (A) Observed dissimilarity plotted against the model predictions for all 1,176 object pairs. Some example object pairs are shown. Pairs involving objects with repeated parts are highlighted in red circles whereas pairs involving objects that share parts on opposite locations are shown as blue squares. (B) Part relations at opposite locations (red) and within-object locations (blue) plotted against part relations at corresponding locations. Solid lines indicate the corresponding best-fitting lines. All part relations are significantly correlated but vary in magnitude. (C) Two-dimensional embedding of part relations at corresponding locations, showing the distance relationships between parts that ultimately drive object dissimilarity.

observed no such difference: Distances in both conditions were strongly correlated ($r = 0.83$, $p < 0.00005$) and did not differ in magnitude (median distances: 1.17 s^{-1} and 1.19 s^{-1} for horizontal and vertical, $p = 0.79$, Wilcoxon signed-rank test, Figure 4D). We conclude that the effect of symmetry does not depend on object orientation.

Experiment 4: Symmetry versus repetition

In Experiment 1, we found that objects with bilateral symmetry were systematically more distinct in visual search compared to model predictions. This effect could be specific to bilateral symmetry or alternatively arise due to repetition of parts in an object. To distinguish between these possibilities, we created a new set of objects from the same parts as in Experiment 1 except that the stem was U-shaped to allow each part to be placed at either end in the same orientation (Figure 5A). This created asymmetric objects with parts that could be repeated within the object.

Method

Participants

Eight subjects (two female, aged 22–33 years) participated in the experiment. All other details were identical to Experiment 1.

Stimuli

A total of 49 objects were created from two parts placed on either end of a U-shaped stem. This allowed parts at either end to be in the same orientation unlike in Experiment 1 in which they had to undergo mirror reflection. The parts were the same as those used in Experiment 1. Each part measured 1° wide by 1° tall (as in Experiment 1), and the center-to-center distance between parts was 2°. The entire object therefore was 3° in width.

Procedure

All other details were identical to Experiment 1.

Data analysis

We fit the linear part summation model and confirmed that the model is not overfitting the data

(cross-validated correlation: $r = 0.86 \pm 0.01$). In addition to the linear model, we also tested a model with extra nonlinear terms. The linear model was not significantly different from the nonlinear model: $r = 0.88$ for linear model and $r = 0.89$ for nonlinear model, $p = 0.092$, $F(210, 900) = 1.15$ for a partial F test comparing the two models. For both models, we transformed the predicted dissimilarities using a sigmoid function to account for saturation in the search dissimilarities. We did not analyze the effect of search asymmetry on model performance as only 12 out of 1,176 pairs showed significant search asymmetry.

Results

As in Experiment 1, subjects performed visual search on 1,176 ($^{49}C_2$) object pairs and were highly consistent in their performance (average corrected split-half correlation between dissimilarities across two random groups of subjects: $r = 0.85 \pm 0.01$, $p < 0.00005$). We then fit the model to the observed dissimilarities as before and obtained striking fits ($r = 0.88$, $F(63, 1113) = 53.52$, $r^2 = 0.77$, $p < 0.00005$; Figure 5A). Estimated part relations at corresponding locations were significantly correlated with relations at opposite locations ($r = 0.97$, $p < 0.00005$) and within objects ($r = -0.62$, $p < 0.005$), suggesting that there is a common set of underlying part relations modulated by location (Figure 5B). Part relations were visualized using multidimensional scaling (Figure 5C). These part relations were similar to part relations in Experiment 1 ($r = 0.79$ between corresponding location terms, $r = 0.85$ for opposite locations, and $r = 0.83$ for within-object terms, $p < 0.00005$ in all cases).

Having established that the model performs well on these objects, we then investigated how dissimilarities between objects with repeated parts are explained by the model. We observed the same pattern as in Experiment 1: Model predictions were strongly correlated with the observed dissimilarities for repeated-part objects ($r = 0.88$, $p < 0.00005$) but were offset by a fixed amount. The slope of the best fitting line between observed and predicted dissimilarities did not differ significantly from one (slope = 1.06 with [0.64, 1.49] as 95% confidence interval), and the offset was significantly different from zero (offset = 0.28 with [0.013, 0.54] as 95% confidence interval). This offset was slightly smaller than the offset observed for symmetric objects in Experiment 1 (0.3 vs. 0.28) but cannot be directly compared because the data is from two different groups of subjects. Nonetheless, to compare the offsets with this caveat, we calculated the difference between the observed dissimilarity and the model prediction for each of the 21 repeated/symmetric object pairs and compared them using an unpaired t test. This revealed a difference that approached significance ($t(40) = 1.75$, $p = 0.087$, unpaired t test).

We conclude that objects with repeated parts are systematically more distinct than expected by the part summation model and that the magnitude of this effect is comparable to that of symmetry. Thus, the effect of bilateral symmetry observed in Experiment 1 might arise due to part repetition rather than bilateral symmetry per se. However, bilateral symmetry might still play a role: For instance, subjects take equally long to detect symmetry or repetition of two displays that are far apart but are faster on symmetry when the displays are joined together (Corballis & Roldan, 1974). More generally, repetition can be considered as a translational symmetry. Importantly, for the purposes of the study, both repetition and symmetry cause objects to become more distinctive than predicted by the part summation model. The fact that they make objects more distinctive by a fixed offset implies that both repetition and symmetry combine additively with local features.

Experiment 5: Three-part objects

The results of Experiments 1–4 show that part relations sum linearly and that part relations at opposite locations have a weaker contribution compared to those at corresponding locations. This could be because parts at opposite locations are further apart compared to parts at corresponding locations, implying that part relations vary with distance. Alternatively, this could be because corresponding locations are fundamentally different from opposite locations in an object. If part relations do not vary with distance, it implies a "bag-of-parts" matching process in which all part pairs are compared in the two objects regardless of location. If part relations vary with distance, it implies a spatially tuned matching process.

To address this issue, we created objects with three parts that vary in their distance from each other. If part relations decay with distance, relations (of the same kind) between nearby parts should have a greater contribution to the overall dissimilarity than relations between far away parts.

Method

Participants

Ten subjects (four female, aged 15–30 years) participated in the visual search experiment. All other details are similar to Experiment 1.

