ELSEVIER

Cognitive Science 28 (2004) 335–358

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Varieties of sameness: the impact of relational complexity on perceptual comparisons☆

James K. Kroger[a,*], Keith J. Holyoak[b], John E. Hummel[b]

[a]*Department of Psychology, New Mexico State University, MSC 3452, Science Hall, Room 220, Sweet/Stewart St., Las Cruces, NM 88003-8006, USA*

[b]*Department of Psychology, University of California at Los Angeles, LA 951563, CA 90095-1563, USA*

Received 2 January 2003; received in revised form 28 June 2003; accepted 30 June 2003

Available online 5 March 2004

Abstract

The fundamental relations that underlie cognitive comparisons—“same” and “different”—can be defined at multiple levels of abstraction, which vary in relational complexity. We compared response times to decide whether or not two sequentially-presented patterns, each composed of two pairs of colored squares, were the same at three levels of abstraction: perceptual, relational, and system (higher order relations). For both 150 ms and 5 s inter-stimulus intervals (ISIs), both with and without a masking stimulus, decision time increased with level of abstraction. Sameness at lower complexity levels contributed to decisions based on the higher levels. The pattern of comparison times across levels was not predictable solely from encoding times. The results indicated that relations at multiple levels of complexity can be abstracted and compared in working memory, with higher complexity levels requiring more processing time. We simulated the impact of relational complexity on response time using Learning and Inference with Schemas and Analogies (LISA), a computational model of relational comparisons based on dynamic binding of elements into roles in a relational working memory.

Keywords: Reasoning; Working memory; Executive control; Perception

1. Introduction

Perhaps the most fundamental psychological relation is *sameness*. The capacity to recognize that two objects, situations, or events are the same with respect to a certain criterion underlies

☆ Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cogsci.2003.06.003.

* Corresponding author. Tel.: +1-505-646-2243; fax: +1-505-646-6212.
E-mail address: jkroger@nmsu.edu (J.K. Kroger).

0364-0213/$ – see front matter
doi:10.1016/j.cogsci.2003.06.003

object recognition, categorization, and analogical reasoning. It is apparent that the concept of sameness can vary enormously in abstraction. At an implicit level, all vertebrates (and many invertebrates as well) can treat distinct but similar entities as the same (e.g., as prey). More complex comparisons can be made by explicitly identifying dimensions of variation, making it possible to treat objects as the same based on multiple criteria. A monkey, for example, can select which object is the same as another on the basis of shape (ignoring location) or on the basis of location (ignoring shape). An adult human is capable of far more abstract comparisons, such as recognizing that West Side Story has the "same" relational structure as Romeo and Juliet, despite the considerable surface differences. It has been suggested that human capacity to recognize sameness at abstract levels is closely linked to the evolution of prefrontal cortex (Holyoak & Kroger, 1995). Increasing ability to detect sameness of entities based on similar relationships among their components occurs as species's cortical development becomes more advanced (e.g., Holyoak & Thagard, 1995; Premack, 1983), and as humans mature from childhood into adulthood (e.g., Gentner & Rattermann, 1991; Piaget, Montangero, & Billeter, 1977; Smith, 1989).

1.1. Levels of relational complexity

The levels of abstraction at which sameness can exist can be defined in terms of *relational complexity*. A number of theorists have proposed variants of a three-level taxonomy of relational complexity based on the predicate–argument structure of propositions (Gentner, 1983; Halford, 1993; Halford & Wilson, 1980; Holyoak & Thagard, 1995; Premack, 1983). Gentner distinguished sameness at the levels of attributes (one-place predicates), first-order relations (multi-place predicates with objects as arguments), and higher order relations (multi-place predicates with at least one argument in the form of a proposition). Halford (1993) defined complexity in terms of the number of role-filler bindings that must be maintained simultaneously in order to represent an idea or solve a problem, and linked relational complexity to working-memory capacity. Premack (1983) proposed that a more abstract level of sameness distinguishes the reasoning ability of chimpanzees trained in manipulation of symbolic tokens from that of their untrained conspecifics. Holyoak and Thagard (1995) extended Premack's analysis to account for the further gap that separates the reasoning ability of symbol-trained chimpanzees from that of humans.

