Record: 1

Article ID: xlm20129

Title: Time Course of Comparison

Author(s): Robert L. Goldstone *Psychology Department, Indiana University*
Douglas L. Medin *Psychology Department, Northwestern University*

Source: Journal of Experimental Psychology: Learning, Memory, And Cognition. Vol. 20 (1) January 1994, pp. 29-50. American Psychological Association

ISSN: 02787393

Digital Object ID: 10.1037//0278-7393.20.1.29

Article Type: Journal Article

Abstract: Similarity as interactive activation and mapping (SIAM), a model of the dynamic course of similarity comparisons, is presented. According to SIAM, when structured scenes are compared, the parts of one scene must be aligned, or placed in correspondence, with the parts from the other scene. Emerging correspondences influence each other in a manner such that, with sufficient time, the strongest correspondences are those that are globally consistent with other correspondences. Relative to globally inconsistent feature matches, globally consistent feature matches influence similarity more when greater amounts of time are given for a comparison. A common underlying process model of scene alignment accounts for commonalities between different task conditions. Differences between task conditions are accounted for by principled parametric variation within the model.

PDF Image: PDF Full Text

Database: PsycARTICLES

Time Course of Comparison

By: Robert L. Goldstone

Psychology Department, Indiana University

Douglas L. Medin

Psychology Department, Northwestern University

Experiment 2 was presented at the 14th annual conference of the Cognitive Science Society, Bloomington, Indiana, August 1992. Experiments 1 and 2 were reported in Robert L. Goldstone's doctoral thesis submitted to the Department of Psychology, University of Michigan, 1991. We wish to express special thanks to Dedre Gentner for innumerable suggestions and advice throughout the research project. This research was funded by National Science Foundation Grant BNS 87-20301 awarded to Dedre Gentner and Douglas L. Medin. Many useful comments on earlier incarnations of the research were provided by Evan Heit, Keith Holyoak, Roger Ratcliff, Steven Sloman, Ed Smith, and Keith Smith.

Correspondence may be addressed to: Robert L. Goldstone, Psychology Department, Indiana University, Bloomington, Indiana 47405. Electronic mail may be sent to rgoldsto@ucs.indiana.edu.

The similarity of some situations is immediately apparent. The basis for the similarity of the *The Odyssey* and *The Iliad* is clear at once (both are Greek epic poems). Recognizing the commonalities between *The Odyssey* and *The Wizard of Oz,* however, is a relatively slow process. The time course of similarity assessments provides a useful tool for investigating the process by which entities are compared. These time course data inform questions concerning how entities are mentally represented and have important implications for theories of similarity. In this article, we present a model of similarity comparison that makes specific time course predictions, which were tested in three experiments. Before turning to that model, we first outline the need for a consideration of similarity processes.

Process Models of Similarity

Most traditional models of similarity have not addressed the temporal aspects of processing. Instead, they have modeled the similarity of two things by algebraic formulas. In the basic description of both featural (Tversky, 1977) and dimensional (Carroll & Wish, 1974; Torgerson, 1965) models, very little attention is paid to the actual mechanisms by which similarity is computed (although Tversky did discuss processing principles that fix some terms). In both of these approaches, there are no obvious grounds for predicting qualitative shifts in similarity as processing time increases. In the model we present in this article, qualitative shifts are expected: Early in processing, local matches strongly influence similarity; with time, global consistency becomes more important. We argue that the process by which correspondences between two entities are established is nontrival and importantly constrains models of similarity.

It is not surprising that featural and multidimensional scaling models have paid little attention to the comparison process. After all, measurements of similarity range from direct ratings or judgments to perceptual and memorial confusions (the more similar two things are, the more confusable they should be). Instead of expecting some generic comparison process, one might anticipate task-specific processing mechanisms. Consider, for example, same-different reaction times as an index of similarity. The assumption is that the more similar two things are, the longer it will take to say that the things are different. Several process models of these tasks have been proposed (for reviews, see Farell, 1985; Nickerson, 1972). One candidate model (Egeth, 1966) of same-different judgments posits that stimuli are compared attribute by attribute in a serial fashion and that the process is terminated by the detection of a difference with respect to any relevant attribute. This model is sometimes augmented by a fast "identity reporter" that emits a response if the stimuli are the same, based on wholistic cues (Bamber, 1969).

Other process models are relevant to direct similarity ratings. According to the "anchoring and adjustment" process (Lopes & Johnson, 1982; Tversky & Kahneman, 1974), subjects anchor their judgment at the similarity value associated with some salient dimension. The similarity rating is then adjusted (often insufficiently) by the other dimensions such that after each adjustment the new similarity lies between the old similarity and the similarity value of the dimension most recently considered. Parducci's (1965; also see Krumhansl, 1978) range-frequency theory incorporates two processing principles: (a) that people who are asked to assign a numeric value as a rating tend to divide the range into a fixed number of intervals of equal breadth and (b) that people tend to establish intervals that are used with equal frequency. If many events fall within a particular interval, finer discriminations are made and the interval is subdivided.

It should be clear that some of the processes hypothesized for similarity ratings are of limited relevance to a same-different task, and vice versa. Given that different processes are involved in tasks that all purportedly measure similarity, one may suggest that the most powerful and useful generalizations about similarity should be stated at an abstract, relatively process-independent level. Many researchers do find a high degree of correlation between similarity as measured by same-different judgments and similarity as measured by ratings (Corter, 1987; Getty, Swets, Swets, & Green, 1979; Podgorny & Garner, 1979; Sergent & Takane, 1987); however, these same researchers find that the two tasks are not completely sensitive to the same information (see also Beck, 1966). Few explanations have been put forth to explain the dissociations between the measures.

An alternative strategy to ignoring task differences (or treating different tasks as unrelated) is to develop a single unified model of similarity that accounts for both similarities and differences among tasks. Although the argument against a single, generic comparison process is fairly compelling, developing such a model may play an extremely useful function in linking these various measures of similarity. Indeed, we have been surprised at just how successful a model using a single mechanism to predict both similarity judgments and same-different reaction times can be. The key idea is that comparison processes unfold over time and same-different tasks typically involve shorter time periods than do similarity ratings.

The model we present in the next section provides an account for the strong correlation between similarity measures and also their systematic dissociation. According to this model, different measures of similarity are correlated to the extent that the tasks all require a process that places the parts of the compared entities into alignment. Dissociations between similarity measures are accounted for by variations in task-relevant parameters. In specific, different tasks can be modeled by performing different amounts of processing. Tasks that are performed quickly by subjects are modeled by allowing fewer processing cycles in a computational model.

Similarity as Interactive Activation and Mapping

Brief Description of the Model

We recently developed the similarity as interactive activation and mapping (SIAM) model of the dynamic course of similarity comparsions as a model of how people judge the similarity of structured scenes (Goldstone, 1994; Goldstone & Medin, in press). SIAM shares architectural commonalities with McClelland and Rumelhart's (1981) interactive activation model of word perception and is highly related to the structure mapping engine (SME) (Falkenhainer, Forbus, & Gentner, 1989) and the analogical constraint mapping engine (ACME) (Holyoak & Thagard, 1989) models of analogical reasoning. The primary assumption of SIAM is that determining the similarity of two scenes involves placing the scenes' components into alignment or correspondence (e. g. , see Gentner 1983, 1989; Markman & Gentner, 1993, Medin, Goldstone, & Gentner, 1993). Similarity is determined by a process of interactive activation among feature, object, and role correspondences. The degree to which features from two scenes are placed in correspondence depends on how strongly their objects are placed in correspondence. Reciprocally, how strongly two objects are placed in correspondence depends on the correspondence strength of their features. A similar pattern of simultaneous mutual influence occurs between objects and their roles.

SIAM's network architecture is composed of nodes that excite and inhibit each other. A node represents the hypothesis that two entities correspond to one another in two scenes. A feature-to-feature node represents the hypothesis that two features correspond to each other. There is one node for every pair of features that belong to the same dimension; if each scene has O objects with F features each, there are O^2F feature-to-feature nodes. SIAM presupposes both that the object membership of features is known and that corss-dimensional featural alignments are not considered. These two assumptions are convenient simplifications. For example, it may be that the object membership of features is driven by bindings to spatial locations (e. g. , Nissen, 1985). Our idea is that such bindings take place on a time scale much shorter than the situations we model, although SIAM could readily be extended in this direction. There is other work (e. g. , Melara & Marks, 1990) that demonstrates that cross-dimensional feature correspondences are considered and do influence judgments. For example, a loud tone and a bright light might be aligned. So far, we have not attempted to extend SIAM to cross-dimensional matches, in part because we wish to avoid having O^2F^2 feature-to-feature nodes.

Even the requirement of O^2F feature-to-feature nodes may seem overly profligate. A version of SIAM has been constructed that creates feature-to-feature nodes only when a featural correspondence has been noticed by the system. Although this version typically creates a number of feature-to-feature nodes that increases linearly with *OF,* it is not considered here because its processing mechanism is far more complicated.

Network activity starts by features being placed in correspondence according to their similarity. As the activation of a feature-to-feature node increases, the two features referenced by the node are placed in stronger correspondence. After this occurs, SIAM begins to place objects into correspondence that are consistent with the feature correspondences. Object-to-object nodes represent the hypothesis that two objects correspond. As objects begin to be placed in correspondence, activation is fed back down to the feature mismatches that are consistent with the object alignments. In this way, object matches influence activation of feature matches and feature matches influence the activation of object matches concurrently. Role-to-role nodes operate to place scene parts in correspondence that play the same role within a scene. Role correspondences influence object correspondences, and vice versa. (For empirical justifications of the inclusion of roles in the presentation, see Gentner, 1983, 1989; Goldstone, Gentner, & Medin, 1989; Goldstone, Medin, & Gentner, 1991.)

Activation spreads in SIAM by two principles: (a) Nodes that are consistent with one another send excitatory activation to each other, and (b) nodes that are inconsistent inhibit one another. Nodes are inconsistent if they yield many-to-one mappings, and are consistent otherwise. For example, there is a many-to-one mapping between AB and AA; the first element of AB maps onto both elements of AA. As such, the object-to-object node that represents the hypothesis "First element of AB corresponds to first element of AA" inhibits the "First element of AB corresponds to second element of AA" object-to-object node and vice versa. An object-to-object node that places A and B into correspondence excites all feature-to-feature nodes that put A's features into correspondence with B's features, and vice versa.

Scenes are described to SIAM in terms of relations that take objects as arguments, and objects that contain feature slots that are filled with particular feature values. On each time cycle, activation spreads between nodes. The network's pattern of activation determines both the similarity of the scenes and the alignment of the scenes' features, objects, and relation arguments. With time, nodes that have high activity will be weighted highly in the similarity assessment and their elements will tend to be placed in alignment.

Behavioral Characteristics of the Model

The details of SIAM's implementation are presented elsewhere (Goldstone, 1994). For the present purposes, we focus on several important features of SIAM's operation. First, SIAM predicts that feature matches increase similarity more if they belong to objects that correspond to each other. Two scenes may have a feature match between noncorresponding objects. For example, in Scene A, a boy with a white shirt throws a football; in Scene B, a boy with a brown shirt throws a baseball. The matching "brown" feature between the football in Scene A and the boy's shirt in Scene B is a match between noncorresponding objects. The objects do not correspond because they are featurally dissimilar and because they play different roles in the scenes. Such a feature match is called a match out of place (MOP). Conversely, a feature match between corresponding objects is called a match in place (MIP). If both of the boys have green pants, this shared color feature is a MIP. SIAM predicts that MIPs increase scene similarity more than MOPs do because they receive excitation from strong object-to-object correspondences. Because of the boys' featural and relational similarity, the node that places the two boys in correspondence is highly activated. This object-to-object node then sends activation back down to the nodes that place the boys' features into correspondence. The influence of a feature match (or mismatch) in a similarity assessment is directly related to its activation level.

In previous applications, SIAM has successfully described a number of detailed phenomena concerning the influences of MIPs and MOPs and their distribution on similarity judgments (Goldstone, 1994). SIAM has thus far been used to model similarity ratings, category inferences, mapping judgments, feature detection, and ease-of-comparison judgments. In this article, we describe the application of SIAM to indirect measures of similarity; this application required a consideration of the time course of alignment.

Experiment 1

SIAM is unique among models of similarity in being dynamic. As SIAM executes more cycles of activation adjustment, feature-to-feature nodes become increasingly influenced by object correspondences. At first, how strongly two features correspond to each other depends mostly on the features' similarity; two features tend to be put into correspondence if they are identical or highly similar. With more time, feature correspondences depend increasingly on object correspondences. Specifically, features tend to be placed in strong correspondence if they belong to objects that are placed in correspondence. In turn, objects are placed in correspondence if they are featurally similar to each other. After several cycles have passed, object correspondences reflect objects' featural similarity and feed activation back down to the feature correspondences with which they are consistent.

One prediction of this temporal process is that the relative importance of MIPs compared with MOPs increases with processing time. MIPs and MOPs are originally equally salient, but object correspondences send more activation to MIPs than to MOPs. That is, feature matches between corresponding objects show more of a benefit from object correspondences than do feature matches between noncorresponding objects.

In Experiment 1, we tested the influence of time on feature match salience by using a same-different judgment task with a deadline. Subjects had to decide whether two scenes contained the same butterflies within a specified time limit. The rate of incorrect responses on "different" trials was assumed to be directly related to the scenes' similarity. Each scene was composed of two butterflies, and each butterfly had four critical features. Sample displays are shown in

Two sample displays from Experiment 1. In Panel A, two butterflies are identical (four matches in place [MIPs]), and two butterflies agree on three of four features (three MIPs). In Panel B, corresponding butterflies share six feature (six MIPs), and noncorresponding butterflies share one feature (one match out of place [MOP]—the spotted body pattern).