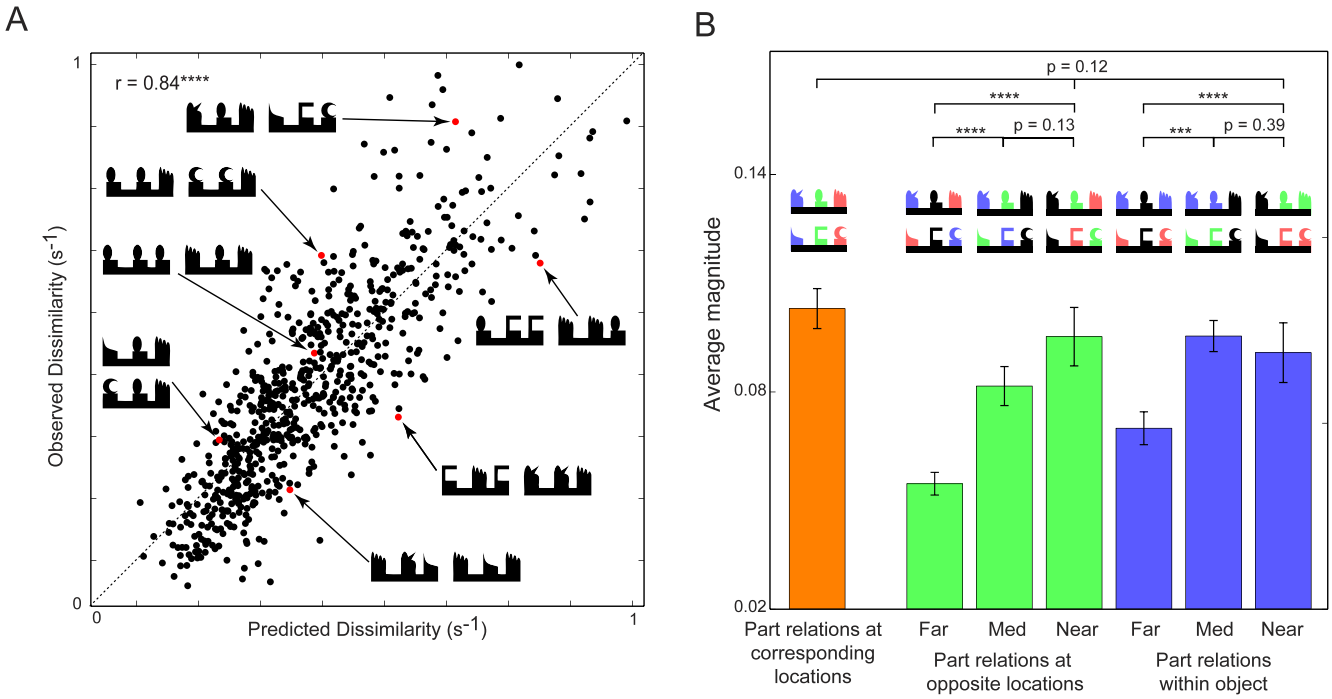

Figure 6. Effect of distance on part relations (Experiment 5). (A) Observed dissimilarity versus predicted dissimilarity for objects with three parts. Example object pairs are indicated in red. (B) Magnitude of part relations (mean ± *SEM*) at different distances and spatial locations. Asterisks represent statistical significance of the comparison in magnitude for various pairs of terms with conventions as before. Example part comparisons are shown above each bar. Purple, green, and pink colors indicate parts being compared in each case.

Stimuli

The stimuli were three-part objects with each part attached to the branches of a W-shaped stem (Figure 6A). We chose the location of the branch point on the stem such that two parts were close together and the third part was twice as far away. There were six possible parts at each location with the result that there were 216 possible objects and a huge number ($^{216}C_2 =$ 23,220) of possible object pairs.

Procedure

Out of the possible 23,220 pairs of objects, we chose a subset of 700 random pairs and measured perceived distances as in previous experiments using visual search.

Data analysis

We asked whether observed distances between objects could be accounted for by a model in which part relations sum linearly. However, for the object pairs used in this experiment, the part relations could be between parts at three possible distance levels: near (i.e., between the near and middle parts), medium (i.e., between the middle and far parts), and far (i.e., between the near and far parts). Accordingly, we grouped part relations into corresponding, opposite, and within-object terms with three subgroups (near, medium, and far distance parts) of terms within each group. Each set of part relations contained $^6C_2 = 15$ terms, and the complete model contained a total of 106 (seven groups of 15 terms each and a constant term). The predicted dissimilarities from the linear part summation model were transformed using a sigmoid function to account for saturation in search dissimilarities (1/RT). We also confirmed that the model was not overfitting using a cross-validated measure of performance (mean ± *SD* of cross-validated correlation: $r = 0.78 \pm 0.04$). Further, we found that there were only six out of 700 pairs that exhibited significant search asymmetry. Hence, we did not explore the effect of search asymmetry on model performance in detail.

Results

We collected visual search data for 700 object pairs from 10 subjects. The subjects were highly consistent in their dissimilarities (average corrected split-half correlation between two random groups of subjects [mean ± *SD*]: $r = 0.83 \pm 0.01$, $p < 0.00005$). Upon fitting the observed dissimilarities using the part summation model, we obtained striking fits ($r = 0.84$, $F(105, 595) = 12.69$, $r^2 = 0.71$, $p < 0.00005$; Figure 6A). The estimated part relations yielded several interesting insights. First, all groups of part relations were significantly correlated

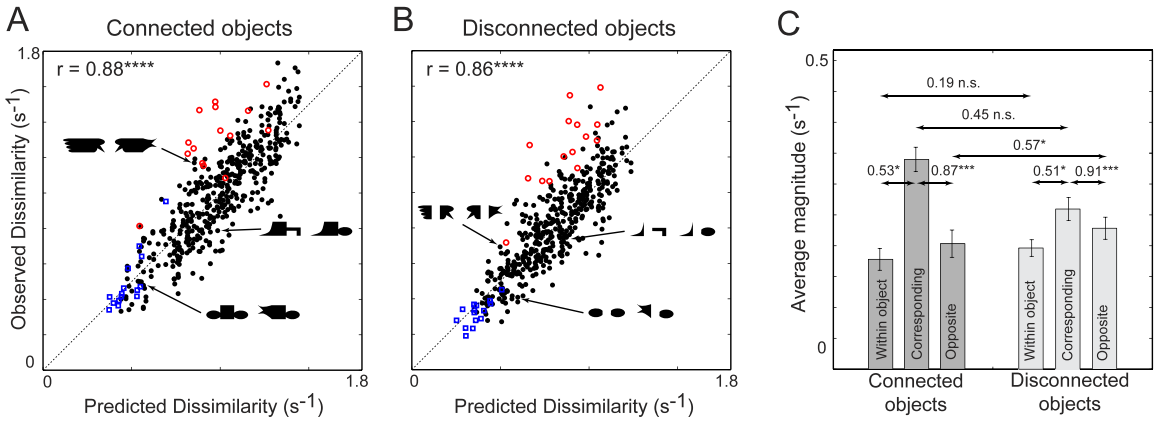

Figure 7. Disconnected parts (Experiment 6). (A) Observed dissimilarity versus predicted dissimilarity for pairs of connected objects. Object pairs with global attributes are highlighted (symmetric object pairs: red, mirror-related pairs: blue). (B) Similar plot as in (B) for pairs of disconnected objects. (C) Average magnitude of part relations for connected objects (dark gray) and disconnected objects (light gray). Numbers indicate pair-wise correlation coefficients. Error bars indicate standard deviation.

(median correlation between groups: $r = 0.81$, median p value: $p = 0.000051$), suggesting that there is a common set of part relations that is modulated by location. Second, the magnitude of the part relations estimated at the different locations varied systematically (Figure 6B): Part relations at corresponding locations were strongest as before compared to all other terms and approached significance for some comparisons (Figure 6B). Importantly, the magnitude of part relations for the far part was systematically smaller than the near and medium parts for both opposite and within-object location terms. We conclude that part matching is spatially tuned and decays with distance.

Experiment 6: Disconnected parts

The results in the preceding sections were based on objects whose part identity was unambiguous. In Experiments 6–8, we manipulated part identity systematically. In this experiment, we investigated how part relations change when parts are disconnected from each other.

Method

Participants

Seven subjects (four females) took part in the experiment. All other details are similar to Experiment 1.

Stimuli

We selected a subset of six parts from Experiment 1 and created 36 two-part objects from it. We then created two variants for each object: the normal one in which the two parts were connected by a stem (Figure 7A) and a variant one in which the stem was deleted and the two parts were now spatially separated by the same distance as before (Figure 7B).

Procedure

Subjects performed visual search task on 630 ($^{36}C_2$) pairs of connected objects and 630 pairs of disconnected objects. The trials involving connected objects were randomly interleaved between trials involving disconnected objects. In case of searches involving disconnected objects, the spacing between items in the array (3°) was larger than the separation between the two parts (1°). This ensured that the two isolated parts still grouped together by spatial proximity cues.

Data analysis

We fit a linear part summation model to the observed data as explained in the previous experiments. We confirmed that the model was not overfitting using cross-validation (average cross-validated correlation: $r = 0.87 \pm 0.03$ and $r = 0.85 \pm 0.02$ for connected and disconnected objects, respectively). For connected objects, the linear part summation model was not significantly different from a model with extra nonlinear terms: $r = 0.88$ for linear model and $r = 0.91$ for nonlinear model, $p = 0.07$, $F(105, 477) = 1.24$ for a partial F test comparing the two models. This was true for disconnected objects as well: $r = 0.86$ for linear model and $r = 0.89$ for nonlinear model, $p = 0.31$, $F(105, 477) = 1.07$ for a partial F test comparing the two models. Finally, very few searches ($n = 5$)

exhibited a statistically significant asymmetry across both groups, so we did not analyze them separately.