Consider a basic comparison task such as match-to-sample. If an apple is presented as the sample, a monkey can be trained to select an apple (rather than, say, a hammer) as the match; if the sample is then varied, the animal will continue to select the alternative that is the same shape as the sample. This task can be performed by making a *perceptual* match between two objects. Monkeys (macaques and capuchins) can learn versions of the match-to-sample rule based on identity of object shape (D'Amato, Salmon, Loukas, & Tomie, 1986), color (Fujita, 1982) and sounds (Wright, Shyan, & Jitsumori, 1990), and then apply it to novel stimuli within the same modality (but generally not across modalities, as would be required for transfer from identity of shape to identity of color; D'Amato, Salmon, & Colombo, 1985).

Chimpanzees are yet more facile in learning and generalizing the identity relation (Nissen, Blum, & Blum, 1948; Oden, Thompson, & Premack, 1988a, 1988b, 2001; Thompson, Oden, & Boysen, 1997). Premack (1983) described a pairwise version of the match-to-sample task, in which the sample is a pair of objects (e.g., apple–apple), and the alternatives are also pairs (e.g., hammer–hammer vs. shoe–flower). Symbol-trained chimpanzees are able to choose the "same" alternative in the pairwise task. This task is more complex because the match must be made on a relation, rather than directly between physical objects. That is, it is necessary to code apple–apple as "same objects," or *O-same*, and to recognize that hammer–hammer is also O-same, whereas shoe–flower is O-different. The task thus requires a *relational* match.

Holyoak and Thagard (1995) observed that the match-to-sample task can be further generalized to items based on *pairs* of pairs, and that this further increment in relational complexity is required to represent sameness of relations so that analogical matches can be recognized. For example, the relation between apple–apple (O-same) and that between hammer–hammer (O-same) are themselves the same relation, or *R-same* (a higher order relation between relations). Recognizing this abstract sameness would allow a match to shoe–flower (O-different) and bottle–bell (O-different), as the relation between the latter relations is again R-same, even though there is no overlap either of objects or first-order relations. This deeper level of sameness requires a *system* match, and there is no evidence that any species other than humans in capable of this type of abstraction.

1.2. Relational complexity and processing time

The present study introduces a speeded matching task related to the variants of match-to-sample described above. We compared the time required to make matches between visual displays at either the perceptual, relational, or system level, while holding the physical characteristics of the displays as constant as possible. We hypothesized that higher complexity levels would require more processing time. By varying sameness at lower complexity levels, we also sought to determine whether each type of match is made independently, or whether multiple levels of representation can cooperatively contribute to a decision. We examined the impact of relational complexity on decision time at both short (Experiments 1 and 3) and longer (Experiments 2 and 4) inter-stimulus intervals (ISIs).

Most models of sameness or similarity judgments, whether based on discrete features (Tversky, 1977) or geometric representations (Shepard, 1962), have focused on comparisons between static representations of the situations being compared, and have not differentiated alternative representational levels that may emerge over the course of processing. In contrast, a variety of computational models focusing on analogical reasoning have emphasized inter-dependencies between processing at multiple levels of abstraction (e.g., Falkenhainer, Forbus, & Gentner, 1989; Hofstadter & Mitchell, 1994; Holyoak & Thagard, 1989; Keane & Brayshaw, 1988). At a more specific level, some models derived from work on analogical reasoning can yield predictions about the time course of comparisons. In particular, Goldstone's SIAM model (1994; Goldstone & Medin, 1994a) uses an interactive-activation algorithm (similar to that of the ACME model of analogy; Holyoak & Thagard, 1989) to account for qualitative shifts in the basis for perceptual similarity judgments over time. In SIAM, processing proceeds in a local-to-global fashion, with similarity first being driven by direct feature overlap, and later increasingly influenced by systematic relational correspondences. Empirical studies of speeded similarity judgments have yielded results consistent with SIAM's predictions (Goldstone & Medin, 1994b).