Figure 1. A symbolic representation is shown below each of the butterflies. For example, the display in Figure 1A contains the butterflies AAAA and BBBB, with the letters referring to different values along the four dimensions (body shading, head type, tail type, and wing shading). The butterflies on the right in Figure 1A are identical to those on the left, with the exception of the body shading of one butterfly. Because the butterfly BBBB is not identical to BBBD, a subject should respond that the scenes have different butterflies.

Both Figure 1A and 1B show displays that have seven feature matches. In Figure 1A, all seven feature matches appear on butterflies that are in true correspondence; consequently, the display has seven MIPs. Butterflies are in true correspondence if their correspondence is part of the consistent set of butterfly correspondences that maximizes the number of MIPs. Butterfly correspondences are consistent if they do not create a many-to-one mapping between butterflies. In Figure 1B, six feature matches are MIPs, and one feature match is a MOP. Butterfly AAAB corresponds to AAAA, but AAAB also has a feature match in common with BBBB (they both have spotted bodies).

The following empirical questions were addressed in Experiment 1: Are the scenes in Figure 1B as similar to each other as the scenes in Figure 1A are? How do the displays' relative similarity vary with processing time? SIAM predicted that the similarity of the scenes in Figure 1A would increase relative to the similarity of Figure 1B's scenes with time. MIPS, relative to MOPs, should increase similarity more as object correspondences begin to influence feature correspondences. Experiment 1 compared displays that varied systematically in their number of MIPs and MOPs. Each of the displays was presented under three different deadline conditions. The similarity of each display, measured by the percentage of false "same" judgments, was predicted to be an interactive function of deadline and distribution of MIPs and MOPs.

Method

Subjects.

Thirty-three undergraduate students from the University of Michigan served as subjects in order to fulfill a course requirement.

Materials.

Each display contained two scenes divided by a 1-cm black line. Scenes were composed of two butterflies, which in turn were composed of four features: wing shading (22 different values, including striped, spotted, checkerboard, black, brick, etc.), head style (triangle, square, circle, or M-shaped); tail style (radiating lines, zigzag, cross lines, or line with ball); and body shading (the same range of values as wing shading). None of the wing shadings on any of the four butterflies had the same value as the body shading of any butterfly. The display area was 17 cm high × 21 cm across. Each individual butterfly was approximately 6 cm × 4 cm. Viewing distance was not controlled but was approximately 60 cm.

Three different spatial layouts were used. In the same-positions layout, butterflies that corresponded to each other (according to their feature overlap) were placed in the same relative locations in their respective scenes (abstractly, AB ? AB). In the opposite-positions layout butterflies that did not correspond to each other were placed in the same relative locations (abstractly, AB ? BA). Figure 1 shows butterflies in an opposite-position layout. In Figure 1A, butterfly BBBB corresponds to butterfly BBBD, but BBBB is in the same relative spatial location as butterfly AAAA. In the unrelated-positions layout, neither of the butterflies of one scene had the same relative location as either of the butterflies of the other scene. Within each scene, the two butterflies were always placed diagonal to each other, each butterfly occupying one of four locations (upper left corner, upper right corner, lower left corner, and lower right corner).

Design.

Table 1 lists the 13 different displays presented to subjects. Each display differed in terms of its number of MIPs and MOPs. The butterflies in one scene (the initial scene) can be represented by AAAA and BBBB, with each letter representing a particular value on one of the four butterfly dimensions. The 13 displays changed features from the initial scene to create the other scene (the changed scene). For example, the changed scene's butterflies for Method 13 are AAAA and BBBD, signifying that the initial and changed scenes differed only on a single dimension value. Figure 1A is an instantiation of Method 13; the difference between the scenes was that the spotted body of one butterfly became scaly in the other scene. Figure 1B depicts Method 11; BBBD differed from BBBB by one value (spotted body became scaly), AAAB differed from AAAA by one value (striped body became spotted), and there was a single MOP between BBBB and AAAB (they both had spotted bodies).

Procedure.

The subjects' task was to press a key with one hand if the butterflies in one scene were the same as the butterflies in the other scene and to press a key with the other hand if the two scenes' butterflies were different. They were instructed to ignore butterfly positions and base their decision on the features of the butterflies. If the subject did not respond S or D within the time indicated by the deadline, "OVERTIME" appeared on the screen. Otherwise, either a check or an X appeared on the screen, depending on whether the subject's response was correct or incorrect. After 2 s, the screen was erased. After another second, the next trial commenced.

Each subject was given a total of 494 trials. On half of these trials, each of the butterflies in one scene matched a featurally identical butterfly in the other scene. These comprised the "same" trials. The remaining trials consisted of 19 repetitions of each of the 13 displays. One of these 19 repetitions was a practice block set at the slowest deadline. The other 18 repetitions consisted of six blocks at each of the three deadlines, randomly ordered. Short, medium, and long deadlines corresponded to 1 s, 1. 84 s, and 2. 68 s. For these respective deadlines, subjects were instructed that they would be required to make "VERY fast responses," "moderately fast responses," and "fairly slow responses. "

Results

Both the main effect of match type (MIP vs. MOP) and the interaction between match type and deadline were of interest. Because of the deadline procedure, error responses rather than response times were the most sensitive measure of similarity. The analyses focused on incorrect responses on "different" trials as a measure of similarity. Error rate and overtime responses were combined to form a measure of the total incorrect response rate. The analyses were not significantly changed when error trials and overtime trials were analyzed separately.

There was a clear influence of both MIPs and MOPs on similarity as measured by incorrect responses on "different" trials. One simple method of analyzing the data was to extract the 3 × 3 factorial design embedded within the experiment's design. As such, Displays 3-5 and 7-12 were analyzed because they constituted every level of MIPs between 4 and 6, combined with every level of MOPs between zero and two. Averaging over number of MOPs and deadline, 4 MIPs = 24% incorrect responses, 5 MIPs = 33%, and 6 MIPs = 45%, $F(2,64) = 7. 1$, $MS_e = 4. 5$, $p < . 05$. For MOPs, 0 MOPs = 31%, 1 MOP = 35%, and 2 MOPs = 36%, $F(2,64) = 3. 6$, $MS_e = 6. 1$, $p < . 05$. Subjects were more likely to call two scenes the same incorrectly when the scenes contained many MIPs and MOPs. MIPs were much more influential than MOPs in determing error rates. Increasing the number of MIPs by two increased the rate of incorrect answers by 21 percentage points, compared with a 5 percentage point increase obtained by increasing the number of MOPs by two.

There was a reliable interaction between both MIPs and MOPs and deadline. As subjects were given more time to respond, the influence of MIPs increased and the influence of MOPs decreased. The interaction between number of MOPs and deadline is shown in

Results of Experiment 1. The number of matches out of place had a greater influence on the rate of incorrect "different" responses for short deadlines than for medium or long deadlines.

Figure 2, $F(4,128) = 2. 9$, $MS_e = 8. 9$, $p < . 05$. The shortest deadline showed the greatest sensitivity to differences in the number of MOPs in a display. Two additional MOPs increased incorrect responses by 10%, as compared with 4% for the longest deadline, $F(1,32) = 4. 4$, $MS_e = 6. 2$, $p < . 05$. The interaction between number of MIPs and deadline is shown in

Results of Experiment 1. The number of matches in place had a greater influence on the rate of incorrect "different" responses for long and medium deadlines than for short deadlines.

Figure 3, $F(4,128) = 5. 7$, $MS_e = 6. 3$, $p < . 05$. The long and medium deadlines showed the greatest sensitivity to MIPs, where sensitivity was measured by the difference between displays that had 4 and 6 MIPs. At the shortest deadline, the influence of two additional MIPs was not significantly different from the influence of two additional MOPs.

Reaction times were significantly greater for "same" responses (1. 25 s) than for "different" responses (1. 17 s), paired $t(32) = 2. 9$, $p < . 05$. Correct "same" responses were fastest when corresponding objects were placed in the same locations (same location = 1. 21 s, unrelated location = 1. 25 s, and opposite location = 1. 28 s). "Same" error rate data reflect the same trend (same location = 83% correct "same" responses, unrelated location = 79%, and opposite location = 76%). "Different" response speed and error rate were not significantly affected by object location.

Discussion

Both MIPs and MOPs increased similarity, but MIPs increased similarity more. Longer deadlines decreased the influence of MOPs and increased the influence of MIPs on the error rate measure of similarity. At the shortest deadline, MIPs had only slightly more influence on similarity than MOPs. At the longer deadlines, MOPs became much less influential, and MIPs became more influential. A single MIP was always more influential than 2 MOPs at the longer deadlines, but often 2 MOPs were more influential than 1 MIP at the shortest deadline.

SIAM's account of these results is that the influence of object-to-object correspondences on feature-to-feature correspondences increases with processing time. MIPs become more important with time because they are matches that belong to aligned objects. As subjects begin to place two objects into correspondence, feature matches between the objects receive more attention. At the same, feature matches between unaligned objects receive less attention. The more attention a particular feature match receives, the more it influences similarity, when similarity is measured by the percentage of trials in which different scenes are judged the same.

If a model of similarity is to provide an account of the same-different judgment task, it has to consider temporal aspects of the task, and not simply the scenes' featural descriptions. Process models such as SIAM have an advantage over other models of similarity in this regard. SIAM is a dynamic model; scene correspondences and similarity assessments develop over time. The experimental manipulation of deadlines has a natural analogue in SIAM—the number of cycles of activation adjustment that SIAM completes. In the Theoretical Analysis section of this article, we show that SIAM gives a quantitative as well as a qualitative account of the results of Experiment 1.

Experiment 2

Experiment 1 provided evidence that the influence of object correspondences on feature correspondences increases with time. Feature matches that are compatible with the optimal object correspondences count more than feature matches that are not, and this difference grows with processing time. In Experiment 2, we explored whether there is an increased influence of object correspondences on each other with time. SIAM initially begins to place objects in correspondence on the basis of their featural overlap; the more featural commonalities two objects have, the more strongly they will be placed in correspondence. However, the strength of an object correspondence is also influenced by its consistency with other object correspondences. If two objects from one scene correspond to a single object in another scene, the two correspondences are inconsistent and decrease each other's strength. SIAM predicts that object correspondences become increasingly influenced by other object correspondences with time, as activation between nodes spreads.

One prediction of this temporal processing is that feature matches that are inconsistent with the set of globally consistent correspondences tend to influence similarity less with time. Globally consistent feature matches should become more influential with time. Global consistency is determined by the entire pattern of correspondences. A set of mappings is globally consistent if it (a) yields only one-to-one mappings and (b) maximizes the number of MIPs-matching features that belong to corresponding entities. [1]

To test the influence of processing time on globally consistent and inconsistent feature matches, we used Experiment 1's task and materials in Experiment 2. Subjects decided whether two scenes contained the same butterflies within a specified amount of time. Sample scenes and their symbolic representations are shown in

Abstract design of stimuli used in Experiment 2. Each scene was composed of two objects with four features (represented by letters). The target scene was compared with the base scene and with two transformations of the base scene. If the mutual consistency of butterfly alignments was not taken into account, the left butterfly of the base scene (BABA) corresponded to both butterflies from the target scene. If the global consistency of object alignments was considered, the right butterfly of the target scene was placed in optimal alignment with the right butterfly of the base scene.

Figure 4. The target scene was composed of butterflies AAAA and BBBB, with the letters referring to different values along the four dimensions.

The target scene was compared with a base scene and two derivatives of the base scene. Both of the butterflies in the target scene had more matching features in common with the left butterfly than the right butterfly in the base scene. Thus, if we were only constrained by the locally preferred mappings, we would map both target butterflies onto the left butterfly in the base scene. However, if the global consistency of object mappings is maintained, then this many-to-one mapping is not permitted. The best globally consistent mapping is to match the left butterflies to each other, and the right butterflies to each other. This consistent mapping yields three MIPs and two MOPs. The only other consistent mapping (aligning the left butterfly of one scene with the right butterfly of the other scene) yields two MIPs and three MOPs. In short, BBBB (in the target scene) corresponds to BABA (in the base scene) if we consider only local feature matches, but BBBB corresponds to XXXB if we consider the influence that object correspondences have on each other.

As mentioned, the target scene was also compared with two derivatives of the base scene. Each derivative differed from the base scene by only a single feature. In Figure 4A, one of the locally preferred matches is absent, leaving all of the globally consistent matches intact. In Figure 4B, one of the globally consistent matches is absent, preserving all of the locally determined matches. The experimental question is, Which of these two derivatives is more similar to the target? In other words, does it matter more if a local or global match is removed? SIAM predicts that the answer to this question depends on timing. Early on, local matches are more important; with increasing time, global matches should become more important. The target and Figure 4B have four feature matches between butterflies that are aligned on the basis of local object-to-object similarity, but they have only two globally consistent feature matches. The target and Figure 4A have three locally determined matches and also three globally consistent feature matches. If SIAM's prediction is valid, then it should be possible to make either Figures 4A or 4B more similar to the target butterfly by manipulating the amount of time allowed for a response.

Method

Subjects.

Thirty-three undergraduate students from the University of Michigan served as subjects in order to fulfill a course requirement.

Materials.

Displays consisted of the target scene and one of three other scenes. The target scene is described by AAAA/BBBB, signifying a scene with two butterflies, each with four dimensions such that the butterflies did not have any of the same dimension values. One of the three scenes with which the target scene was compared was the base scene, described by BABA/XXXB. Locally preferred and globally consistent object alignments were in conflict when the target and base scenes were compared. If each butterfly's correspondence was determined independently, then the butterfly characterized by BBBB would most strongly correspond to BABA. BBBB had two matching features in common with BABA and only one in common with XXXB. As such, local object correspondences would place both AAAA and BBBB into correspondence with BABA. However, if global consistency was considered and many-to-one butterfly mappings were suppressed, BBBB would most strongly correspond to XXXB. Globally, the mapping AAAA ? BABA, BBBB ? XXXB yields three MIPs, whereas the mapping AAAA ? XXXB, BBBB ? BABA yields only two MIPs. Thus, depending on whether AAAA's mapping influenced BBBB's mapping, BBBB would be placed in correspondence with either BABA (the locally preferred mapping) or XXXB (the globally consistent mapping).