Results

The subjects were extremely consistent in their dissimilarities (average corrected split-half correlation between two random groups of subjects [mean ± *SD*]: $r = 0.86 \pm 0.01$ for connected objects and $r = 0.83 \pm 0.01$ for disconnected objects, $p < 0.00005$). To assess how dissimilarities change with stem deletion, we compared search dissimilarities for every pair of objects when the parts were connected versus when they were separated. This revealed a strong positive correlation ($r = 0.7$, $p < 0.00005$). However, searches involving connected objects were slightly easier than searches involving disconnected objects (average search times: 1118 ms for connected, 1382 ms for disconnected objects, $z = 15.64$, $p < 0.00005$, ranksum test). We then fit the model containing corresponding, within, and across terms to the dissimilarities for normal and stem-deleted object pairs. Model performance was equally good for connected objects ($r = 0.88$, $F(45, 585) = 36.89$, $r^2 = 0.77$, $p < 0.00005$ [Figure 7A]), and for disconnected objects ($r = 0.86$, $F(45, 585) = 33.43$, $r^2 = 0.74$, $p < 0.00005$ [Figure 7B]). These correlations were not significantly different ($p = 0.14$, Fisher z test). Here too, symmetric objects were systematically more distinct by a constant offset (best-fitting slope: 1.05 with a 95% confidence interval [0.52 1.59]; intercept: 0.34 with a 95% confidence interval [0.12 0.56]).

Part relations at corresponding locations were strongly correlated with those at opposite and within-object locations for both connected objects ($r = 0.87$ and $r = -0.53$, $p < 0.05$ for connected objects; Figure 7C) and for disconnected objects ($r = 0.91$ and $r = -0.51$, $p < 0.05$; Figure 7C). Thus, both connected and disconnected objects have consistent part relations across locations. However, part relations for connected and disconnected objects were only weakly correlated and often not even statistically significant ($r = 0.45$, 0.57, and 0.19 for connected vs. disconnected part relations at corresponding, opposite, and within-object locations, $p = 0.096$, 0.025, and 0.5, respectively; Figure 7C). Thus, part relations are fundamentally different for isolated and embedded objects but sum linearly in both cases.

Experiment 7: Natural parts

The results of the Experiments 1–5 were based on objects with operationally defined parts joined by a stem. In this experiment, we asked whether part relations depend on the specific manner in which objects are broken down into parts. Would our results hold even for any breakdown of objects into fragments? Are some fragments of an object better than others in explaining its relation to other objects?

To address this issue, we drew upon the existing literature in object vision showing that we readily perceive objects to contain certain parts but not others: For instance, people are faster to report that an "8" contains an "o" than a "3" (Hoffman & Singh, 1997; Palmer, 1999). However, this pattern was observed while subjects were explicitly instructed to judge whether fragments belong to objects. Thus it is unclear whether part decomposition occurs in natural vision in the absence of explicit part judgments.

Method

Participants

Twelve subjects (two females) participated in the experiment. All other details are similar to Experiment 1.

Stimuli

To address how part relations depend on the manner in which parts are defined, we designed a set of seven objects that could be broken down into natural or unnatural fragments and recombined to form other objects (Figure 8A). We chose different parts on either end to avoid creating symmetric objects. This resulted in two sets of objects that had these seven objects in common. The first set comprised objects broken down into unnatural fragments (Figure 8B). The second set comprised objects broken down into natural fragments (Figure 8B). Instead of running all pairs of objects in both sets, we sampled five 4×4 object matrices along the diagonal of the full 7×7 object matrix and chose all possible pairs of objects within each 4×4 object matrix. Our particular choice of the matrices led to 492 unique object pairs. This design ensured that all part pairs were sampled sufficiently without having to run all possible object pairs (${}^{49}C_2 = 1{,}176$).

Procedure

We measured perceived distances between the chosen object pairs using visual search. Trials from both sets of objects were randomly interleaved.

Data analysis

We used a linear part summation model with 43 parameters to fit the data (detailed below). This model

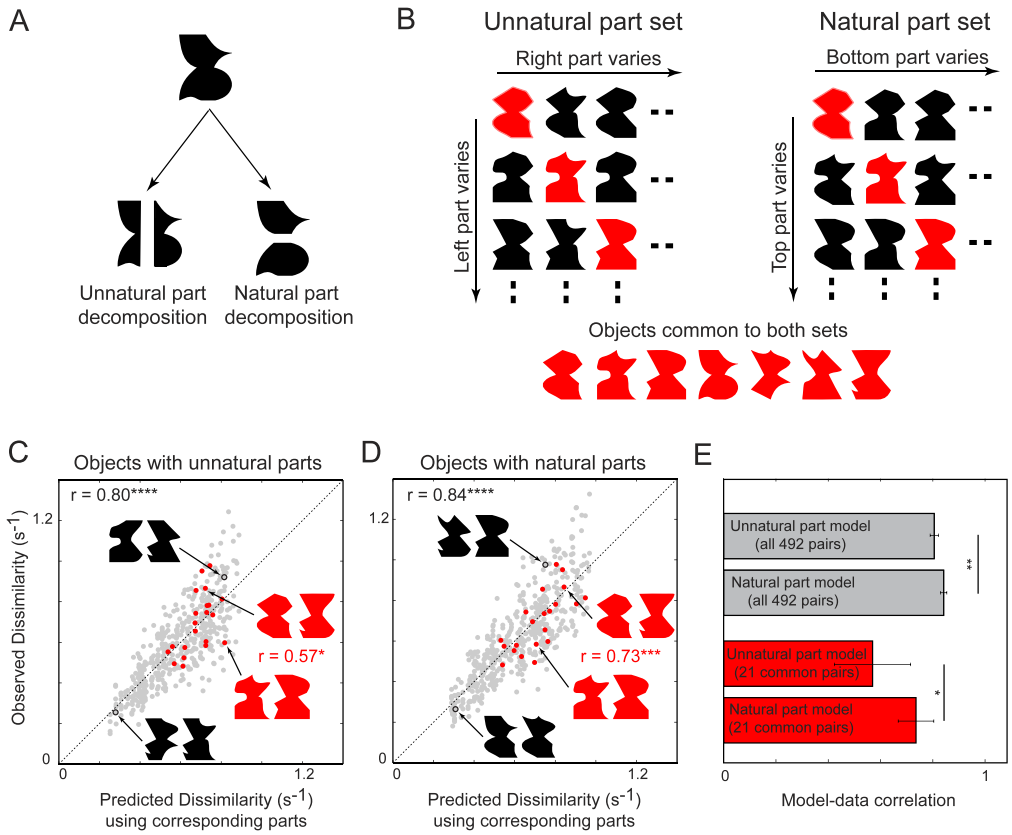

Figure 8. Natural and unnatural parts (Experiment 7). (A) An example object broken into its unnatural parts (left) and natural parts (right). (B) Object sets used in the experiment. A subset of the unnatural part set (left) and natural part set (right). Each set contained a total of 49 objects generated from seven possible parts on either side. Objects common to both sets are shown in red. (C) Observed dissimilarity versus predicted dissimilarity for pairs of objects with unnatural parts. Example object pairs unique to this set are shown in black with the corresponding points marked as black circles. Pairs corresponding to common objects are shown in red. (D) Similar plot as in (B) for pairs of objects with natural parts. (E) Correlation between predicted and observed dissimilarities for both sets of objects on all pairs of objects (gray) and common pairs of objects (red).

did not overfit the data as evidenced by a high cross-validated correlation ($r = 0.77 \pm 0.03$ and $r = 0.81 \pm 0.02$ for unnatural and natural part sets, respectively). The incidence of search asymmetries was very low (20/492 and 14/492 searches for unnatural and natural sets, respectively), and hence, we did not explore this further.

Results

To address how part relations depend on the manner in which parts are defined, we designed a set of seven objects that could be broken down into natural or unnatural fragments and recombined to form other objects (Figure 8A). We chose different parts on either end to avoid creating symmetric objects. This resulted in two sets of objects that had these seven objects in common. The first set comprised objects broken down into unnatural fragments (Figure 8B). The second set comprised objects broken down into natural fragments (Figure 8B). The dissimilarities between the seven common objects in both sets could therefore be predicted using the natural fragments or by the unnatural fragments. This allowed us to ask whether object dissimilarities were explained better by their natural parts or their unnatural parts.

Subjects performed visual search on a total of 984 (492 conditions × 2 natural/unnatural) conditions and were extremely consistent with each other (average corrected split-half correlation between dissimilarities across two random groups of subjects [mean ± *SD*]: $r = 0.89 \pm 0.02$ for unnatural and $r = 0.89 \pm 0.02$, $p < 0.00005$). Here also, we tried to fit the part summation model to the observed data. Because the parts on either end were different in identity, we were only able to use part relations at corresponding locations in the model. Hence the model had only 43 free parameters (21 part relations each on left and right sides of the object and a constant term).