Although SIAM provides a good fit to data from experiments in which relations have been shown to influence similarity judgments, there are reasons to question its ultimate adequacy as an algorithmic model. In particular, like Holyoak and Thagard's (1989) ACME model of analogy, SIAM is not constrained to operate within realistic working-memory limits. In response to this and other limitations of previous models of analogical reasoning, Hummel and Holyoak (1997, 2003; Holyoak & Hummel, 2000a, 2000b) developed the Learning and Inference with Schemas and Analogies (LISA) model, which performs analogical mapping (and other subprocesses of analogical reasoning) using distributed representations of concepts that are dynamically bound into relational structures by patterns of temporal synchrony. In a supplementary archive (http://cognitivesciencesociety.org/supplements/), we present an extension of the LISA model that can account for sameness judgments at multiple levels of abstraction, and in particular the reported effects of relational complexity on human response times making these judgments.

The experiments we report here involve speeded judgments of sameness between simple perceptual patterns. In contrast to the similar paradigm investigated by Goldstone and Medin (1994b), we explicitly required participants to assess sameness at various distinct levels of complexity. The four-square task, which is a direct extension of simpler match-to-sample procedures used in the primate literature, requires participants to make sameness judgments about stimuli consisting of four colored squares arranged together into a larger square. The participant saw one four-square stimulus, which was then replaced by another one; each square in the second stimulus might or might not be the same color as the corresponding square in the first figure. The complexity of the judgment to be made varied across conditions. An example of a four-square trial is shown in Fig. 1.

Primary considerations in designing this task were that the participant should make sameness judgments at varying levels of relational complexity while the stimuli themselves remained identical across conditions; this control would eliminate performance differences attributable to intrinsic perceptual qualities of the stimuli. We also sought to prevent contamination from intrinsic differences in how different conceptual relations might be processed; e.g., the relation *under* (x, y) might take longer to process than *darker* (x, y). The task used here requires participants to make yes–no judgments about only the conceptual relation "sameness" (or its negation, "difference").

The four-square task involves three levels of complexity. In the lowest or attribute-complexity level, the participant simply compares the color of each square in the probe to the color in the sample. If any square has changed colors, the participant indicates there has been a change. In the relation-level condition, the participant determines whether the top squares in the sample are the same color ("O-same," for sameness of objects) or different colors (O-different), and does the same for the bottom pair. If either of these pair relationships is not the same as it was in the sample, the participant must indicate there has been a change. In the system-level condition, participants must compare the relation between the top two squares (O-same or O-different) to the relation between the bottom two squares, deriving an overall relation between relations (R-same or R-different). This overall relation for the probe is compared to that for the sample. Fig. 2 illustrates the nature of judgments made for a trial at each level of relational complexity.

Importantly, the number of "sameness" judgments necessary to complete a trial is constant (at four) across all of the complexity conditions. This control is made possible by the fact that each

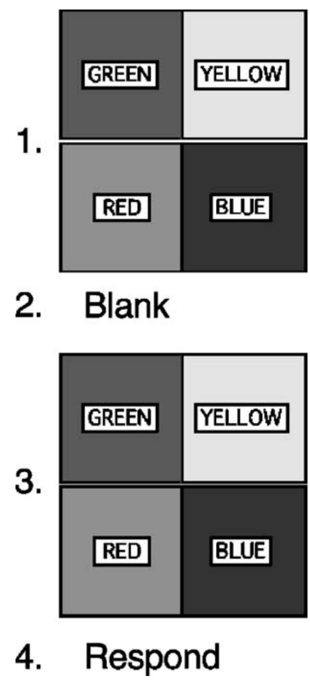

Fig. 1. An example of a four-square trial. The participant first sees one figure which serves as the sample, then a blank screen, then a second figure which is the probe. They are to judge if the probe is the same as the sample or if a change has occurred. The type of change judgment the participant makes is different for the three complexity levels. The probe remains onscreen until the participant responds. The probe figure then serves as the sample for the next trial. It is followed by a blank screen and then another probe figure.