The other two scenes with which the target scene was compared were identical with the base scene except that either one globally consistent or one locally preferred feature match was removed. The abstract description of the scene that kept all of the globally consistent matches, removing one locally preferred match, is XABA/XXXB. Going from the base scene to this scene, one B is changed into an X. This change results in one less matching feature with the target, and the matching feature that is lost belongs to an object that locally but not globally corresponds to BBBB. Conversely, the third comparison scene, BABA/XXXX, preserves all of the locally aligned matches but removes one globally consistent mapping.

Butterfly position, dimension order, dimension values, scene location, and display type were all randomized. Displays were presented on Macintosh SE30 computers.

Procedure.

The procedure was the same as in Experiment 1. The subjects' task was to press a key with one hand if the butterflies of one scene were the same as the butterflies of the other scene and to press a key with the other hand if the two scenes' butterflies were different in any feature.

Subjects were given 608 trials in all. On half these trials, each of the butterflies of one scene matched a featurally identical butterfly in the other scene. These were the "same" trials. The remaining trials consisted of 19 repetitions of each of the three comparisons (target-base, target-local match preserved, and target-global match preserved) and 13 filler trial displays. One of these 19 repetitions was a practice block set at the slowest deadline. The other 18 repetitions consisted of six blocks at each of three deadlines (1 s, 1. 84 s, and 2. 68 s), randomly ordered.

Results

The crossover interaction between deadline and type of display is shown in

Results of Experiment 2. At the slow deadline, the scene with more global matches preserved was more confusable with the target scene. At the fast deadline, the scene with more local matches preserved was more confusable with the target scene.

Figure 5. The 3 × 3 interaction was significant, $F(4,128) = 3.5$, $MS_e = 0.68$, $p < .05$, as was the 3 (deadlines) × 2 (local vs. global matches kept) interaction, $F(2,64) = 4.4$, $MS_e = 0.66$, $p < .05$. When subjects were forced to respond within a short deadline, the display that preserved the locally preferred match was more often incorrectly responded to as "same" than was the display that preserved the globally consistent match. The opposite effect was found when subjects were given longer to respond. The following four mean error rates were of particular interest: slow deadline/global matches kept = 5%, slow deadline/local matches kept = 3%, fast deadline/global matches kept = 18%, and fast deadline/local matches kept = 21%. A planned comparison of these four data showed a significant interaction between deadline and type of scene on error rate, $F(1,128) = 4.16$, $MS_e = 6.3$, $p < .05$. The overall times to correctly respond "different" for the different displays were not significantly different (base = 1. 147 s, global match kept = 1. 137 s, and local match kept = 1. 135 s).

In addition to the interaction between display type and deadline, there was an interaction between butterfly position and display type, using error rate on "different" trials as a dependent measure, $F(4,128) = 3.02$, $MS_e = 0.58$, $p < .05$. For base, global-match-kept, and local-match-kept scenes, respectively, the error rates were 15%, 17%, and 10% in the same-positions displays; 14%, 9%, and 13% in the opposite-positions displays; and 17%, 11%, and 10% in the unrelated-positions displays. Scenes with preserved global matches were more often incorrectly thought to be the same when (globally) corresponding butterflies were in the same relative positions than when they were in unrelated positions. Conversely, more incorrect "same" judgments were made for scenes with preserved local matches when corresponding butterflies were given unrelated, rather than the same, positions.

Discussion

Measuring similarity by the percentage of trials on which scenes with different butterflies were judged to be the same, the obtained results are consistent with SIAM's prediction. More incorrect "same" judgments were found for short deadlines when local, as opposed to global, matches were preserved. More incorrect "same" judgments were found for the longest deadline when global, as opposed to local, matches were preserved. This is consistent with SIAM's dynamic account of similarity. The influence of one object-to-object mapping on another takes time to develop, and until it develops, object-to-object mappings are largely determined by feature-to-feature matches. Locally consistent matches are more important than globally consistent matches for similarity early in processing (fast deadline). Later in processing, globally consistent matches gain in importance relative to locally consistent matches. It takes time to set up the influence that one object-to-object mapping has on another object-to-object mapping, and until this happens, error data show the influence of two-to-one mappings. At first, both butterflies of the target are mapped onto one butterfly of the other scene, but with time the influence of one mapping redirects the other mapping.

SIAM also provides an account of the interaction between butterfly positions and the preservation of local or global matches. When corresponding butterflies are in corresponding locations (same-positions displays), location information (represented by role-to-role nodes) bolsters globally consistent object-to-object correspondences. When corresponding butterflies are placed in opposite locations (opposite-positions displays), location information hinders the development of the globally consistent correspondences. As such, opposite-positions displays are correctly predicted to show more of an influence of locally determined object correspondences than same-positions displays.

Experiment 1 indicated that, with time, objects are placed in correspondence on the basis of their featural similarity and object alignment influences the salience of feature matches. Experiment 2 indicated that featural similarity cannot completely predict object correspondences. Objects tend to be aligned if they share many features; however, object alignment also depends on the similarity of other pairs of objects.

It might be expected that an account of the deadline results could be provided without assuming object alignment, if it was assumed that (a) both simple features and complex conjunctions of features (e. g. , striped body and checkered wings) are represented, and (b) simple features are computed more quickly than complex features. The first assumption has proven useful in models of categorization (Gluck & Bower, 1988; Hayes-Roth & Hayes-Roth, 1977). The second assumption is plausible given work by Treisman (reviewed by Treisman, 1987) that suggests that simple feature identification is fast and parallel, whereas identification of conjunctions of features is slow and serial. For the particular stimuli used in Experiment 2, however, this-account makes two wrong predictions. First, it predicts that scenes that preserve all of the local matches will always be judged as more similar than scenes that preserve the global matches. The target scene and Figure 4A have one complex feature (round head and checkered wings) and four simple features in common. The target scene and Figure 4B scene have two complex features (round head with checkered wings and wavy body with radiating-lines tail) and four simple features in common. The simple and complex features model predicts the display with two complex and four simple features to be the most similar at every deadline. Second, the model predicts that the greater similarity of the display with locally preserved matches will be more pronounced later in processing than early in processing. This prediction derives directly from the assumption that complex features become more important than simple features with increased processing. This second prediction is directly opposed to the empirical results. Instead of increasing its advantage with processing time, the display with the locally preserved matches lost its similarity advantage to the display with globally preserved matches as the deadline increased.

The problem with the simple and conjunctive features account is that it essentially weights a feature match more if it belongs to highly similar objects. Adding a simple feature match between two objects will also add conjunctive feature matches if the two objects have other matching features. However, as shown earlier, object alignment in Experiment 2 depended not only on object similarity, but also on the development of globally consistent object correspondences. When object similarity is dissociated from object alignment, as it was in Experiment 2, then feature match salience depends increasingly on object alignment, and not object similarity, with time.

Experiment 3

Evidence from the first two experiments indicates that object correspondences are determined by feature matches and other object correspondences and that the influence of object correspondences on similarity (as measured by confusion errors) increases with time. In SIAM, object correspondences depend on an additional third information source: relations between objects. Objects that play the same role in two scenes may be placed in correspondence even if the objects are not featurally similar. The influence of roles on object correspondences was not considered in Experiments 1 and 2 because the instructions explicitly discounted role information. Subjects were instructed to respond "same" when the same butterflies were present in two scenes, even if the relative positions of the butterflies were not identical. Thus, subjects were instructed to ignore the spatial relation of one butterfly to the other and base their decision only on the featural properties of the butterflies.

The purpose of Experiment 3 was to manipulate instructionally the relevance of spatial relations to the task. One group of subjects was asked to respond "same" only when the same objects occupied the same roles in their scenes, with "role" defined in terms of corresponding spatial position. The second group responded "same" when the same objects were present in two scenes, regardless of their spatial relations. The former group shall be referred to as the role-relevant group and the latter group as the role-irrelevant group. The design is similar to that of a study by Proctor and Healy (1985), who referred to these groups as order-relevant and order-irrelevant, respectively.

In SIAM, the difference between the role-relevant and role-irrelevant groups can be modeled by varying the influence of role-to-role nodes on object-to-object nodes. Similarity can still be measured by feature (mis)matches weighted by their activation. The instructional manipulation is assumed to affect only whether object correspondences are influenced by role correspondences. As role correspondences increasingly influence object correspondences, objects will be increasingly aligned on the basis of their role in the scene. In turn, object correspondences will influence the salience of feature (mis)matches. In this way, relation information can affect the similarity of two scenes, even if similarity is only a function of feature (mis)matches and their activation.

Modeling the difference between the role-relevant and role-irrelevant groups by varying the influence of role correspondences, SIAM predicts that the different instruction groups would yield different patterns of scene similarity. In Experiments 1 and 2, whether a feature match was a MIP or a MOP depended on the featural similarity and consistent correspondences of objects. In Experiment 3, a MIP was defined as a feature match that occurs between objects that are placed in correspondence according to their roles. Thus, in

Two sample displays from Experiment 3. Each different pattern represents a different color. A match in place (MIP) was defined as a matching color between squares that occupied the same role in their cross. A match out of place (MOP) was defined as a matching color between squares that occupied different roles.

Figure 6A, the two crosses (the center is empty) share 2 MIPs; the shared striped and checkerboard shadings belong to squares that occupy the same relative positions. Each square is considered to be an object. In Figure 6B, the two crosses have 3 MOPs in common; the three shared shadings belong to squares that occupy different positions. SIAM predicts that whether a feature match is a MIP or a MOP should matter more for the role-relevant than for the role-irrelevant group. If object correspondences are not determined at all by role correspondences, there would be no difference between the influence of a MIP and that of a MOP. Conversely, if roles do influence object correspondences, objects in similar roles will be placed in relatively strong correspondence, and their feature matches will be particularly salient.

One possible ramification of SIAM's prediction is that the similarity of two scenes relative to other scenes may not remain fixed under the two instruction conditions. If roles have very little influence on object correspondences, then the scenes in Figure 6B may be more similar to each other than the scenes in Figure 6A are to each other. Such a result would indicate that 2 MOPs increase similarity more than 1 MIP. Conversely, if roles have a substantial influence, the scenes in Figure 6A may be more similar than the scenes in Figure 6B. If the relative similarities of the scenes in Figures 6A and 6B reverse with task instructions, it would not be possible to claim that the two tasks tap into the same fundamental construct of similarity. Such a claim has been made by Proctor and Healy (1985; also see Ratcliff, 1981). SIAM, in contrast, allows for the basis for similarity to change with time and, by conjecture, as a function of instructions. Proctor and Healy argued that their empirical results are most consistent with a model that bases both role-relevant and role-irrelevant responses on the same information. The implications of the results from Experiment 3 for this claim are explored in the discussion.

Method

Subjects.

Sixty-two undergraduate students from Indiana University served as subjects in order to fulfill a course requirement.

Materials.

Displays consisted of two crosslike configurations of four squares. Unlike the squares shown in Figure 6, the true squares varied in their colors, not in their shading pattern. The squares were filled with one of eight highly discriminable colors: blue, red, yellow, gray, turquoise, brown, green, and orange. Each square had 2-cm sides. The positions of the two crosses were randomized under the constraint that the two squares were always more than 6 cm apart. Viewing distance was not controlled but was approximately 60 cm.

Design.

Each subject saw 12 repetitions of a randomized block with 25 trials. On 12 trials within the block, displays consited of two identical crosses with the same four colors in the same positions. On the remaining 13 trials, each of the methods shown in Table 2 was used once. Each cross can be represented by a four-letter string, with each letter referring to a particular color and each letter position referring to a particular square position (i. e. , top, bottom, left, and right positions). One of the crosses (the initial cross) can always be described by the string ABCD. Each of the 13 methods altered the initial cross in a particular manner in order to create a changed cross. The initial and changed crosses shared between 0 and 4 MIPs and MOPs. A MIP was defined as a color match between squares that were in identical relative positions in their crosses. A MOP was defined as a color match between squares that did not have the same relative position. For example, the crosses ABCD and BCGD shared three colors in common (B, C, and D), but only one color (D) was in the same position in the two crosses. Figure 6A is an instantiation of Method 12, and Figure 6B instantiates Method 2.

Where possible, every combination of MIPs and MOPs that resulted in four or fewer cumulative color matches was included. Not all combinations of MIPs and MOPs were possible. For example, no display with 3 MIPs and 1 MOP could be created unless a two-to-one mapping was permitted. The particular colors represented by the letters, the square positions represented by the letter positions, and the order in which the methods were presented were all randomized.

Procedure.

On each trial, the two crosses were simultaneously presented. Subjects were instructed to press a key with one hand if the two crosses were the same and to press a second key with the other hand if the two crosses were different. The 31 role-relevant subjects were instructed that the two crosses were the same only if they contained the same colors in the same positions. The 31 role-irrelevant subjects were instructed that the two crosses were the same as long as they contained the same colors, regardless of the colors' positions. Both groups of subjects were told to respond as quickly as possible without sacrificing accuracy.

Results

The data of primary interest are shown in

Error and response time (RT) results of Experiment 3. MIP = match in place.

Figure 7. The similarity of the two crosses was measured by the percentage of trials in which the two crosses received a "same" response and by the time required to respond correctly that the two crosses were different.

Error data.