Model performance was slightly worse on the unnatural part set ($r = 0.80$, $F(42, 450) = 19.34$, $r^2 = 0.64$, $p < 0.0005$; Figure 8C), compared to the natural part set ($r = 0.84$, $F(42, 450) = 25.18$, $r^2 = 0.71$, $p < 0.0005$; Figure 8D). This difference was statistically significant as assessed using bootstrap resampling of the correlation coefficients ($p < 0.0005$, Wilcoxon signed-rank test on 492 bootstrap-derived estimates of correlations). Across all bootstrap-derived samples, the natural part model correlations were higher than the unnatural part model correlations about 99% of the time (Figure 8E). However this comparison is based on different sets of objects, and the difference might be due to the objects being different rather than because of natural or unnatural fragments. We therefore compared the two models on the 21 object pairs common to both sets. The natural fragment model performed slightly but significantly better than the unnatural fragment model ($r = 0.73$ for natural, $r = 0.57$ for unnatural; $z = 3.79$, $p < 0.005$, Wilcoxon's ranksum test based on 21 bootstrap derived estimates of correlations; Figure 8E). Across all bootstrap-derived samples, the natural part model correlations were higher than the unnatural part model correlations about 90% of the time. Because the unnatural part model was still reasonably successful in explaining perceived distances, we surmise that the underlying process involves contour matching rather than part matching. However, using natural parts confers a slight advantage in explaining object distances. We conclude that the contour matching process is modulated by part decomposition but not determined by it.

Experiment 8: Holistic objects

In the previous experiment, we showed that the manner in which an object is broken down into parts influences the quality of fit between part relations and object relations. In this experiment, we wondered whether the part summation model would work for objects that cannot be reduced into any perceptually obvious parts (hereafter denoted as holistic objects). We were motivated by our observation that objects such as circles or squares contain no perceptually obvious parts and appear instead as irreducible wholes. To investigate this issue, we created a set of objects that appeared to us to contain no distinctive or perceptually obvious parts (Figure 9A). More formally, these objects contain no salient curvature minima that would lead to unambiguous part parsing (Hoffman & Singh, 1997). Using this set of objects, we asked can dissimilarities between holistic objects be understood in terms of their parts?

Method

Participants

Seven subjects (two female) participated in this experiment. All other details are similar to Experiment 1.

Stimuli

We created seven contours that could be placed on either the left or right side to make a set of 49 holistic objects (Figure 9A).

Procedure

Subjects performed visual search on all possible 1,176 ($^{49}C_2$) pairs of holistic objects.

Data analysis

We fit a linear part summation model with 64 parameters as detailed in Experiment 1. The model did not overfit the data as evidenced by a high cross-validated correlation ($r = 0.88 \pm 0.01$). We also found that the linear model was not significantly different from a nonlinear model: $r = 0.89$ for linear model and $r = 0.9$ for nonlinear model, $p = 1$, $F(210, 900) = 0.22$ for a partial F test comparing the two models. In addition, the incidence of search asymmetry was very low (four out of 1,176 pairs), and hence, we did not explore this further.

Results

We collected visual search data for all 1,176 pairs of holistic objects. Subjects were extremely consistent in their dissimilarities (average corrected split-half correlation between two random groups of subjects [mean $\pm$ SD]: $r = 0.87 \pm 0.01$, $p < 0.00005$). As before, we fit the part summation model to the observed data and obtained excellent fits ($r = 0.88$, $F(63, 1113) = 50.56$, $p < 0.00005$, $r^2 = 0.77$; Figure 9B). Here too, part relations were consistent across locations (corresponding vs. opposite: $r = 0.91$, $p < 0.00005$ and corresponding vs. within-object: $r = 0.79$, $p < 0.00005$; Figure 9C). Thus, distances between holistic objects also can be understood in terms of their parts. Likewise, symmetric objects were systematically more distinct by a constant offset (best-fitting slope: 0.9 with a 95% confidence interval [0.63 1.18]; intercept: 0.47 with a 95% confidence interval [0.11 0.82]).

The results of the previous experiment are counter-intuitive because they suggest that distances between holistic objects can also be explained using their parts. There are two possible interpretations: First, it is

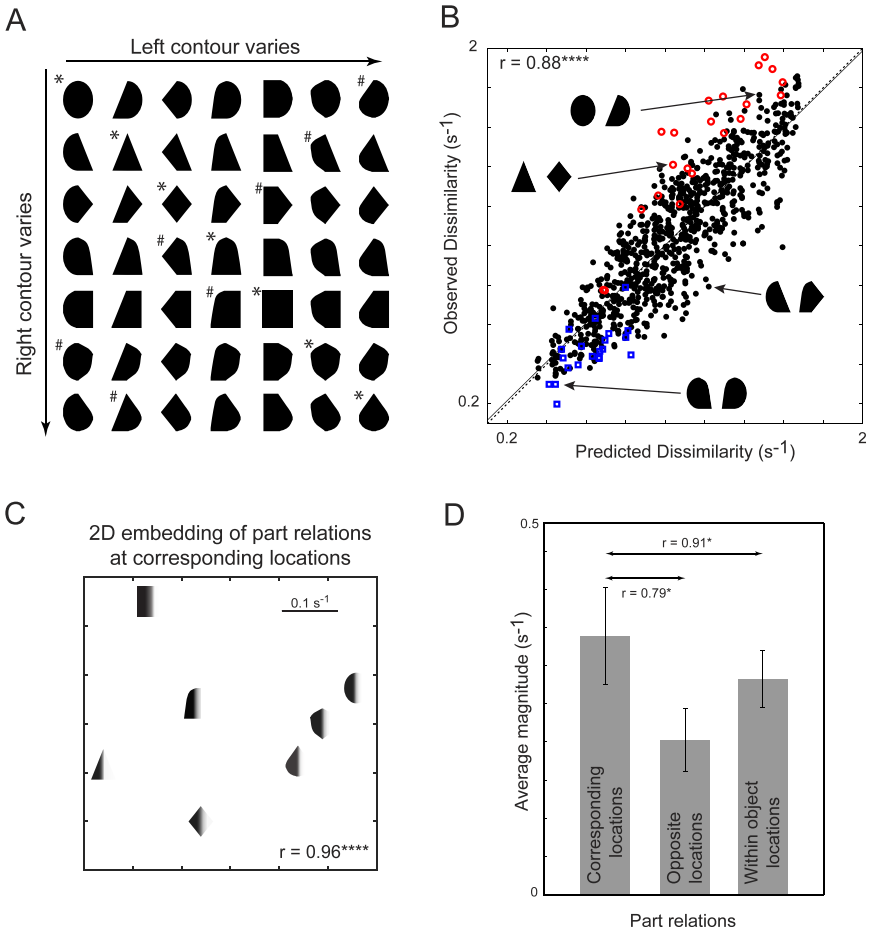

Figure 9. Holistic objects (Experiment 8). (A) The 49 holistic objects used in the experiment. * and # indicate symmetric and asymmetric objects used in the context experiment (Experiment 9). (B) Observed dissimilarities plotted against predicted dissimilarities. Pairs of symmetric objects and mirror-related objects are indicated in red and blue, respectively. (C) Two-dimensional embedding of part relations at corresponding locations. The correlation coefficient represents the correlation between the estimated part relations and the 2-D distances in this plot. (D) Average magnitudes of estimated part dissimilarities for different part relations and their correlations. Error bars denote standard deviation.

possible that the salience signal that drives searches involving holistic objects depends on a contour-matching process that is independent of whether the object is decomposed into parts. Second, it is possible that the viewing of holistic objects with other objects that share their parts resulted in the objects themselves being perceived as containing parts. This would imply that the perception of these objects as being holistic is pliable by the context in which they occur. We assessed this possibility in Experiment 9.

Experiment 9: Context

The findings of Experiment 8 show that dissimilarities between holistic objects can be explained using their parts. However the searches in Experiment 8 frequently involved holistic objects that shared a part of their contour with the distracters, and this context could have caused subjects to perceive the holistic objects as containing these shared parts. To assess this possibility, we measured dissimilarities between the holistic objects in an independent group of subjects who never saw objects with shared parts.

Method

Participants

Eight subjects (one female) participated in the experiment. Two of these subjects had participated in Experiment 8 (holistic objects), but their data did not differ qualitatively with the other subjects. Therefore, the analyses below are based on data from all subjects.