stimulus serves both as the probe in one trial and as the sample in the following trial. For each trial, the number of comparisons needed are those required for the new stimulus (probe) plus comparisons between the sample and the probe, the relations of which have been determined in the previous trial (see Fig. 2). It may be seen from Fig. 2 that the number of comparisons *between* the sample and probe decreases with increasing relational complexity (from four in the attribute-level condition to two in the relation-level condition and one in the system-level condition); conversely, the number of comparisons *within* the probe increases across the three complexity levels (0, 2, and 3, respectively). The total number of comparisons to complete a trial (after presentation of the probe) is therefore constant at four across conditions, and the judgment made is constant ("no change" or "change"). Assuming the time required to make each subsidiary judgment is roughly constant, then any difference in RT between conditions can be attributed to the increased processing required to mediate representations formed by participants at different levels of relational complexity. Indeed, it might be hypothesized that cross-pattern comparisons, which necessarily depend on a working-memory representation of the sample, are more cognitively demanding than within-probe comparisons, which can be made based solely on the visible probe pattern. Since the number of cross-pattern comparisons declines with increasing complexity, this factor runs counter to the predicted increase in decision time with complexity level.

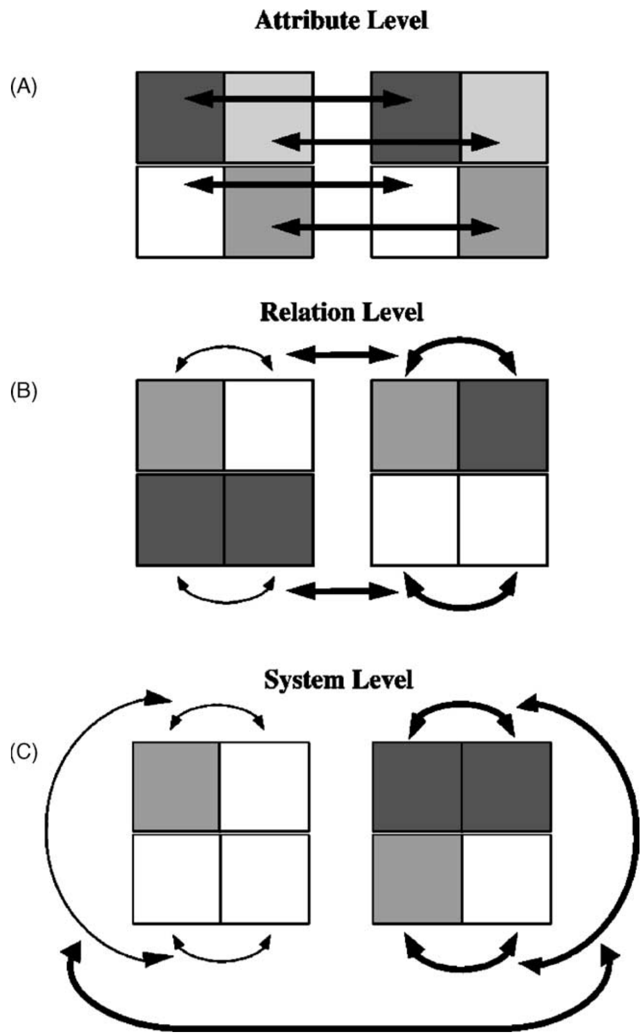

Fig. 2. Examples of trials at each level of relational complexity. Subsidiary comparisons logically necessary to calculate the answer for each trial are indicated in thick bold; comparisons completed on previous trial (when current sample served as probe) are indicated in thinner bold. In each case the correct answer is "no change."