For most "different" judgments, the error rate was substantially lower for the role-relevant than for the role-irrelevant group. Analyzing only the nine methods of changing scenes that created a 3 × 3 factorial design, we found that both MIPs and MOPs influenced the error rate of the role-relevant group: $F(2,60) = 5.2$, $MS_e = 1.4$, $p < .05$, for MIPs, and $F(2,60) = 4.3$, $MS_e = 1.7$, $p < .05$, for MOPs. In addition, there was a MIPs × MOPs interaction, $F(4,120) = 3.5$, $MS_e = 2.1$, $p < .05$. The interaction indicates that MOPs had a greater influence on error rate when there were 2 MIPs, rather than no MIPs or one MIP. This may have been due to ceiling effects (low error rates) when crosses had few MIPs. Considering only methods that yielded no MIPs or one MIP, we found no evidence for any influence of MOPs on error rate. For the role-relevant group, MIPs were more influential than MOPs. We could compare all displays that had the same number of cumulative matches (MIPs + MOPs). For example, displays with 3 MIPs/0 MOPs, 2 MIPs/1 MOP, 1 MIP/2 MOPs, and 0 MIPs/3 MOPs all had a total of three feature matches. Comparing such displays, we found a strong influence of match distribution such that greater confusability occurred as the number of MIPs increased. There was a significant linear trend between the number of matches that were MIPs and confusion errors, $F(3,90) = 7.1$, $MS_e = 1.7$, $p < .05$. Very few errors occurred for 3 MOPs/0 MIPs displays, but many confusions arose for 0 MOPs/3 MIPs displays. For role-relevant subjects, a color match's influence on the confusability of the crosses depended significantly on whether it was a MIP or a MOP.

This same dependence was not found for the role-irrelevant group. Considering only "different" judgments, we found no significant linear trend between number of matches that were MIPs at a given number of cumulative matches and error rate, $F(3,90) = 1.9$, $MS_e = 1.8$, $p > .10$. For example, error rates were roughly equal for 2 MIPs/0 MOPs, 1 MIP/1 MOP, and 0 MIPs/2 MOPs displays. There was, however, an influence of type of match on "same" judgments, $F(3,90) = 3.2$, $MS_e = 2.4$, $p < .05$. For role-irrelevant "same" judgments, there were always four matching colors; confusion errors increased as the number of color matches that were MIPs increased.

For the role-irrelevant group, the same 3 × 3 factorial design could not be used, because the display with 2 MIPs and 2 MOPs demanded a "same" judgment, whereas the other displays demanded "different" judgments. Examination of the 2 × 3 (no MIP and one MIP vs. no MOP, one MOP, and two MOPs) factorial design revealed significant influences due to both MIPs and MOPs: $F(1,30) = 6.2$, $MS_e = 2.5$, $p < .05$, for MIPs, and $F(2,60) = 5.9$, $MS_e = 2.0$, $p < .05$, for MOPs. Compared with the role-relevant group, there was a greater influence of MOPs, $F(2,60) = 4.9$, $MS_e = 1.8$, $p < .05$. Whereas MOPs had an influence for the role-relevant group only when two MIPs were present, they had a strong influence for the role-irrelevant group at all MIP levels.

Response time data.

Data were also obtained on the time required to respond correctly that the two crosses were the same or different. With a few exceptions, the response time data pointed to the same conclusions as the error data. For the role-relevant group, MIPs and MOPs both influenced response time, with the influence of the former much greater than that of the latter. Concentrating on the 3 × 3 factorial, response time varied significantly as a function of the number of MIPs, $F(2,60) = 8.1$, $MS_e = 51$, $p < .05$, and MOPs, $F(2,60) = 4.6$, $MS_e = 60$, $p < .05$. The response time measure was somewhat more sensitive to MOPs than the error measure. The differential influence of MIPs could again be determined by comparing displays with the same number of total color matches. Once again, there was a significant linear trend that related the number of MIPs at a particular level of total matches to response time, $F(3,60) = 5.7$, $MS_e = 64$, $p < .05$.

For the role-irrelevant group, examining the 2 × 3 (Factor 1: no MIP and one MIP; Factor 2: no MOP, one MOP, and two MOPs) subdesign revealed significant influences of both MIPs, $F(1,30) = 4.7$, $MS_e = 48$, $p < .05$, and MOPs, $F(2,30) = 4.3$, $MS_e = 52$, $p < .05$, on response time. Compared with the role-relevant group, there was a greater influence of MOPs, $F(2,60) = 4.3$, $MS_e = 62$, $p < .05$. Considering only "different" judgments, there was no significant linear trend between the number of matches that were MIPs for a particular number of total matches and response time, $F(3,90) = 1.4$, $MS_e = 55$, $p > .05$. The influence of type of match on "same" judgments was significant, $F(3,90) = 4.8$, $MS_e = 51$, $p < .05$.

Relation between error and response time data.

With a few exceptions, the error and response time data mirrored each other well. As response time for a display increased, error rate for the display usually increased as well. This pattern counters a general speed/accuracy tradeoff. However, there was one noteworthy exception. For both order-relevant and order-irrelevant groups, "same" judgments were slower but more accurate than "different" judgments. For the order-irrelevant group, "same" judgments produced lower error rates ($M = 4.6\%$) than "different" judgments (16.0%) but produced response times that were comparable to "different" judgments (1192 ms for "same" judgments vs. 1226 ms for "different" judgments). For the order-relevant group, "same" judgments (1423 ms) required more time than "different" judgments (1030 ms), but they were also fairly accurate (5.0% and 6.3% error rates for "same" and "different" judgments, respectively).

Discussion

The most important result from Experiment 3 is that the role-irrelevant and role-relevant groups produced quite different assessments of the relative similarity of different displays. By both response time and error measures, the role-relevant group was more influenced by MIPs and less influenced by MOPs than was the role-irrelevant group. The role-irrelevant group showed no differential influence due to MIPs versus MOPs on "different" judgments. Crosses that shared 0 MIPs/3 MOPs were as similar as crosses that shared 3 MIPs/0 MOPs. In contrast, the role-relevant group showed marked differences in the influences of MIPs and MOPs for the same "different" displays.

The instruction manipulation caused shifts in the rank order of display similarity. There were several cases in which the crosses of Display A were more similar to each other than were the crosses in Display B for the role-relevant group but not for the role-irrelevant group. For example, the crosses of Display 2 (see Table 1) were more similar than the crosses of Display 12 for role-irrelevant, but not role-relevant, subjects. Similar reversals occurred for Displays 2 and 8 and for Displays 4 and 11. In all cases, the display with more MIPs was more similar for the role-relevant group, whereas the display with more MOPs was more similar for the role-irrelevant group.

The influence of task instructions on display similarity provides difficulties for the claim that a single similarity estimate underlies response time and error rate performance in both role-relevant and -irrelevant tasks. Proctor and Healy (1985) indicated that their empirical data were consistent with this claim (also see Ratcliff, 1981), although they qualified the claim on empirical grounds not related to the present results. Proctor and Healy required subjects to say whether two strings of three letters were the same or different according to role-relevant or role-irrelevant rules. Their basic evidence in favor of a single pool of similarity information was the typical "mirroring" of the tasks; task variables that made role-irrelevant subjects slow in responding that two strings were the same also speeded role-relevant subjects' "different" judgments. In general, Proctor and Healy's data suggest that pairs of letter strings have a single measure of similarity and that the different tasks simply have different "cutoff" points as to what level of similarity is required to respond "same." If a pair of letter strings is close to the cutoff point, the response is slow and error prone.

Because the present results clearly oppose the claim that a single similarity estimate is sufficient to handle both instruction groups, the question remains, Why do Proctor and Healy's (1985) study and the present experiments yield discrepant outcomes? Essentially, the answer seems to lie in the particular displays used. Proctor and Healy may have failed to find evidence that role-relevant subjects were more influenced by the difference between MIPs and MOPs than were role-irrelevant subjects because their design did not permit the comparison of displays that (a) matched MIPs against MOPs and (b) required the same response for each instruction group. In our terminology, Proctor and Healy's three-letter displays fall into four patterns of match distribution: 3 MIPs/0 MOPs, 1 MIP/2 MOPs, 0 MIPs/3 MOPs, and 2 MIPs/0 MOPs. Only the last pattern is categorized as "different" by role-irrelevant subjects; only the first pattern is categorized as "same" by role-relevant subjects. In Experiment 3, there were trade-offs between MIPs and MOPs that did not change a subject's correct response. For example, we were able to determine whether 1 MIP or 2 MOPs increased the similarity of displays more for both role-relevant and -irrelevant subjects. This is not possible with the four display types used by Proctor and Healy, because the two instruction groups gave the same response to only two of the displays (the display with 3 MIPs/0 MOPs was called "same" by both groups, and the display with 2 MIPs/0 MOPs was called "different" by both groups), and these displays varied only on number of MIPs, not number of MOPs.

In short, the present results show that when MIPs and MOPs are allowed to vary independently in displays that receive the same response by both groups, they are not equally important for the two groups. The influence of MIPs and MOPs varies as a function of the subjects' task. Interestingly, it also appears that the influence of MIPs and MOPs is not identical for "same" and "different" judgments. For the role-irrelevant group, whether a color match was a MIP or a MOP influenced both the speed and accuracy of "same" judgments but did not influence either the speed or accuracy of "different" judgments. A potentially similar finding is that subjects are more influenced by relational properties for similarity than for difference judgments (Medin, Goldstone, & Gentner, 1990). For example, a scene with two squares is judged to be more similar to and more different from (for different subjects) two squares than a square and a circle. Medin et al. proposed that the same-shapes relation influences similarity judgments more than difference judgments. Similarly, Experiment 3 suggests that the relational role of a color matters more for speeded "same" judgments than for "different" judgments. This result, together with the tendency for slow but accurate "same" judgments, suggests that "same" and "different" judgments may be based on different processes (for a review of dual-process models, see Farell, 1985). In summary, MIPs and MOPs appear not to be equally weighted for role-relevant and role-irrelevant tasks and also appear not to be equally weighted for "same" and "different" judgments.

Theoretical Analysis

All three experiments provide general support for the role of alignment and evolving global constraint satisfaction associated with SIAM. Before considering the implications of these results for theories of similarity, we wish to present a more detailed analysis of SIAM. We begin with quantitative modeling, contrasting SIAM's predictions with those of a number of alternative models that attempt to address similarity in a less dynamic manner.

Experiment 1

Experiment 1 allows for a quantitative analysis of SIAM because of the fairly large number of displays. Thirteen displays were presented at each of three deadlines, yielding 39 data points to model. Because the full version of SIAM has too many parameters relative to the number of data points, default values were assigned to all parameters internal to SIAM (as described by Goldstone, 1994). [2] However, free parameters are still required to apply SIAM to the paradigm of speeded responses under deadline. SIAM's estimates are incorporated in a boundary-crossing model of same, different, and overtime responses (for another example of a boundary-crossing model, see Busemeyer & Rapoport, 1988).

The boundary-crossing model used to supplement the similarity estimate of the similarity as interactive activation and mapping (SIAM) model. If SIAM's estimate crosses a lower boundary, a "different" response is given; if it crosses an upper boundary, a "same" response is given. Parameters are as follows: M = 0. 4, short-distance = 0. 1, long-distance = 0. 3, cycles per second = 10, and K = 1. 12.

Figure 8 is a graphic depiction of the boundary-crossing model. Initial upper and lower boundaries are chosen. If SIAM's estimated similarity for a display, combined with Gaussian random noise (with a mean of 0 and a variance of N), [3] falls above the upper boundary, a "same" response is given. If SIAM's estimated falls below the lower boundary, a "different" response is given. If neither of these events occurs, SIAM enters another cycle of activation adjustment, and the boundaries are adjusted.

Boundaries need to be adjusted because similarity generally increases with cycles in SIAM, and a more stringent criterion for responding "same" is required as the amount of processing increases. A subject may quickly respond "same" if two scenes appear moderately similar. However, if the subject examines the scenes for a long time, virtual identity should be required before the subject responds "same. " Similarly, a "different" response may be made after a long period of processing even though the scenes are highly similar—a single difference suffices for a "different" response.

To model the effect of processing time on the decision criteria, we assume the bounds to vary according to where boundary*C* is the upper or lower boundary used to obtain SIAM's estimate at Cycle *C, I* is the initial value of the boundary, and *L* is the cycle when the upper and lower boundaries converge to 1. 0. This equation dictates that both the upper and lower boundaries will increase linearly, with the lower boundary increasing with a greater slope. Ideally, the subjects' deadline would be at *L.* However, subjects frequently give overtime responses. These are accommodated by including another parameter, cycles per second. The cycles-per-second parameter converts the number of seconds that subjects are given for their decision to the number of cycles of activation adjustment that SIAM is allowed. For example, if the number of cycles per second is set at 20, SIAM will model the data from the three experimental deadlines (1 s, 1. 84 s, and 2. 68 s) by setting a deadline *D* at 20, 37, and 54 cycles, respectively. [4] If SIAM's estimate has not exceeded the upper or lower boundary by *D,* the response is considered to be overtime. If *D* were equal to *L,* no overtime response would be possible. If *L* is set equal to *D* multiplied by the constant *K, L* can be greater than *D,* allowing for trials on which similarity has not exceeded either boundary by *D.*

The instructional manipulation of deadline is directly modeled by *D,* but deadline also must influence the speed of an average response. To this end, the initial upper and lower boundaries also depend on deadline. A mean value between the upper and lower boundaries, *M,* is parametrically set. (If *M* is relatively high, there will be a bias to respond "different. " If *M* is low, there will be a bias to respond "same".) Three parameters—short distance, medium distance, and long distance—refer to the distances of the initial upper and lower deadlines to *M* for the short, medium, and long deadlines, respectively. In general, we anticipate that the upper and lower boundaries will be closer to *M* for the short than for the long deadline. The closer the boundaries are to *M,* the sooner they will be crossed by SIAM's estimate.