Stimuli

We chose all seven vertically symmetric objects and seven asymmetric objects from Experiment 8 (marked in Figure 9) along with seven two-part objects from Experiment 1. The asymmetric and two-part objects were chosen such that none of the parts were shared between two objects on the same side.

Procedure

Subjects performed visual search involving pairs of objects in each stimulus set in separate blocks that were counterbalanced across subjects. Each block consisted of searches involving 21 object pairs with two trials each as before. In all, subjects performed 126 correct trials of visual search (3 blocks × 42 trials per block). All other details are as before.

Results

The main goal of this experiment was to test whether the experimental context of viewing objects with shared parts influenced the perceived dissimilarity between objects. To this end, we measured the dissimilarities of three groups of seven objects each: holistic objects with vertical symmetry (marked with a star in Figure 9), holistic asymmetric objects (marked with a # in Figure 9), and two-part objects from Experiment 1. Importantly, objects in each group were chosen so that they did not share any part or contour with other objects.

Subjects were highly consistent in their search performance across all three groups of searches (corrected split-half correlations: symmetric object pairs: $r = 0.92 \pm 0.03$, $p < 0.00005$; asymmetric pairs: $r = 0.94 \pm 0.01$, $p < 0.00005$; two-part object pairs: $r = 0.78 \pm 0.06$, $p < 0.00005$). We then asked whether predicted dissimilarities from the part summation model (based on models estimated in the previous experiments from an independent set of subjects) would predict the dissimilarities observed in this experiment. Model predictions were striking across all three groups with no qualitative difference (model-data correlations: symmetric pairs: $r = 0.88$, $r^2 = 0.77$; asymmetric pairs: $r = 0.92$, $r^2 = 0.85$; and two-part pairs: $r = 0.88$, $r^2 = 0.77$; all correlations $p < 0.00005$). Thus, object dissimilarities were unaffected by the experimental context in which objects were observed. We conclude that holistic objects are explained by the part summation model because the dissimilarities are fundamentally driven by a contour-matching process.

Experiment 10: Shape and texture

The results of the preceding sections were all based on testing shape attributes. Here, we wondered whether this result would generalize to other object properties as well. Specifically, we asked whether dissimilarities between objects that differed in both shape and texture could be explained using shape and texture dissimilarities.

Method

Participants

Eight subjects (four female) participated in the experiment. All other details were identical to Experiment 1.

Stimuli

A total of 36 stimuli were created by combining six shapes with six textures in a combinatorial fashion. This resulted in a total of ${}^{36}C_2$ search conditions involving every pair of stimuli. We also measured dissimilarities between all possible pairs (${}^{6}C_2 = 15$) of shapes and between all possible pairs of textures (${}^{6}C_2 = 15$). For the shape-only conditions, shapes were shown as silhouettes with a uniform white fill. For the texture-only pairs, the textures were shown as squares filled with the corresponding textures. Examples of these searches are shown in Figure 8.

Procedure

All other details were identical to Experiment 1.

Data analysis

We used a linear model with 31 parameters (15 shape dissimilarities, 15 texture dissimilarities, and a constant term) to fit the observed data. We confirmed that the linear model is not overfitting the data (cross-validated correlation coefficient: $r = 0.9 \pm 0.01$). In addition, we also found that there was only one search asymmetry and, hence, did not explore the effect of search asymmetry on model performance.

Results

We measured 630 (${}^{36}C_2$) pair-wise dissimilarities between 36 objects differing in both shape and texture using visual search (Figure 10). In addition, we measured 15 (${}^{6}C_2$) shape–shape dissimilarities and 15 texture–texture dissimilarities to confirm

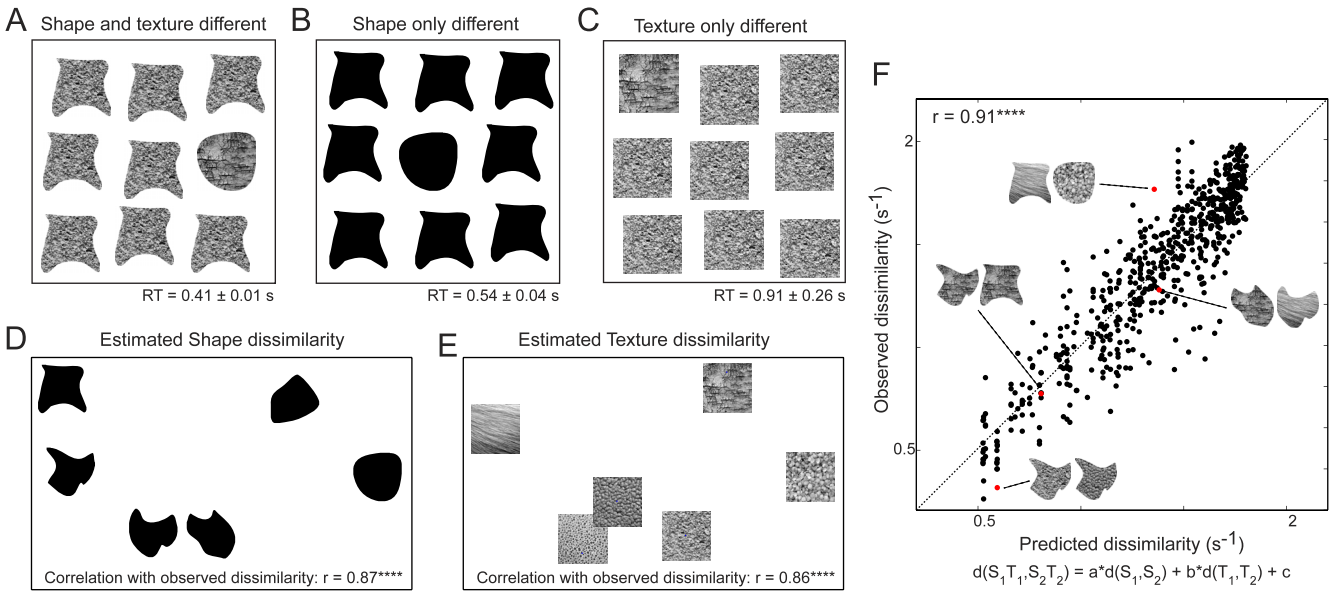

Figure 10. Shape and texture sum linearly (Experiment 10). (A) Example search array with a target differing in both shape and texture. Search reaction time (mean ± *SD*) is shown at the bottom right. In the experiment, we used a 4 × 4 array of objects. (B) Example search with a target differing only in shape. In the actual experiment, all shapes were white on a black background. (C) Example search with the target differing only in texture. (D) Two-dimensional embedding of shape relations estimated by the model. These shape–shape distances were strongly correlated with the observed shape–shape dissimilarities. (E) Similar plot as in (D) for texture dissimilarities. (F) Observed dissimilarity plotted against predicted dissimilarity for all 630 object pairs.

model predictions. Subjects were extremely consistent in their responses (average split-half correlation between dissimilarities across two random groups of subjects: $r = 0.88 \pm 0.01$, $p < 0.001$). An example search in which the target differed in both shape and texture is shown in Figure 10A. It can be seen that this search is slightly easier than searches in which the target differs only in shape (Figure 10B) or in texture (Figure 10C). Thus, both shape and texture differences combine in visual search, and we set out to investigate the precise functional manner in which they combine using a similar model as before. In the model, the net dissimilarity between two objects different in both shape and texture is the sum of the dissimilarity between the shapes of the two objects and the dissimilarity between the textures of the two objects. The model parameters (31 parameters, 15 each for shape and texture and a constant term) were estimated using linear regression as before. Observed dissimilarities were explained extremely well by the model ($r = 0.91$, $F(30, 630) = 66.11$, $r^2 = 0.83$, $p < 0.001$; Figure 10F). To visualize the underlying shape and texture relations, we performed multidimensional scaling as before. These revealed systematic patterns of shape and texture distances, which underlie the observed dissimilarities (Figure 10D, E). We then compared the shape and texture relations estimated by the model with those observed using the shape-only and texture-only conditions in visual search. These model parameters were strongly correlated with their observed counterparts ($r = 0.87$ for shape–shape dissimilarities; $r = 0.86$ for texture–texture dissimilarities; $p < 0.001$). We conclude that shape and texture sum linearly in object vision.