2. Experiment 1

2.1. Method

2.1.1. Participants

Twenty-eight undergraduate students in an introductory psychology class at the University of California, Los Angeles (UCLA) served in the experiment as part of a class requirement.

2.1.2. Materials

The experiment was controlled by an Apple Macintosh computer with a color monitor. Participants sat enclosed in a 4 ft by 4 ft cubicle containing the computer on a desk and one chair. A script written in the MacProbe programming language presented all instructions and stimuli and recorded responses and response times (RTs) for each trial. Stimuli consisted of a group of four colored squares (blue, green, red, or gray) arranged as shown in Fig. 1, presented against a black background. The actual overall figure was 1.6 in. high and 1.5 in. wide on the computer screen. Participants sat approximately 30 in. from the screen, with their index fingers resting on the keyboard keys representing the "change" and "no-change" answers.

2.1.3. Design and procedure

Participants were greeted by the experimenter and asked to sit in the chair in the experiment cubicle. The experimenter explained to the participant that all of the instructions for the experiment would be presented on the screen of the computer in front of them, and told them to press any key to begin when they were ready. The experimenter then closed the door to the cubicle and left the participant alone.

After some initial demographic questions had been answered, the computer script began animated instructions for the first block of trials, which were in the attribute-level condition. Text on the screen accompanied an example of a trial, explaining how to compare each figure to the previous one for that condition. Participants were instructed to indicate for each figure whether it demonstrated the "designated change" relative to the previous figure (except that no response was made for the first figure). In the attribute-level condition, the designated change was defined as any change in the color of one or more squares from the previous figure to the current figure (e.g., the top left square changing from red to green). Participants were instructed to press the "n" key (no change) if all four squares were the same color as they had been in the previous figure; otherwise they were to press the "c" key (change).

Each stimulus remained on the screen until the participant pressed a response key. Two seconds after the response, the figure was replaced by a black screen for 150 ms, after which the next stimulus appeared. One trial consisted of a presented figure and the participant's response, indicating whether or not each figure was changed in the designated manner from the previous figure. The stimuli were presented in a continuous fashion, so that each figure (after the first) was first compared to the preceding figure on one trial (i.e., was the probe), and then served as the comparison figure (i.e., the sample) for the next trial.

After the explanatory instructions were finished, three examples of trials at the complexity level of the current block were presented with correct answers indicated. Finally, the participant was instructed to practice until they answered five successive trials correctly. A series of practice trials followed, with either the word "correct" or the words "incorrect—start over" displayed following each trial. When five successive trials had been correctly completed, the legend "5 successes" appeared on the screen. The participant was then instructed that real trials would begin, and a block of 20 trials was presented.

Following completion of the 20 attribute-level trials, instructions, practice, and a trial block were presented for each of the two remaining level-of-complexity conditions. In the relation-level condition, a "change" was defined in terms of the relations between the colors of the squares within each of the two pairs (top and bottom). For example, if both squares in a pair were red, then a relational change would require that the corresponding pair in the next figure be of different colors (e.g., green–blue would define a change, whereas green–green would not). Similarly, if an initial pair were different (e.g., red–green), then green–green would count as a change, but green–blue would not. If either or both the top or bottom pair relationship had changed, participants were to press the "c" key.

In the system-level condition, a change was defined in terms of the higher order relation between the sameness relations for top and bottom pairs. For example, if the initial figure consisted of red–red above green–blue, then the higher order relation between these two relations would be different. Accordingly, if the next figure presented consisted of red–red above green–green, forming a higher order relation of same, this would count as a change. In contrast, if the next figure were red–gray above green–green, the higher order relation would remain different (i.e., no change).