In summary, the data from Experiment 1 are modeled by running a default version of SIAM, supplemented by a seven-parameter process model of "same"-"different" decisions. SIAM was run 100 times for each of the 39 display-deadline combinations. From these 100 runs, percentages of correct "different" responses, incorrect "same" responses, and overtime responses were determined. The parameters *N, M,* short distance, medium distance, long distance, cycles per second, and *K* were fit so as to minimize the root mean square error of SIAM's predictions of the empirically observed percentages. [5] The other parameters (*D* and *L*) can be derived from the number of cycles per second and *K.*

The best fitting SIAM estimate has a correlation of (r) . 973, with N = 0. 34, M = 0. 63, short distance = 0. 12, medium distance = 0. 20, long distance = 0. 23, cycles per second = 12, and K = 1. 72. The overall fit is extremely good but perhaps is best gauged when compared with competing models. Other theoretically motivated but simple methods for computing scene similarity/confusability were developed. These models all involve a linear regression with confusability/similarity as the predicted variable.

First, we can assign separate parameters for the influence of total number of matching features at each of the three deadlines. This four-parameter model obtains a correlation of . 87.

Second, we can consider a model that explicitly distinguishes between features matches that are in and out of place. Such a model agrees with our main thesis that object alignment is an important determinant of the influence of matching features between the scenes on similarity. This model is instantiated by a linear regression with number of MIPs, number of MOPs, deadline, and intercept as predictors of confusions. This model obtains a correlation of . 92. The regression weights for number of MIPs and number of MOPs were 11. 1 and 2. 9, respectively. Both of these values were significantly different from 0, $t(37) > 2.6$, $p < .05$, corroborating that both MIPs and MOPs have a significant influence on confusions. In addition, the regression weights for MIPs and MOPs were significantly different, $t(37) = 4.3$, $p < .05$, indicating that MIPs have significantly greater influence on confusions within this model than do MOPs. This differential weighting of MIPs and MOPs reveals that a large part of the superior fit of this second model is due to its distinction between MIPs and MOPs.

The third and last model that was tested also makes a distinction between MIPs and MOPs. In addition, it allows for the possibility that the influences of MIPs and MOPs vary independently with deadline. This model is able to instantiate SIAM's prediction that MOPs are relatively influential for similarity early in processing and MIPs are relatively influential later in processing. A linear regression was applied with separate predictors for number of MIPs and number of MOPs at each of the three deadlines. The resulting best fitting equation was as follows: Percent errors on "different" trials = 7. 4 (fast-deadline MOPs) + 14. 1 (fast-deadline MIPs) + 0. 9 (medium-deadline MOPs) + 10. 2 (medium-deadline MIPs) + 0. 5 (slow-deadline MOPs) + 9. 2 (slow-deadline MIPs).

The preceding linear regression achieved a correlation of . 956, and the set of six predictors, when added to any of the previous models, significantly increased the quality of the models' fits. This third model contains several points of interest. First, the weights of the MIPs terms are uniformly larger than weights of the MOPs terms. This reconfirms the benefit of an analysis that distinguishes between feature matches that do and do not belong to aligned objects. Second, the weights of both MIP and MOP terms decrease with increasing deadline. This is due to the low error rates that are observed for trials with slow deadlines. Third, and most important, the weights of MOP terms decrease more rapidly with longer deadlines than do the weights of MIP terms. Whereas deadline has a significant influence on linear regression weights when only the MOP terms are considered, $F(2,37) = 3.7$, MS_e = 0. 02, $p < .05$, there is no corresponding influence of deadline when only the MIP terms are considered, $F(2,37) = 1.4$, MS_e = 0. 03, $p > .10$. Thus, it appears that the quality of this model's fit is attributable to its flexibility in allowing MIPs and MOPs to vary in their relative influence with processing time. The four error overestimates that exceed 2 standard deviations occur for Display 13 at a slow deadline, Display 13 at a medium deadline, Display 4 at a fast deadline, and Display 5 at a fast deadline. The two error underestimates that exceed 2 standard deviations occur for Display 10 at a fast deadline and Display 9 at a slow deadline. A general characterization of the pattern of residuals seems to be that the model underestimates the influence of large number of MIPs at slow deadlines and overestimates the influence of MIPs at fast deadlines.

SIAM's fit is superior to this third model. The predicted and observed values for the third model and SIAM are shown in Table 1. Adding a term for SIAM's estimate significantly increases the third model's fit, $F(2,37) = 4.1$, MS_e = 0. 02, $p < .05$, and adding an combination of the terms from the model to SIAM's estimate does not increase SIAM's fit. Although the difference in correlations is not large (. 973 for SIAM vs. . 956 for the third model), when correlations are close to 1. 0 small differences signify large differences in the quality of model fits. Both the third model and SIAM have seven degrees of freedom.

SIAM does not overestimate the same patterns that the third model does. SIAM is able to accommodate the high error rates for Display 13 at slow and medium deadlines. Display 13 contains seven MIPs; it is important to note that two of the scenes' objects are identical. Because two objects have identical features and no mismatching features, SIAM's object-to-object node that represents the alignment of these two objects will be strongly activated. As a consequence, the individual feature matches shared by the two objects will receive a large amount of excitatory support from the object correspondences. This, in turn, results in a high similarity estimate, even after many cycles of activation adjustment.

SIAM also provides an account of the high error rates given to displays with few MIPs (Displays 4 and 5) at fast deadlines. The third model is forced to predict fairly low error rates for these displays because the parameter for MIPs at fast deadline must receive a much higher value than MIPs at slow deadline in order to explain the very different confusabilities of Displays 11 and 13 or Displays 8 and 12 (for example). But, given that the MIPs-at-fast-deadline parameter is weighted heavily, displays with very few MIPs must be given low similarity/confusability assessments. SIAM is not forced into this dilemma because it does not assume that all MIPs are equally quickly placed in alignment. When scenes share many MIPs (e. g. , Display 13), the proper object alignments are determined very quickly. Objects that share many MIPs will soon be placed in alignment and will feed activation back to nodes that represent the feature matches. However, if scenes share relatively few MIPs (e. g. , Displays 4 and 5), no objects will be placed in strong alignment immediately. Therefore, the MIPs will not receive significant fed-back activation from object correspondences. In short, SIAM is not committed to the assumption that all MIPs at a given deadline receive the same weight. The large influence of MIPs relative to MOPs can take a long time to develop if the objects are not placed in strong alignment, but it can develop rapidly if the objects are placed in strong alignment.

Experiment 2

The results from Experiment 2 are not amenable to detailed quantitative fits by SIAM because there are too few data points. There were only three different display types, each shown at three deadlines. SIAM qualitatively models the results well. Under virtually all parameter settings, SIAM predicts that with time (number of cycles of activation adjustment in SIAM), scenes with more globally consistent feature matches become increasingly similar compared with scenes with more locally determined feature matches. As discussed earlier, the results are problematic for models that claim that simple features and conjunctive features are both encoded and that simple features are registered more quickly than conjunctive features, as well as models that assert that features align simply on the basis of their objects' similarity without regard for other objects' similarities.

Experiment 3

To model the results from Experiment 3, we must incorporate role-to-role correspondences into SIAM. Scenes are described by four objects that contain one color feature each. The four objects serve as arguments to the following schema: cross (upper-arm role, lower-arm role, left-arm role, and right-arm role). Thus, one scene might be represented as follows: cross (Object 1 [color = blue], Object 2 [color = red], Object 3 [color = yellow], Object 4 [color = gray]). As described earlier, SIAM tends to place objects in correspondence that occupy identical roles in their crosses and that have identical colors. The first tendency is controlled by the parameter role to object weight, and the second is controlled by the parameter feature to object weight. In the current modeling, feature to object weight is assigned a default value of 1. 00, and role to object weight is fit as a free parameter.

The parameter role to object weight is the critical parameter used for modeling the difference between Experiment 3's role-irrelevant and role-relevant groups. The assumption made in the modeling is that the difference between these groups can be characterized in terms of the weight given to roles in determining feature matches. Hypothetically, the role-relevant group's behavior should be better modeled by a higher value for role to object weight than is given for the role-irrelevant group. Thus, two parameters, role to object weight when roles are relevant and role to object weight when roles are irrelevant, are used to model the results from the two groups.

To accommodate the same-different error and response time data, we augmented SIAM with a simplified version of the processing model used to model the results from Experiment 1. Once again, upper and lower boundaries are initially set equal to *M* + distance and *M* - distance, where both *M* and distance are free parameters. Experimentation also showed that a single *M* value for both role-irrelevant and role-relevant groups was insufficient. Consequently, a separate value of *M* was found for each group. Having separate *M*s allows SIAM to model different "same" and "different" criteria for the groups. When SIAM's estimate, combined with Gaussian random noise (with a mean of 0 and a variance of *N*), exceeds the upper boundary, a "same" response is emitted. When the estimate and noise fall below the lower boundary, a "different" response is emitted. As before, unsatisfactory fits were obtained when the boundaries were assumed to remain constant with processing. Instead, it is assumed that both boundaries rise linearly, simulating the intuition that more demanding evidence is required for a "same" judgment as processing continues. The parameter *L* stands for the cycle number at which the upper and lower boundaries meet at 1. 0. [6] On each cycle of activation adjustment, the upper and lower boundaries are incremented. In addition to recording SIAM's response ("same"-"different"), we also record the number of cycles required to cross a boundary.

All in all, seven parameters (*N, L, M* irrelevant, *M* relevant, distance, and role to object weight for role-irrelevant and role-relevant groups) were freely fit so as to minimize root mean square error. SIAM was run 100 times on 20 displays presented in Experiment 3. Only displays that were assigned a correct response of "different" for both groups were modeled. Averaging across the 100 separate runs, the percentage of correct "different" responses, the percentage of incorrect "same" responses, and the average number of trials required for each type of response were recorded. When applied to the error data (the percentage of incorrect "same" judgments on "different" displays), the best fitting SIAM estimate had a correlation of . 944, with *N* = 0. 29, *M* roles relevant = 0. 70, *M* roles irrelevant = 0. 64, *L* = 38, distance = 0. 17, role to object weight when roles are irrelevant = 1. 5, and role to object weight when roles are relevant = 2. 4. For purposes of comparing this fit with that of other models to be developed, it is also informative to fit a five-parameter version of SIAM that sets the parameter role to weight when roles are irrelevant to its default value of 1. 0 and sets *N* to the previously found value of 0. 34. When this is done, the correlation only drops to . 942. The additional two parameters do not significantly add to the fit of the five-parameter version.

When applied to the response time data (only "correct" trials), the best fitting SIAM estimate had a correlation of . 980, with *N* = 0. 26, *M* roles relevant = 0. 74, *M* roles irrelevant = 0. 67, *L* = 33, distance = 0. 13, role to object weight when roles are irrelevant = 1. 2, and role to object weight when roles are relevant = 2. 7. Again, the five-parameter version of SIAM did not fit the data significantly worse than this model.

We developed two alternative linear regression models to appraise SIAM's fit. The first model predicts the percentage of incorrect "same" judgments on "different" displays as a weighted function of number of MIPs, number of MOPs, group (a variable that codes whether role-relevant or role-irrelevant instructions were given), and an intercept term. The model achieved a correlation of . 856 for the error data and . 912 for the response time data (with different parameter values for the two fits).

The second model makes a distinction between MIPs and MOPs and further distinguishes between the relative influence of MIPs and MOPs for the two groups. This model can capture SIAM's basic prediction that the influence of MIPs relative to MOPs is greater when roles are relevant than when roles are not relevant. Thus, this five-parameter model assigns separate values for intercept, role-relevant number of MIPs, role-relevant number of MOPs, role-irrelevant number of MIPs, and role-irrelevant number of MOPs.

The best fitting equation for predicting error rate—percent "same" responses = 0. 24 + 4. 39 (role-relevant MIPs) + 0. 477 (role-relevant MOPs) + 8. 04 (role-irrelevant MIPs) + 7. 69 (role-irrelevant MOPs)—obtained a correlation of . 921. The difference between the weights of MIP and MOP terms is much greater for the role-relevant than for the role-irrelevant displays. This corroborates the empirical result that the difference between MIPs and MOPs is much more important when roles are relevant.

The best fitting equation for predicting response time—"different" response time (milliseconds) = 821 + 163 (role-relevant MIPs) + 27 (role-relevant MOPs) + 195 (role-irrelevant MIPs) + 219 (role-irrelevant MOPs)—obtained a correlation of . 972 and even more dramatically depicts an interaction between the relative weights of MIPs and MOPs and the group instructions.

SIAM's estimate, when added to this second model, significantly increases the fit of this model, considering either error or response time data results. Conversely, this second model only increases the fit of SIAM's prediction of error rate. The pattern of residuals from the second model is diagnostic. The three largest overestimates of the model on the error data occur on displays with two MIPs and 0 MOPs (Display 12 in Table 2) role irrelevant, 2 MIPs and 0 MOPs role relevant, and 1 MIP and 1 MOP role irrelevant. Similarly, the three largest overestimates of the model on the response time data occur on displays with 2 MIPs and 0 MOPs role irrelevant, 2 MIPs and 0 MOPs role relevant, and 1 MIPs and 0 MOPs role relevant. The three largest underestimates of the model for both the error and response time data occur on displays with 0 MIPs and 0 MOPs role irrelevant, 3 MIPs and 0 MOPs role relevant, and 1 MIP and 2 MOPs role irrelevant. Thus, the second model mispredicts both the error and response time data in almost identical patterns. An interpretation of the pattern of residuals, obtained by consulting Figure 5, would be that the model has difficulty simultaneously explaining the large difference between 3 MIPs and 2 MIPs and the much smaller difference among 0, 1, and 2 MIPs. The compromise parameter values that best fit the data underestimate the difference between 3 MIPs and 2 MIPs by overestimating the similarity of 2-MIPs displays and underestimating the similarity of 3 MIPs displays. Likewise, the model overestimates the differences between smaller number of MIPs by underestimating the similarity of 0-MIPs displays and overestimating the similarity of 2-MIPs displays.