Experiment 11: Global properties I

In Experiment 1, we found that symmetry—a global attribute—combines additively with local part relations. In this and the next experiment, we investigated whether this result would generalize to other global attributes. We took pairs of three-part objects that differed in at least one part. We then asked how the perceived distance between two objects would change when one of the objects differed by a global attribute. We considered two global attributes in this experiment: a change in global orientation and a change in the length of the stem connecting one of the parts to the rest of the object. We consider these properties global because they affect the overall appearance of the object without affecting local part identity. We specifically chose relatively subtle changes in appearance in order to study how these properties combine with local properties in visual search.

Method

Participants

Sixteen subjects (eight female, aged 20–30 years) participated in this experiment.

Stimuli

The stimuli consisted of objects with three parts in a T-like configuration. We selected six of the seven parts used in Experiment 1. Parts on the left and right sides of the object were created as in the previous experiment. In addition, the third part at the bottom of the object was created by 90° clockwise rotation of the right-side equivalent of the part. We selected 60 random pairs of three-part objects for further manipulation. For each pair, we created additional variants in which one object in the pair differed in one of two global attributes: orientation (6° or 9°) or stem length of the middle part (50% or 75% increase). These manipulations can be seen for a few example pairs in Figure 11.

Procedure

All other details were identical to Experiment 1 with the exception that subjects performed four correct trials per search condition (2 with either object as target × 2 with target on the left and right). There were a total of 300 (60 normal + 60 × 4 variant conditions) unique pairs of objects, and subjects had to perform a total of 1,200 correct trials (300 conditions × 4 repetition per condition).

Results

To investigate the impact of global attribute differences on object dissimilarities, we compared search times for object pairs that did or did not differ in global attributes. This allowed us to infer in a model-free manner the impact of global attributes on object pairs differing in local features. Subjects were extremely consistent in their performance (average corrected split-half correlation between dissimilarities across random groups of subjects across all 300 searches: $r = 0.94 \pm 0.01$, $p < 0.001$). We selected for further analysis the object pairs that showed a significant difference in RT between the normal and variant conditions ($p < 0.05$ for a main effect of variant type in an ANOVA with subject and variant type as factors). This yielded 32 pairs for the 6° change and 50 pairs for the 9° orientation and 19 and 39 pairs for the 50% and 75% changes in length. We speculate that the remaining pairs did not show an effect because of signal-to-noise issues—in other words, that they would also show a significant effect with a larger number of individual trials. Testing this systematically is beyond the scope of this study given the limited number of trials ($n = 4$) per search.

For the orientation variants, search times for the variant pairs were consistently smaller than the normal pairs (Figure 11A), indicating that the orientation change contributed to the overall dissimilarity. But interestingly, the perceived distances (1/RT) of the variant pairs differed from the normal pairs by a fixed constant offset that increased with orientation (Figure 11B). To further confirm this trend, we examined the slope and offset of the best-fitting line for both levels of orientation and length after removing four extreme points from each set to prevent the best-fitting line from being biased by the outliers (we obtained qualitatively similar results on varying this choice). The slopes hovered around unity (i.e., with confidence intervals always including one) for changes in orientation as well as stem length (Figure 11C). In contrast, the offset changed systematically with an increase in length or orientation (Figure 11D). The difference in offset was significant for both the orientation change and the length change ($p < 0.001$, Wilcoxon signed-rank test on bootstrap-derived offset values equal in number to the data points).

Our finding that dissimilarity increases systematically with increasing orientation is reminiscent of the classic mental rotation experiments in which subjects take longer to match more rotated views of the same object (Shepard & Metzler, 1971). The systematic change in offset with increasing orientation or with changing stem length implies that there is a global registration process that brings the objects into alignment, and the net dissimilarity is driven both by the amount of alignment required and the local part differences (as evidenced by the correlation between normal and variant pairs).

Experiment 12: Global properties II

Here we investigated how a change in another global property—the relative position of parts in an object—would modify the overall dissimilarity between objects.

Method

Participants

Eight subjects (seven female, aged 20–31 years) participated in this experiment.

Stimuli

The stimuli consisted of objects with two parts joined together in a U-like configuration (Figure 12). We

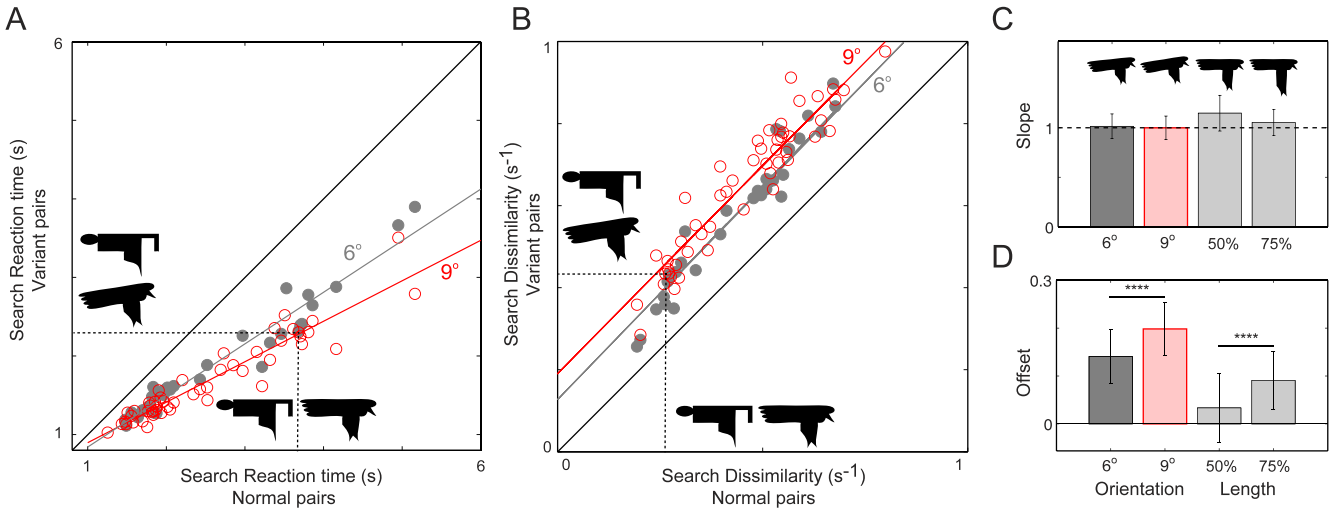

Figure 11. Orientation and stem length combine additively with local features (Experiment 11). (A) Search times for variant object pairs plotted against search times for normal pairs. The pairs are chosen such that the search times for variant object pairs are significantly different from normal pairs ($p < 0.05$). Variant object pairs contained the same local parts as the normal pairs but differed in global orientation that was either 6° (gray filled circles) or 9° (red unfilled circles). Lines of the corresponding color indicate the best-fitting line for the data. (B) Same data as in (A) but now using search dissimilarity (1/RT). The resulting plot shows that introducing a global orientation difference increases dissimilarity by a constant value. This offset is larger for 9° than for 6° differences. (C) Slopes of the best-fitting lines from (B) for two levels of orientation change (dark gray for 6° and pink for 9°) and for two levels of change in length of the middle stem (50% and 75% increase in length, light gray bars). Error bars indicate 95% confidence intervals from the linear fit. (D) Offsets of the best-fitting lines from (B) for the same conditions as in (C).

chose six out of the seven parts used in Experiment 1 to create stimuli. Objects were created by attaching parts measuring 0.6° onto a rectangular base (3.5° × 0.6°). There were two types of object pairs: In each normal object pair, the two objects had parts separated by 1.2°. Each variant pair contained the same parts, but one object had parts separated by 1.2° and the other had parts separated by 2.4°. Thus, the variant pair has the same part differences as the normal pair but, in addition, differed in a global property. We selected 100 normal and variant pairs for further testing.

Procedure

All details were identical to Experiment 1 except that subjects performed eight correct trials per search condition (2 with either object as target × 2 with target on the left and right × 2 repeats). There were a total of 200 object pairs (100 normal, 100 variant), and subjects had to perform a total of 1,600 correct trials (200 conditions × 8 repetitions).