Six more blocks of trials followed, each preceded by a short paragraph indicating the designated change for that block. The nine 20-trial blocks (three for each level-of-complexity condition) formed a Latin square design that controlled for order effects. Successive trial blocks were separated by a rest period of approximately 15 s during which the words "You may now rest" were displayed. After 10 s of rest, three beeps were heard and the words "Get ready" appeared.

An equal number of no-change and change trials were randomly ordered within each level-of-complexity block. In the attribute-level condition no-change trials, all four squares were the same color as in the previous figure. In the change trials, the number of squares changing colors was equally distributed between one and four. Each square changing color was selected randomly among the four squares. When a square changed colors from one trial to the next, the new color was randomly chosen among the other three colors.

In the relation-level condition change trials either the top-pair relation, the bottom-pair relation, or both pair relations changed. These changes were randomly distributed, and were accomplished by changing the color of from one to four squares. When a pair's relation changed, either one or two squares changed colors according to random selection. When a pair's relationship did not change, half of the time one or both squares changed color; when the relation was "same" both squares changed or did not change. Squares that changed colors were randomly selected within the constraints of random pair selection.

Square colors in the system-level condition were determined in a similar manner. Since judgments at the relational and system levels of complexity did not depend directly on whether squares were the same color as in the previous figure, it would have been possible to change the colors of all the squares on any given trial. However, we wished to induce participants to perform the desired comparisons at greater levels of complexity while keeping the number and patterns of actual color changes among the four squares generally consistent with events in the attribute trials. This control was intended to block alternate strategies that participants might have developed on the basis of number or pattern of squares changing colors.

2.2. Results and discussion

All analyses were performed on mean correct RTs averaged across the three trial blocks in each condition. As depicted in Fig. 3, mean correct RTs increased monotonically with complexity level (collapsing over change and no-change trials, means of 1,095, 1,340 and 1,727 ms, respectively, for the attribute, relational, and system conditions), $F(2, 74) = 63.86$, MSE = 120,556, $p < .0001$. Analyses of accuracy for the complexity levels (91, 88, and 86%, respectively) showed a monotonic decrease, $F(2, 37) = 9.996$, MSE = 40.831, $p < .0001$, indicating that the RT pattern cannot be attributed to a speed-accuracy trade-off. Overall, decisions were faster for no-change than for change trials (1,282 ms vs. 1,493 ms), $F(1, 37) = 48.70$, MSE = 52,034, $p < .0001$. Trial type interacted with level of complexity, $F(2, 74) = 8.83$, MSE = 20,720, $p < .001$, reflecting a smaller difference between no-change and change trials for the attribute condition.

The increase in decision time as complexity level increased supports the hypothesis that manipulating representations at higher levels of relational complexity places progressively greater demands on working memory. This pattern is actually opposite to that predicted by the minimal number of cross-pattern comparisons required at each level. Recall that at the attribute level a "no change" (match) decision requires four such comparisons (one for each individual square); at the relational level a "no change" decision requires two such comparisons (one for each pair, based on the relations O-same or O-different); and at the system level a "no change" decision requires one comparison (for the overall higher order relation between pairs, R-same or R-different). Our results indicate that the effect on response times of the number of cross-pattern comparisons, which depend on a working-memory representation of the sample, are small in contrast to the impact of relational complexity on decision times. It might be argued that comparing attributes (e.g., red and blue) is intrinsically easier than comparing more abstract first-order or higher order relations (e.g., O-same and O-different at the relation level, or R-same and R-different at the system level). However, the increase in comparison times across the three levels is roughly linear (see Fig. 3), and thus