SIAM has a positive-feedback process that naturally accommodates the nonlinear influence of MIPs on similarity. When scenes share 3 MIPs, the MIPs serve to support and excite each other. This is due to within-layer excitatory connections between nodes. Object-to-object, and role-to-role, correspondences that are consistent will be mutually excitatory. A MIP that comes from a display with 2 other MIPs will receive excitation from two within-layer sources, whereas a single MIP will not receive any strong within-layer excitation.

Summary of Fits

In this theoretical analysis, we have made several important points. First, models for this task need to distinguish between MIPs and MOPs and to take into account both locally and globally determined matches. Second, the contributions of these factors is dynamic and changes with processing time. Beyond that, the greater success of SIAM than the alternative models shows that a detailed process model is needed to capture the full influence of MIPs, MOPs, and time on performance.

General Discussion

General Advantages of SIAM

SIAM's estimates significantly increase the fit of the other models, but, with one exception (the error data from Experiment 3, when compared with the five-parameter model) the other models do not significantly increase SIAM's fit if they are included in SIAM. Furthermore, SIAM is able to avoid some of the systematic errors of even the most successful alternative models. SIAM naturally accommodates the nonlinear influence of MIPs on similarity (from Experiment 3) and the fact that the importance of a MIP depends on the strength of object alignments (from Experiment 1). There are, however, a number of advantages to SIAM, apart from its superior quantitative fits.

The most important advantage of SIAM is that it provides an account of alignment. The other models either do not make a distinction between aligned and unaligned features or use hand-coded representations of MIPs and MOPs. SIAM's input is simply the description of two scenes. SIAM, in the course of its processing, determines whether a feature match is a MIP or a MOP. By contrast, feature matches are labeled "in place" or "out of place" for the other models that make a distinction between MIPs and MOPs. SIAM's input, therefore, is in a less preprocessed form. SIAM organizes a display into element-to-element correspondences instead of relying on a preprocessor to identify MIPs and MOPs.

In addition, SIAM is able to predict both response time and error data by a single process model. The other models could fit both types of data, but only by reestimating their parameters. SIAM's fit decreases modestly when it is forced to predict response time and error data simultaneously, but the fits are still comparable to those obtained from the best fitting model that maximizes only a single set of data.

Third, SIAM provides an explanation of why processing time has the effect it does. The best fitting models for Experiment 1's results all allow differential weighting of MIPs and MOPs, depending on the deadline. However, only SIAM provides a mechanism for this interaction. According to SIAM, MIPs become more influential with time relative to MOPs because the proper (globally consistent) object alignments take time to develop.

Fourth, SIAM is more constrained than the other models. Measuring the flexibility of a model simply by counting the number of free parameters is crude at best. Although SIAM has as many free parameters as the most complex alternative models, it is also considerably more constrained in its predictions. Whereas the other models could have modeled the result that MIPs become relatively less influential with time, SIAM could only have modeled the interaction that was empirically obtained. Similarly, SIAM predicts that the difference between deadlines of 1 and 1. 84 s must be greater than the difference between 1. 84 and 2. 68 s despite their constant interval. The asymptotic boundaries on node activations causes activations to change most quickly early in processing. On the other hand, the other models considered could have assigned any regression weights to the different deadlines. SIAM makes correct predictions with respect to nonlinear influences of MIPs, the time course of locally and globally consistent matches, and the influence of task instructions on MIPs and MOPs. Although some of these results are compatible with the other models considered, they are not genuine predictions of the alternative models.

Applications of SIAM

Task manipulations in SIAM have been modeled by making specific changes to a limited number of parameters. Experiment 1's deadline conditions were modeled by manipulating the slopes of the lower and upper boundaries, making them converge after many of few cycles. Experiment 3's task manipulation was modeled by allowing the influence of roles on objects to vary (and by altering the initial lower and upper boundaries). The research strategy has been to use the same basic processing model with slight variations in order to account for different tasks. Consequently, one powerful aspect of SIAM is that it potentially provides a unified account of different techniques of assessing similarity. Different measures of similarity (e. g. , similarity ratings and speeded "same"-"different" judgments) are strongly correlated, but they also seem to be different measures. For the most part, researchers have not attempted to provide formal accounts of the differences between tasks. Usually, they have ignored or discounted the differences between their similarity measures.

SIAM provides an alternative to (a) treating different measures of similarity as measuring the same construct and (b) treating different measures as completely independent. SIAM predicts strong correlations between measures insofar as they all involve a common process of alignment. SIAM predicts dissociations to the extent that measures use different parameter values or secondary processes. This modeling approach is not able to accommodate all of the task differences between different comparison tasks. However, the approach is useful because it permits analysis of tasks along particular dimensions. The dimensions that were parametrically varied in the simulations were the amount of processing time permitted and the importance of roles. In addition, both same-different response times and error rates can be modeled if SIAM is augmented with a boundary-crossing process. Other similarity tasks may be best modeled by variations of other SIAM parameters. Thus, SIAM provides a framework for comparing and contrasting tasks. Although we wish to avoid the strong claim that the requirements of all comparison tasks can be instantiated in SIAM, SIAM does provide a method for organizing the systematic relations between some tasks. The following four applications illustrate the use of SIAM in accounting for regularities between tasks.

Corter (1987). Corter presented the same materials in both similarity rating and same-different tasks. Although most of the materials from his Experiments 3 and 4 were arbitrarily created, a subset of the materials was designed to be similar to a particular stimulus. The four stimuli—A, B, C, and D in

Sample stimuli from

Figure 9—were designed to be similar to T. Interestingly, T was rated as most similar to C but was most often confused with B.

SIAM has a plausible, although post hoc, account for this difference in similarity measures. SIAM would be expected to execute fewer cycles of activation adjustment in the same-different task than in the similarity rating task because of the different time pressures. If similarity is estimated after few cycles have been executed, there will be little chance for relations to influence similarity. In this situation, T might be expected to be more similar to B than C. Treisman and Paterson (1984) have argued that line segments with specific orientations are psychological primitives, and Pomerantz (1986) has argued that oriented angles are primitives. Thus, the diagonal line and 45° angle shared by T and B would be encoded in SIAM as simple features. The more complex, vertically oriented horseshoe shape shared by T and C, on the other hand, is too perceptually complex to be a single feature and would consequently be represented as a relation or schema. Although the top bar of C and the bottom bar of T are not aligned strongly by their location similarity, they would be aligned by a shared role. Both bars are the middle argument of the schema horseshoe shape (left leg, horizontal connecting bar, right leg). The relational similarity of T and C would influence similarity more with time. As a result, T and C would seem more similar with similarity judgments than with speeded judgments. It is interesting to note that D may not receive as high a similarity rating as does C when compared with T because of a two-to-one mapping. The connecting bar of D aligns with either T's bottom bar or its right bar, depending on whether the alignment is feature or role driven. To the extent that these correspondences compete, similarity ratings are expected to suffer. SIAM cannot explain all of the results in Figure 9, but it does provide an account of the largest discrepancies between the similarity measures.

Beck (1966). Beck found a strong dissociation between similarity and perceptual grouping. A tilted T was judged to be more similar to an upright T than an X, but the tilted T was also more likely to be perceptually grouped with the X than the upright T. Again, similarity judgments are assumed to involve more cycles of activation adjustment than grouping decisions; subjects make the former judgment much more slowly. The T shapes can be represented as follows: T shape (bar subtended in middle, subtending bar). Specifically, the representations of the tilted T, T, and X may be as follows: T shape (135° bar, 45° bar), T shape (0° bar, 90° bar), and X shape (135° bar, 45° bar). By these representations, the tilted T and X would be similar because of their matching diagonal bars, but the tilted T and T would be similar because of their corresponding roles. A similar treatment can be given to Gati and Tversky's (1984) observation that 6 and 9 (in calculator-like fonts) were judged more similar than 6 and 8 although the latter pair was more frequently confused in a faster recognition task.

Palmer (1978). Palmer argued that multisegment figures are not represented in terms of independent line segments. Instead, figures are represented in terms of structural relations between line segments such as "closure" and "good continuation. " Palmer used sets of stick figures (similar to those used by Corter, 1987) that had the same number of shared line segments but different amounts of relational similarity.

Palmer (1978) found that common structural relations influenced similarity, as measured by a same-different task. In addition, the influence of structural commonalities was greater when figures were presented simultaneously than when they were presented sequentially. In simultaneous presentation, both the standard figure and the comparison figure were flashed on the screen at the same time. In sequential presentation, one figure appeared for 1 s, followed by a blank screen for 500 ms, followed by the second figure.

As in previous modeling, we assume that fast "different" judgments are made when SIAM's similarity estimate on "different" trials becomes low quickly. Roughly, structural relations can be represented in terms of role nodes, and object nodes can represent groupings of one or more line segments. In Palmer's (1978) study, sequential decisions were made much more quickly than simultaneous decisions, presumably because of the longer exposure of one the figures. Because of this general response time difference, SIAM would model the sequential task with fewer cycles of activation adjustment than the simultaneous task. Palmer's results can be taken as a special case of SIAM's general enhancement of relational properties with more processing time. Early in processing, SIAM is strongly influenced by featural similarity. On the first cycle, SIAM is influenced only by matching line segments; similarity would simply be a function of the number of overlapping line segments. Later, SIAM is influenced more by role correspondences. Again, this account is speculative, but at least it is testable.

Proctor and Healy (1985). Recall that Proctor and Healy argued that their results were consistent with a model that role information counts in the comparison of letter strings, but that it counts equally for role-relevant (called "order-relevant" by Proctor and Healy) and role-irrelevant comparisons. That is, the same similarity estimate can serve both comparison tasks (with a few caveats that are not relevant to the present discussion). Contrary to this conclusion, Experiment 3 showed that the MIPs were much more heavily weighted (relative to MOPs) for role-relevant displays than for role-irrelevant displays, indicating that the two tasks compute similarity differently. We can ask whether differential weighting of MIPs and MOPs can explain any results from Proctor and Healy's experiment that are difficult to handle by a single-similarity-estimate model. In fact, Proctor and Healy pointed out exactly such a result. In the role-relevant task, subjects responded that ABC and BCA were different more quickly and accurately than they responded that ABC and AXC were different. In the role-irrelevant task, subjects are much more likely to call ABC and BCA correctly the same than they are to call ABC and AXC incorrectly the same. In other words, if an error measure of similarity is used, ABC is more similar to BCA than AXC for role-irrelevant, but not role-relevant, subjects. As Proctor and Healy acknowledged this evidence indicates that the two tasks are not based on exactly the same similarity estimate.

SIAM predicts this task difference, assuming that roles correspond to relative positions within a letter string and that the role-relevant task is modeled by a relatively high role-to-object weight. ABC has 3 MOPs in common with BCA and 2 MIPs in common with AXC. For the role-irrelevant task, scenes with 3 MOPs may be more similar than scenes with 2 MIPs because of the fairly large influence of features on object correspondences. For the role-relevant task, scenes with 2 MIPs may be more similar than scenes with 3 MOPs because of the strong influence of role alignments on object correspondences. Not all of the data reported by Proctor and Healy (1985) can be accounted for by SIAM supplemented with the assumption concerning role-to-object weight. Still, an analysis of their data suggests that the order-irrelevant and order-relevant tasks do not always use the same similarity information. Despite this difference, the pattern of results from both tasks can be modeled by SIAM if a reasonable assumption concerning parameter values is made.

Summary.

We do not wish to claim that similarity ratings, assorted same-different judgment tasks, and other measures of similarity all use exactly the same process. There are nontrivial processing differences between the tasks that cannot be modeled simply by changing parameter values of SIAM. Nonetheless, SIAM can be used as a tool for comparing and contrasting different tasks along particular dimensions (i. e. , how long processing takes, how important roles and simple features are, and whether there are biases to respond in particular ways). To the extent that SIAM can model different conditions of a single task or different tasks, we have a simple and elegant method of analyzing common and distinctive characteristics of similarity measures.

Related Work

The present data and theoretical analyses argue for a prominent role of alignment in the comparison of scenes with multiple parts. In addition, Experiments 1 and 2 provide insight into the time course of the alignment process. Early in processing, the weights of feature matches seem to be based on the locally determined featural similarity of the matches' objects. Later in processing, the weights of feature matches become more sensitive to the global consistency of object correspondences. With globally consistent object correspondences, each object in one scene corresponds to a single object in the other scene. SIAM incorporates this global consistency constraint by a mechanism similar to one used by Holyoak and Thagard (1989). Nodes representing inconsistent correspondences inhibit each other. Object correspondences are first determined by the only information available—feature matches. Once object correspondences are developed, they begin to influence each other, yielding correspondences (depending on parameter values) that are optimally consistent. As object correspondences begin to redirect each other's activation, matches between aligned objects (MIPS) become particularly important for determining similarity (Experiment 2).

SIAM's basic local-to-global processing principle is possessed by the SME (Falkenhainer et al. , 1990) and ACME (Holyoak & Thagard, 1989) models of analogical reasoning. In ACME and SME, there are pressures against developing many-to-one mappings, and pressures in favor of developing mutually consistent mappings. In many respects, the present work is a quantitative model that extends this work in analogy to more quantitatively literal/perceptual forms of similarity (see also Gentner, 1989; Gentner & Ratterman, 1991; for other evidence that perceptual similarity is more sophisticated than might be thought, see Smith & Heise, 1992). In addition, the notion of determining globally consistent correspondences operates in what Marr (1982) called "cooperative algorithms. " For example, in Marr and Poggio's (1979) model of depth perception or Ullman's (1979; also Dawson, 1991) model of apparent motion, correspondences between visual elements are computed in a manner that yields correspondences that are maximally consistent with each other.