Results

Subjects were extremely consistent in their performance (average corrected split-half correlation between dissimilarities across two random groups of subjects: $r = 0.93 \pm 0.01$, $p < 0.00005$). As in the previous experiment, we selected for further analysis only those objects that showed a significant difference in the search times between normal and variant conditions ($p < 0.05$ for main effect of pair type in an ANOVA with subject and pair type as factors). This procedure yielded 51 pairs of objects. Searches involving variant pairs were always easier than searches involving normal pairs as expected because the variant pairs differed both in local and global attributes (Figure 12A). As observed in the previous experiment, search dissimilarities for variant pairs were greater than search dissimilarities for normal pairs by a constant offset (Figure 12B). The slope of the best-fitting line did not differ from unity (slope = 1.08 with 95% confidence interval [0.99 1.17]), and the intercept was significantly different from zero (intercept = 0.11 with 95% confidence interval [0.08 0.15]). Thus, a change in spatial separation of parts introduced a fixed offset to the dissimilarity already present due to local part differences. The fixed offset shows that this global attribute combines additively with local features.

General discussion

To summarize our findings: We have shown that the dissimilarity between two objects is a linear sum of

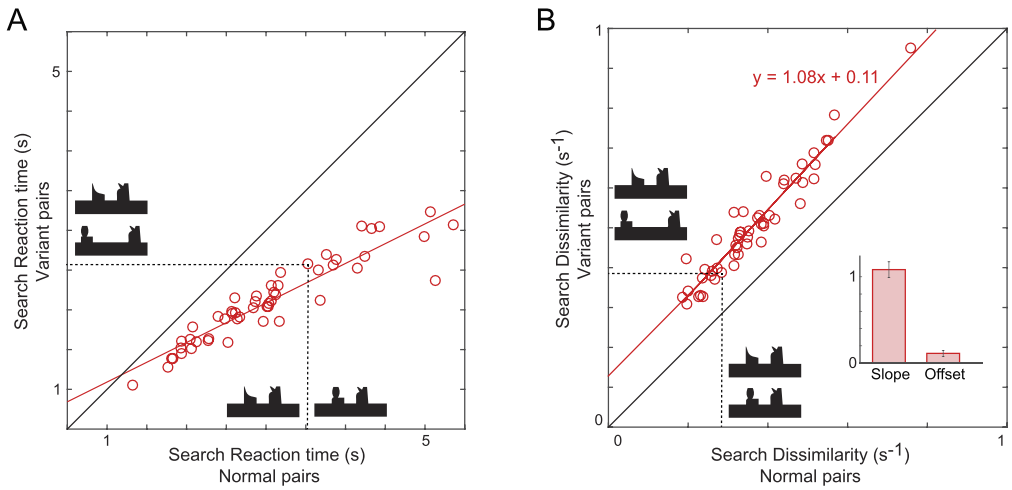

Figure 12. Part location combines additively with local features (Experiment 12). (A) Search times for variant object pairs plotted against search times for normal pairs. Variant object pairs contained the same local parts as the normal pairs but differed in their relative part positions. The red line indicates the best-fitting line for the data. (B) Same data as in (A) but now using search dissimilarity (1/RT). Inset: Slope and offset of the best-fitting line. Error bars indicate 95% confidence intervals.

local feature differences (Experiments 1, 2, and 3). These feature difference operations are spatially tuned as evidenced by weaker influences of more distant parts (Experiment 5). These difference computations are modulated by part decomposition (Experiments 7, 8, and 9). They are fundamentally altered when parts are disconnected, implying that parts behave differently within an object and when isolated (Experiment 6). Symmetry and part repetition make objects appear more distinct than expected through local feature summation and, importantly, combine additively with local feature differences (Experiments 1, 3, and 4). Texture differences add linearly to shape differences (Experiment 10). Finally, we have shown that altering global attributes such as global orientation, stem length, or part position all increase dissimilarity due to local features by a fixed amount, implying that they combine additively (Experiments 11 and 12).

The above findings are indicative of a remarkably simple additive rule by which object attributes combine in visual search. According to this rule, the perceptual distance between two objects is a linear sum of local part differences, texture differences, and global property differences. This model performed remarkably well in a variety of situations and yielded novel insights into the nature of the computations that underlie perceived object dissimilarity. Below we review our key findings in the context of the existing literature.

Finding 1: Local features combine additively

Our main set of findings (Experiments 1–10) show that the net dissimilarity between objects can be understood as a linear sum of part dissimilarities. The model produced extremely striking fits and in most cases approached the consistency of the data itself, suggesting that it explained nearly all the explainable variance in the data. To our knowledge, this is by far the most successful quantitative model of perceived dissimilarity.

The general principle behind our model is consistent with the theoretical framework proposed by Tversky (1977) for conceptual similarity in general. In Tversky's framework, the dissimilarity between two objects is increased by feature differences between the two objects and decreased by features shared by the two objects. Our findings are analogous: Part differences in general sum linearly, and objects with shared parts will be less dissimilar because the corresponding part terms are zero.

Our finding that local part and texture differences sum linearly is consistent with recent studies in monkey inferotemporal neurons showing that color and shape signals (McMahon & Olson, 2009) and part responses within whole objects sum linearly (Sripati & Olson, 2010). It is also consistent with our earlier finding that simple features combine additively in visual search (Pramod & Arun, 2014). Our model provides several additional insights into how part relations combine in object perception. First, we have shown that part relations decay with distance, implying a spatially tuned matching process between parts. Second, we have shown that objects with similar parts tend to become more distinctive. This effect is analogous to the finding that search becomes easy when distracters become homogenous (Duncan & Humphreys, 1989; Vighneshvel & Arun, 2013). Third, we have shown that distances between objects can be explained better using their natural parts compared to their unnatural parts. It is

consistent with the finding that search for a broken object among unbroken objects is harder when the cut occurs at a local concavity minimum than elsewhere (Xu & Singh, 2002). Finally, we have shown that even for objects that appear holistic, their dissimilarities can be understood as a linear sum of their parts. We conclude that, at least in visual search, dissimilarity is driven by a contour-matching process that is spatially tuned.

Finding 2: Global features add to local features

Our second set of findings (Experiments 11 and 12) concern how global properties combine with local properties. By a global property, we mean one that modifies all features of the object rather than making a local change in features. We tested two such properties: global rotation (which modifies the orientation of all features) and global configuration of parts (which modifies the relative positions of all features). A change in global orientation, stem length, or position of parts of one object in a pair led to a fixed increase in search distance regardless of the object pair. This fixed offset increased systematically with orientation and length. Taken together, our findings show that several global attributes combine additively with existing local part differences. This remarkably simple result suggests a unified framework for understanding global attributes in object vision.

Our approach in general can be used as a quantitative framework to study many classic phenomena in object vision. In the classic finding about global precedence, subjects are faster to detect a global shape than the local shape of a hierarchical stimulus (Kimchi, 1992; Navon, 1977; Sripati & Olson, 2009). This effect could be simply due to stimulus size or due to hierarchical level (Kimchi, 1992; Sripati & Olson, 2009). These findings only indicate the relative strength of global and local features but do not explain how they might combine. This issue can be addressed by measuring dissimilarities between stimuli differing in global and local shape and asking whether the part summation model can explain these data.

In the classic example of configural superiority effects, search for a "(" among ")" is hard, but adding the fixed context ")" makes the resulting search for "()" among "))" easy (Kimchi & Bloch, 1998; Pomerantz & Portillo, 2011; Pomerantz & Pristach, 1989; Pomerantz et al., 1977). Our results suggest a systematic method to study emergent features. Consider for instance two objects AB and AC that share a common part A. According to the part summation model, their dissimilarity is given by $d(\mathrm{AB,AC}) = d_{\mathrm{AA}} + d_{\mathrm{BC}} + x_{\mathrm{AC}} + x_{\mathrm{AB}} - w_{\mathrm{AB}} + w_{\mathrm{AC}}$. Any emergent feature present in the objects AB and AC would result in a failure of the model to predict the observed dissimilarity because the model only considers each part in isolation. This approach offers a quantitative framework to identify emergent features and study how they combine with other features.

Finding 3: Repeated parts and symmetry add to other attributes

Our finding that symmetry, a global property, adds a fixed extra distinctiveness to objects was only possible because the part summation model provided an appropriate baseline to factor out the influence of local features. Our finding indicates that symmetry is an independent property that is computed in visual cortex that combines additively with other features. It explains several observations about symmetric objects: (a) symmetric objects are easier to find in repetition detection (McMahon & Olson, 2007) and (b) a deviation from symmetry is easy to detect (Wagemans, 1997; Wagemans et al., 2012a) because loss of symmetry reduces distinctiveness.

We have further found that the fixed increase in dissimilarity for symmetric objects was present not only for bilaterally symmetric objects but also for objects with repeated parts (i.e., a kind of translational symmetry). This is consistent with the fact that symmetry and repetition can be detected equally fast in displays that are spatially separated (Corballis & Roldan, 1974). Humans are extremely good at detecting a variety of symmetries. High-level visual areas but not early visual areas show differences in activation between symmetric and asymmetric displays (Bertamini & Makin, 2014). A recent transcranial magnetic stimulation study has causally implicated the lateral occipital complex in symmetry detection (Bona, Herbert, Toneatto, Silvanto, & Cattaneo, 2014). Thus, the likely locus for the representation of symmetry is high-level visual cortex. Our results add to the existing literature on symmetry by showing that symmetry causes objects to be more distinctive and also combines additively with local part differences.

Finding 4: Perceptual space can be understood without knowing the features

A central goal of object perception is to understand the features and principles governing perceptual space. Our results demonstrate that certain principles governing perceptual space can be inferred without knowing the underlying features. In Experiments 1–10, we have shown that the distance between two objects

AB and CD can be expressed as a linear sum of the constituent part–part dissimilarities d_{AC}, d_{BD}, d_{AD}, d_{BC}, d_{AB}, and d_{CD}. In Experiments 11 and 12, we have shown that if object AB is rotated or its part configuration is altered, the distance between AB and CD increases by a fixed amount. These findings place constraints on how objects may be represented in the brain. For instance, if objects AB and CD are represented as vectors in an underlying multidimensional space, then the distance between these two vectors, according to our findings, is not a simple vector distance. Rather, it is a complex function that is influenced systematically by (a) spatially tuned part-matching processes and (b) changes in global attributes. Thus, our results challenge the commonly held view that objects can be thought of as vectors in some multidimensional space.

Linearity in perceptual space

The central claim of our study—that object attributes combine additively in visual search—raises the interesting possibility that distances in perceptual space are linear. For a one-dimensional function, linearity implies (a) additivity, i.e., $f(x + y) = f(x) + f(y)$, and (b) scaling, i.e., $f(ax) = af(x)$. In our context, if object AB was represented by vectors **a** and **b**, and CD was represented using vectors **c** and **d**, then we have shown that $d(\{\mathbf{a},\mathbf{b}\},\{\mathbf{c},\mathbf{d}\}) = w_1 d(\mathbf{a},\mathbf{c}) + w_1 d(\mathbf{b},\mathbf{d}) + w_2 d(\mathbf{a},\mathbf{d}) + w_2 d(\mathbf{b},\mathbf{c}) - w_3 d(\mathbf{a},\mathbf{b}) - w_3 d(\mathbf{c},\mathbf{d})$, where d is a distance metric on features, and w_1, w_2, w_3 represent the relative weights associated with each type of comparison. Thus, our results show that the perceived distance between two collections of features is a linear sum of pair-wise distances between all features. Thus, we have demonstrated that perceived distance has an additivity property, but we have not demonstrated scaling. This is impossible in our study because we do not assume any explicit parts-based representation, rendering meaningless any notion of scaling.

However, the scaling property can be demonstrated when the underlying features are known. For instance, if orientation is taken as a feature, then we have previously shown that, short of an intercept term, the perceptual distance between two lines differing by an angle of $\Delta\theta$ is given by $d(\Delta\theta) = k\Delta\theta$, where k is a constant (Arun, 2012). This, in turn, implies that $d(a\Delta\theta) = ka\Delta\theta = \mathrm{ad}(\Delta\theta)$, which confirms scaling. More recently, we have shown this to be true for several other features such as length, intensity, and aspect ratio (Pramod & Arun, 2014). Thus, these results together with our present findings confirm both additivity and scaling of distances in perceptual space, indicative of full linearity.

Conclusions

Taken together, our results indicate a remarkably simple additive rule by which a variety of object attributes combine in object vision. This rule can potentially enable the discovery of novel features by requiring them to combine additively with existing ones. It also places powerful constraints on computational models of object vision by requiring emergent features to sum linearly.

Keywords: object recognition, visual search, perceptual space

Acknowledgments

This research was funded by an Intermediate Fellowship from the Wellcome Trust - DBT India Alliance (to SPA).

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

References

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

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology, 63*(4), 516–556.

Bertamini, M., & Makin, A. D. (2014). Brain activity in response to visual symmetry. *Symmetry*, *6*(4), 975–996.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147.

Bona, S., Herbert, A., Toneatto, C., Silvanto, J., & Cattaneo, Z. (2014). The causal role of the lateral occipital complex in visual mirror symmetry detection and grouping: An fMRI-guided TMS study. *Cortex*, *51*, 46–55.

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

Corballis, M. C., & Roldan, C. E. (1974). On the perception of symmetrical and repeated patterns. *Perception & Psychophysics, 16*(1), 136–142.

Duncan, J., & Humphreys, G. W. (1989). Visual search

and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Dunn, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology. Human Perception and Performance*, *9*(2), 242–257.

Garner, W., & Felfody, G. (1970). Integrality of stimulus and dimensions in various and types of information processing. *Cognitive Psychology, 1*, 225–241.

Gottwald, R., & Garner, W. (1972). Effects of focusing strategy on speeded classification with grouping, filtering, and condensation tasks. *Perception and Psychophysics, 11*(2), 179–182.

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *63*(1), 29–78.

Hyman, R., & Well, A. (1967). Judgments of similarity and spatial models. *Perception and Psychophysics, 2*(6), 233–248.

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

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

Krantz, D., & Tversky, A. (1975). Similarity of rectangles - An analysis of subjective dimensions. *Journal of Mathematical Psychology, 12*, 4–34.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences, 200*(1140), 269–294.

McMahon, D. B. T., & Olson, C. R. (2007). Repetition suppression in monkey inferotemporal cortex: Relation to behavioral priming. *Journal of Neurophysiology, 97*(5), 3532–3543.

McMahon, D. B. T., & Olson, C. R. (2009). Linearly additive shape and color signals in monkey inferotemporal cortex. *Journal of Neurophysiology, 101*(4), 1867–1875.

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

Olivers, C. N., & van der Helm, P. A. (1998). Symmetry and selective attention: A dissociation between effortless perception and serial search. *Perception and Psychophysics, 60*(7), 1101–1116.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic gestalts. *Journal of Experimental Psychology: Human Perception and Performance, 37*(5), 1331–1349.

Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 635–649.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3*(3), 422–435.

Pramod, R. T., & Arun, S. P. (2014). Features in visual search combine linearly. *Journal of Vision*, *14*(4):6, 1–20, doi:10.1167/14.4.6. [PubMed] [Article]

Rollenhagen, J. E., & Olson, C. R. (2000, Feb 25). Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science*, *287*(5457), 1506–1508.

Shepard, R. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1*, 54–87.

Shepard, R. N., & Metzler, J. (1971, Feb 19). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

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*(24), 7788–7796.

Sripati, A. P., & Olson, C. R. (2010). Responses to compound objects in monkey inferotemporal cortex: The whole is equal to the sum of the discrete parts. *Journal of Neuroscience, 30*, 7948–7960.

Torgerson, W. (1965). Multidimensional scaling of similarity. *Psychometrika*, *30*(4), 379–393.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Vighneshvel, T., & Arun, S. P. (2013). Does linear separability really matter? Complex visual search is explained by simple search. *Journal of Vision*, *13*(11):10, 1–24, doi:10.1167/13.11.10. [PubMed] [Article]

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences, 1*(9), 346–352.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012a). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138*(6), 1172–1217.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R.,

Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012b). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218–1252.

Wender, K. (1971). A test of independence of dimensions in multidimensional scaling. *Perception & Psychophysics, 10*(1), 30–32.

Wiener-Ehrlich, W. (1978). Dimensional and metric structures in and multidimensional stimuli. *Perception and Psychophysics, 24*(5), 399–414.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception and Psychophysics, 63*(3), 381–389.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495–501.

Xu, Y., & Singh, M. (2002). Early computation of part structure: Evidence from visual search. *Perception and Psychophysics, 64*(7), 1039–1054.