Several researchers have argued that different sources of similarity information are available at different points in processing. Gillund and Shiffrin (1984) proposed a fast process that determines a probe's overall similarity to all items in memory and then gives way to a slower item-by-item similarity computation. Dosher (1984) provided evidence that, when subjects judge whether two words are paired in a list, semantic similarity gives way to retrieval of specific associations. Sekular and Palmer (1992) found that before 200 ms, a visual object tended to prime isolated "mosaic" elements; after 200 ms, the object primed its complete, globally consistent, interpretation. Evidence from Gronlund and Ratcliff (1989) suggests that single-item information can be retrieved more quickly than can information about associations between items. Ratcliff and McKoon (1989) found that discriminating sentences on the basis of their words occurred early in processing, but discriminating sentences on the basis of relation between words occurred only after 600-700 ms. They argued that their results are problematic for models that assume only a single similarity value that underlies judgments at all points in processing.

Consistent with Ratcliff and McKoon's (1989) conclusions, the present results emphasize the explanatory inadequacy of similarity assessments that do not change with processing. In Experiment 2, whether the lower left scene or the lower right scene of Figure 4 was more confusable with the target scene depended on how long subjects were given to make their decisions. Any model that assigns a fixed similarity value to each of the comparisons will fail to capture the ordinal cross-over of confusabilities that is observed as function of deadline. Similarly, in Experiment 1, scenes that shared six MIPs/zero MOPs were more confusable/similar than scenes that shared 5 MIPs/3 MOPs if subjects were given 2. 68 s to respond, but exactly the opposite pattern was found if subjects were given only 1 s to respond. Likewise, in Experiment 3, scenes with 2 MIPs/0 MOPs were more confusable than scenes with 0 MIPs/3 MOPs for the role-relevant, but not the role-irrelevant, group.

In all of these cases, the similarity of two things is not simply a relation between the two things; it is a relation between the two things and the process of comparison. The fact that similarity develops along a systematic time course suggests that similarity does not immediately impinge on our perceptual system. Instead, perceptual and cognitive processes actively build a conception of similarity. Once we dispatch with the assumption that similarity is out there in the world, then the question "How does similarity develop with processing?" becomes crucial. Our experiments indicate that similarity develops by placing, objects in alignment, and that these alignments become increasingly influenced by other object and role correspondences with time.

Footnotes

1

The second half of this definition is somewhat amended when roles are included in the scene description.

2

A nondefault value of 2. 0 was assigned to the parameter object to feature weight. This particular value is not important, as long as it is greater than 1. 0. If the parameter is given a value of 1. 0, no amount of activation cycling will allow nonmatching features to be placed in strong correspondence because of their objects' alignment.

3

The noise parameter is necessary to make SIAM stochastic. When shown the same display, a subject will sometimes say "same" and sometimes say "different. " Without N, SIAM would always give the same response to a particular display at a particular deadline. As N increases, SIAM's determinism decreases.

4

By using a single parameter to convert from subjects' deadlines to cycles, we are making the strong assumption of a ratio scale for number of cycles. Although this is probably too strong an assumption, model fits were not significantly better when two parameters, corresponding to an interval scale, were used to model deadlines: number of cycles for medium deadline and number of cycles added/subtracted to derive short/long deadlines.

5

Model parameters that maximize the likelihood of obtaining the observed data under the model are also obtained, and are arguably (K. Smith, personal communication; May 11, 1991) more appropriate for data with category-dependent and -independent variables. The best fitting maximum likelihood parameters closely approximate the least mean square estimates. We use the latter for the present modeling in order to have greater comparability with the other models developed.

6

It might appear to be more natural to dedicate parameters for initial-upper, initial-lower, slope-upper (the rate at which the upper bound increases), and slope-lower. However, the effect of any combination of values for these four parameters can be exactly duplicated by the three parameters L, M, and D, as long as slope-lower is greater than slope-upper. This is accomplished by setting M = (initial-upper + initial-lower)/2, D = (M - initial-lower), and L = (initial-upper - initial-lower)/(slope-lower - slope-upper). Thus, the parameters L, M, and D were chosen so as to be compatible with Experiment 1 and to minimize the number of free parameters.

References:

1. Bamber, D. (1969). Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception & Psychophysics*, 6, 169-174.

2. Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics*, 1, 300-302.

3. Busemeyer, J. R. & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, 32, 91-134.

4. Carroll, J. D. & Wish, M. (1974). Models and methods for three-way multidimensional scaling.In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.),*Contemporary developments in mathematical psychology*(Vol. 2, pp. 57-105). San Francisco: Freeman.

5. Corter, J. E. (1987). Similarity, confusability, and the density hypothesis. *Journal of Experimental Psychology: General*, 116, 238-249.

6. Dawson, M. R. W. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, 98, 569-603.

7. Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, 16, 519-555.

8. Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1, 245-252.

9. Falkenhainer, B., Forbus, K. D. & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.

10. Farell, B. (1985). "Same"-"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, 98, 419-456.

11. Gati, I. & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology*, 16, 341-370.

12. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

13. Gentner, D. (1989). The mechanisms of analogical learning.In S. Vosniadu & A. Ortony (Eds.),*Similarity and analogical reasoning*(pp. 199-241). New York: Cambridge University Press.

14. Gentner, D. & Ratterman, M. J. (1991). Language and the career of similarity.In S. A. Gelman & J. P. Byrnes (Eds.), Perspectives on thought and language: Interrelations in development (pp. 225-277). London: Cambridge University Press.

15. Getty, D. J., Swets, J. A., Swets, J. B. & Green, D. M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics*, 26, 1-19.

16. Gillund, G. & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 19, 1-65.

17. Gluck, M. A. & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247.

18. Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 3-28.

19. Goldstone, R. L., Gentner, D. & Medin, D. L. (1989). Relations relating relations.In*Proceedings of the 11th Annual Conference of the Cognitive Science Society*(pp. 131-138). Hillsdale, NJ: Erlbaum.

20. Goldstone, R. L. & Medin, D. L. (in press). Similarity, interactive activation, and mapping: An overview.In K. Holyoak & J. Barnden (Eds.),*Advances in connectionist and neural computation theory: Vol. 2. Connectionist approaches to analogy, metaphor, and case-based reasoning.* Norwood, NJ: Ablex.

21. Goldstone, R. L., Medin, D. L. & Gentner, D. (1991). Relations, attributes, and the non-independence of features in similarity judgments. *Cognitive Psychology*, 23, 222-264.

22. Gronlund, S. D. & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 846-858.

23. Hayes-Roth, B. & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior*, 16, 321-338.

24. Holyoak, K. J. & Thagard, P. (1989). Analogical mapping by constrain satisfaction. *Cognitive Science*, 13, 295-355.

25. Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 450-463.

26. Lopes, L. L. & Johnson, M. D. (1982). Juding similarity among strings described by hierarchical trees. *Acta Psychologica*, 51, 13-26.

27. Markman, A. B. & Gentner, D. (1993). Structural alignment during similarity comparison. *Cognitive Psychology*, 25, 431-467.

28. Marr, D. (1982). *Vision.* San Francisco: Freeman.

29. Marr, D. & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London*, 204, 301-328.

30. McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375-407.

31. Melara, R. D. & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 398-414.

32. Medin, D. L., Goldstone, R. L. & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, 1, 64-69.

33. Medin, D. L., Goldstone, R. L. & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.

34. Nickerson, R. S. (1972). Binary classification reaction time: A review of some studies of human information-processing capabilities. *Psychonomic Monograph Supplements*, 4((17), Whole No. 65).

35. Nissen, M. J. (1985). Accessing features and objects: Is location special?In M. I. Posner & O. F. Marin (Eds.),*Attention and Performance XI*(pp. 241-273). Hillsdale, NJ: Erlbaum.

36. Palmer, S. E. (1978). Structural aspects of visual similarity. *Memory & Cognition*, 6, 91-97.

37. Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, 72, 407-418.

38. Podgorny, P. & Garner, W. R. (1979). Reaction time as a measure of inter-intraobject visual similarity: Letters of the alphabet. *Perception & Psychophysics*, 26, 37-52.

39. Pomerantz, J. R. (1986). Overview of visual form perception.In E. C. Schwab & H. C. Nusbaum (Eds.),*Pattern of speech and visual form: Theoretical issue, models, and research.* (Vol. 2, pp. 1-30). San Diego, CA: Academic Press.

40. Proctor, R. W. & Healy, A. F. (1985). Order-relevant and order-irrelevant decision rules in multiletter matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 519-537.

41. Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, 88, 552-572.

42. Ratcliff, R. & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, 21, 139-155.

43. Sekuler, A. B. & Palmer, S. E. (1992). Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General*, 121, 95-111.

44. Sergent, J. & Takane, Y. (1987). Structures in two-choice reaction-time data. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 300-315.

45. Smith, L. B. & Heise, D. (1992). Perceptual similarity and conceptual structure.In B. Burns (Ed.),*Percepts, concepts, and categories: The representation and processing of information*(pp. 233-272). Amsterdam: Elsevier.

46. Torgerson, W. S. (1965). Multidimensional scaling of similarity. *Psychometrika*, 30, 379-393.

47. Treisman, A. M. (1987). Properties, parts, and objects.In K. Boff, L. Kaufman, & J. Thomas (Eds.),*Handbook of perception and performance*(pp. 159-198). New York: Wiley.

48. Treisman, A. & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 12-31.

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

50. Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

51. Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Received: July 7, 1992. Revised: February 17, 1993. Accepted: March 15, 1993.

Source: Journal of Experimental Psychology: Learning, Memory, And Cognition. Vol. 20 (1) January 1994, pp. 29-50
Accession Number: xlm20129 **Digital Object Identifier:** 10.1037//0278-7393.20.1.29

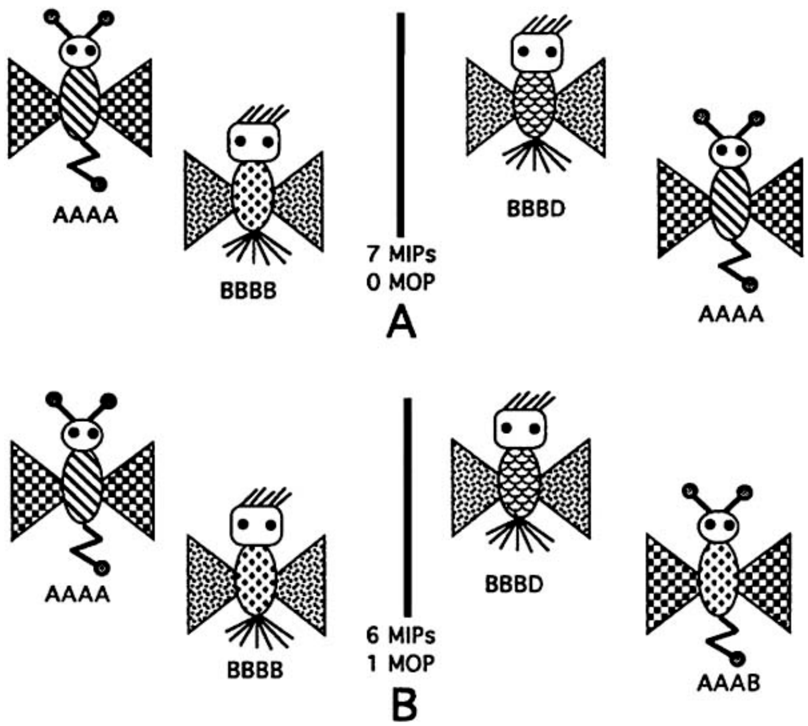
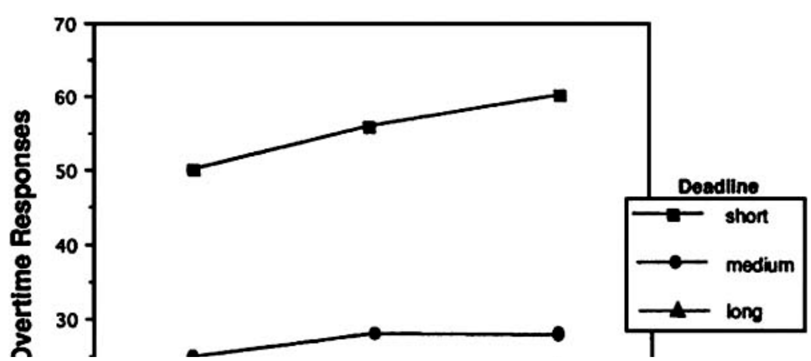
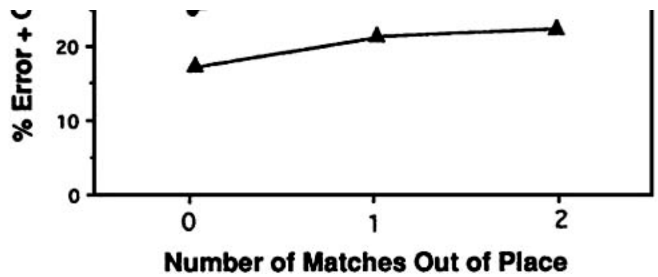


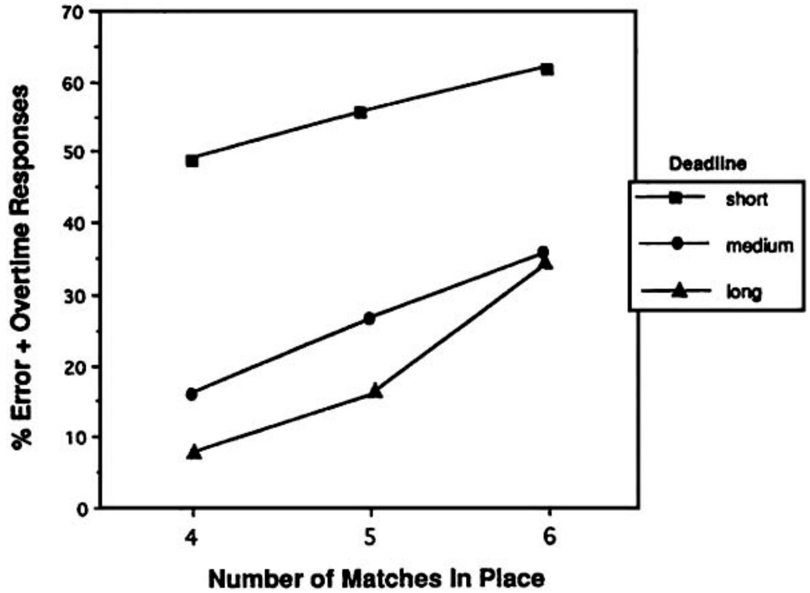
Table 1
Design, Results, and Simulations (Error and Overtime Rate) of Experiment 1

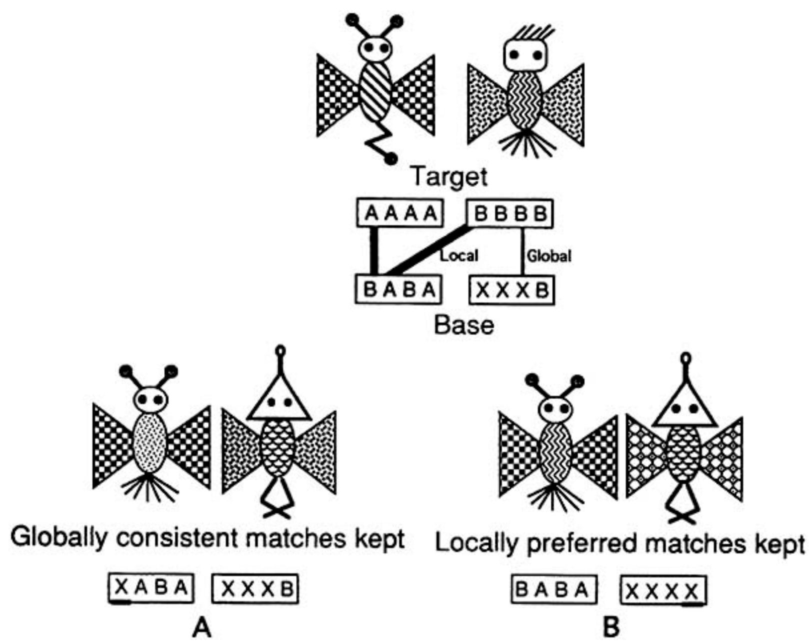
Method no. and changed scene[a]	No. of MOPs	No. of MIPs	Deadline: Short *O*	*M*	*S*	Medium *O*	*M*	*S*	Long *O*	*M*	*S*
1: AABB/BBAA	4	4	60	63	60	27	21	19	20	15	13
2: AABC/BBAA	3	4	62	55	56	20	20	21	16	14	18
3: AABC/BBAD	2	4	53	48	55	13	19	14	12	14	12
4: AABC/BBDD	1	4	52	41	46	20	18	21	9	13	11
5: AACC/BBDD	0	4	43	33	42	15	18	21	6	13	6
6: AAAB/BBAA	3	5	70	69	68	31	30	32	18	23	24
7: AAAB/BBDA	2	5	59	62	66	26	30	28	16	23	19
8: AAAB/BBDD	1	5	55	55	53	25	29	24	18	22	24
9: AAAC/BBDD	0	5	53	47	46	31	28	30	14	22	17
10: AAAB/BBBA	2	6	66	76	67	38	40	38	38	32	31
11: AAAB/BBBD	1	6	62	69	61	39	39	38	36	31	28
12: AAAC/BBBD	0	6	57	61	59	32	38	33	30	31	31
13: AAAA/BBBD	0	7	74	76	75	58	48	60	53	40	46

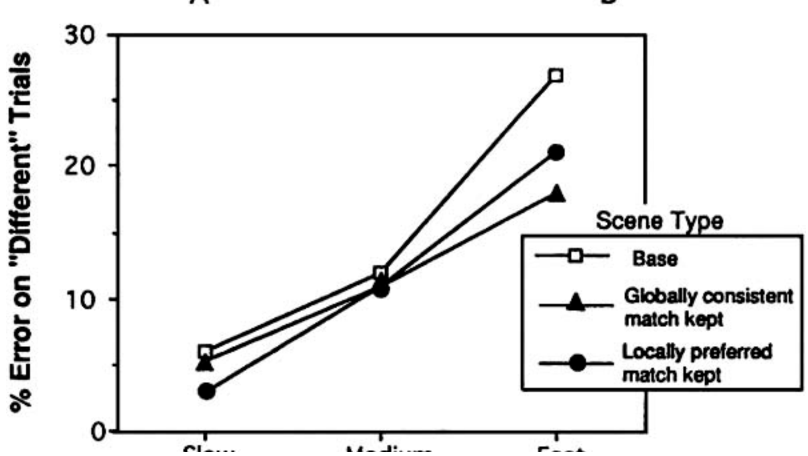
Note. MOP = match out of place; MIP = match in place; *O* = observed percentage of incorrect responses on "different" trials; *M* = predicted values from best fitting (seven parameter) linear regression; and *S* = predicted values from the presented model of similarity as interactive activation and mapping.
[a]Initial scene was AAAA/BBBB.




% Error +
0
10
20
0
1
2
Number of Matches Out of Place


% Error + Overtime Responses
0
10
20
30
40
50
60
70
4
5
6
Number of Matches In Place
Deadline
short
medium
long


Target
AAAA
BBBB
Local
Global
BABA
XXXB
Base
Globally consistent matches kept
XABA
XXXB
A
Locally preferred matches kept
BABA
XXXX
B


% Error on "Different" Trials
0
10
20
30
Scene Type
Base
Globally consistent match kept
Locally preferred match kept

Slow Medium Fast

Deadline

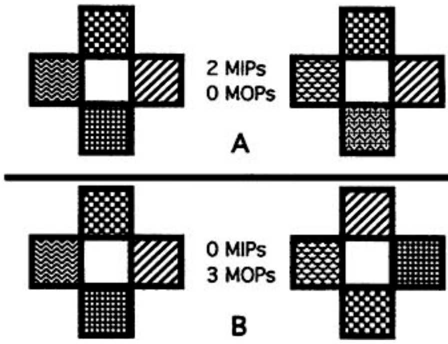
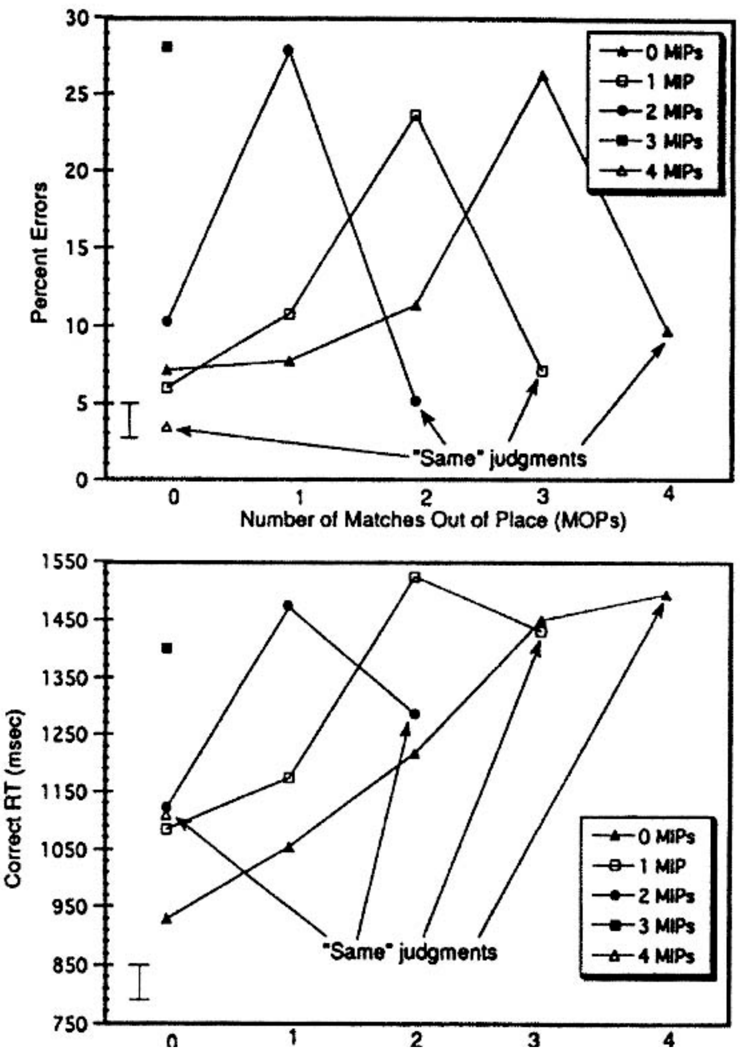


Table 2
Design of Experiment 3

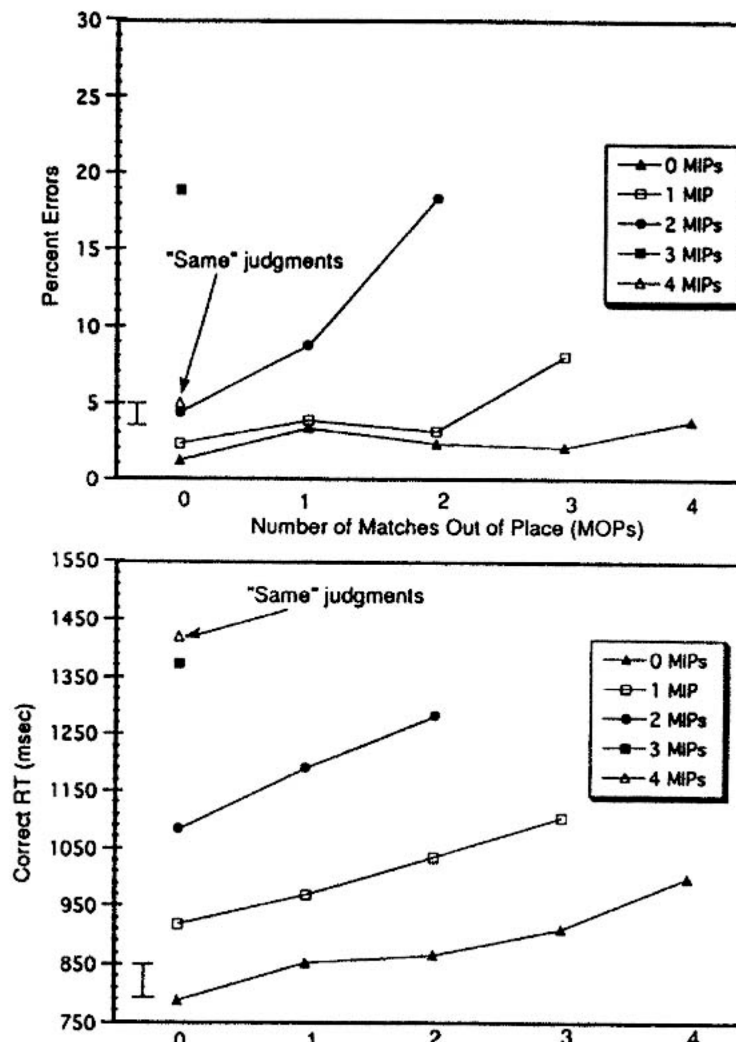
Method no. changed scene	No. of MOPs	No. of MIPs
1: BCDA	4	0
2: BACH	3	0
3: BCAD	3	1
4: BAGH	2	0
5: BCGD	2	1
6: BACD	2	2
7: EFDH	1	0
8: BFGD	1	1
9: ABDH	1	2
10: EFGH	0	0
11: EFCH	0	1
12: EFCD	0	2
13: ABCH	0	3

Note. Initial scene was ABCD. MOP = match out of place; MIP = match in place.

Role-irrelevant Group

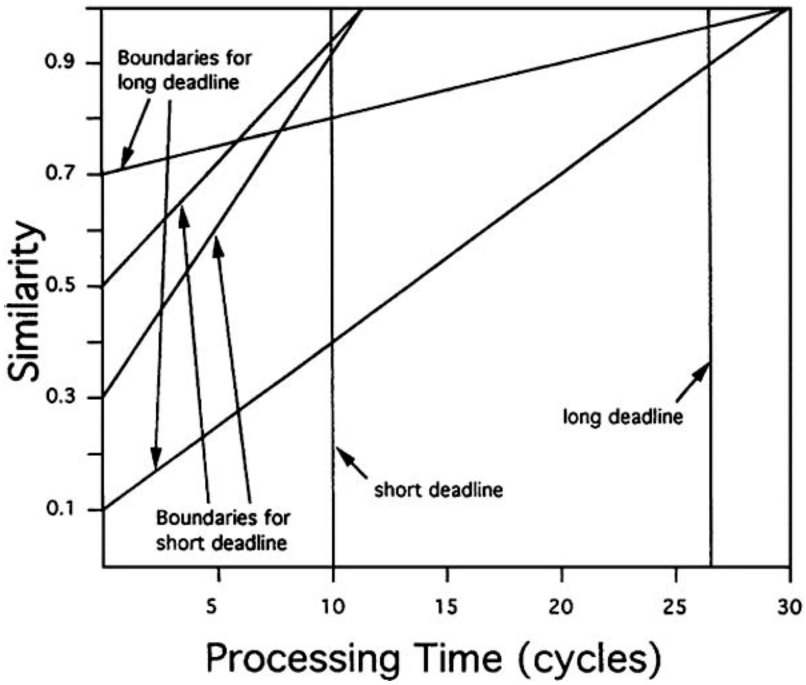


Role-relevant Group



Number of Matches Out of Place (MOPS)

Number of Matches Out of Place (MOPs)



$$\text{boundary}_C = \frac{(L - C)\,I + C}{L},$$

T

	Similarity Rating	Confusion %
A	4.64	0.119
B	4.60	0.524
C	5.48	0.190
D	4.32	0.286