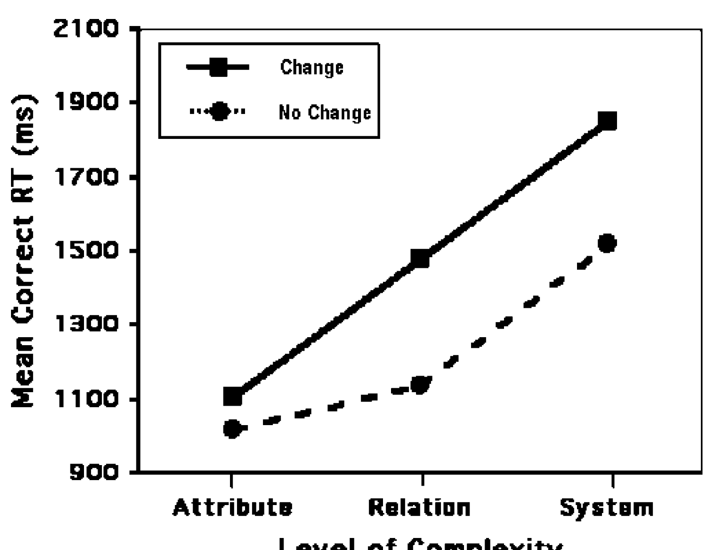

Fig. 3. Mean correct RTs for the three levels of complexity in Experiment 1. Means for no-change trials are shown separately from means for change trials.

does not suggest any major discontinuity between the attribute level and the more abstract levels.

Further analyses were performed to determine whether decisions at each complexity level were made independently of one another, or whether decisions were based on cooperative processing across multiple complexity levels, as suggested by both the SIAM (Goldstone, 1994) and LISA (Hummel & Holyoak, 1997) models. One possibility is that decisions at each level are independent of one another. At each level, decisions could be based solely on the most abstract relevant code, which in effect summarizes all lower level information. For example, a system decision could be made by deriving the higher order relation for the sample, doing the same for the probe, and then comparing the two higher order relations. Alternatively, decisions at the higher levels may be inter-dependent with those at lower levels, in the sense that information from multiple levels contributes to a decision at the nominal level defining the decision. Since the participant must process the stimuli at lower levels to form the higher order representations, rather than responding solely on the basis of the most abstract relevant code, a decision could be based on a mapping between full representations at each level. In other words, information about the stimuli at the designated level and all lower levels might be available for processing, and be utilized by participants in determining a response. For example, in an inter-dependent processing model, a system decision could be made by mapping multiple representational levels, with each stimulus being encoded not only by a single higher order relation, but also by the first-order relations for each pair, and the color attributes of each individual square. A stepwise procedure may be used to construct a representation at the necessary level of relational complexity, with lower levels being fully evaluated before the subsequent level is constructed. Nonetheless, once representations at the highest level have been formed, comparisons may operate on multiple levels simultaneously, creating inter-dependence among levels.

If in fact information at lower complexity levels is being utilized in higher complexity conditions, we should be able to detect an effect on RT when the lower level information is in agreement with the higher level answer. For example, if at the system level the sample and probe are the “same” and the relation-level match is also “same,” then RT should be shorter than if the relation-level match is “different.” Thus, the single-level and inter-dependent processing models can be distinguished by examining whether or not lower level matches have any effect on the time to assess higher level matches. Such cross-level effects would support an inter-dependent model in which cooperative processing occurs. We selected a subset of trials from the two higher complexity conditions for closer inspection. Only no-change trials were included. (Because no trial could be a change trial at a higher level and a no-change trial at the attribute level, the desired comparisons were not available for change trials.) The trial types can be described by a three-letter code signifying whether the figures match (S, for “same,” equivalent to no-change) or mismatch (D, for “different,” equivalent to change) at the attribute, relational and system levels. For example, SSS trials yield a match at all three levels; DDS trials yield a mismatch at the attribute and relational levels, but a match at the system level. Fig. 4 illustrates these types. Note that if all four squares match across the sample and probe at the attribute level (SSS), then the figures are necessarily the same at both the relational and the system level. Similarly, if both pairs match at the relational level, then the figures must match at the system level. (Neither converse holds.)

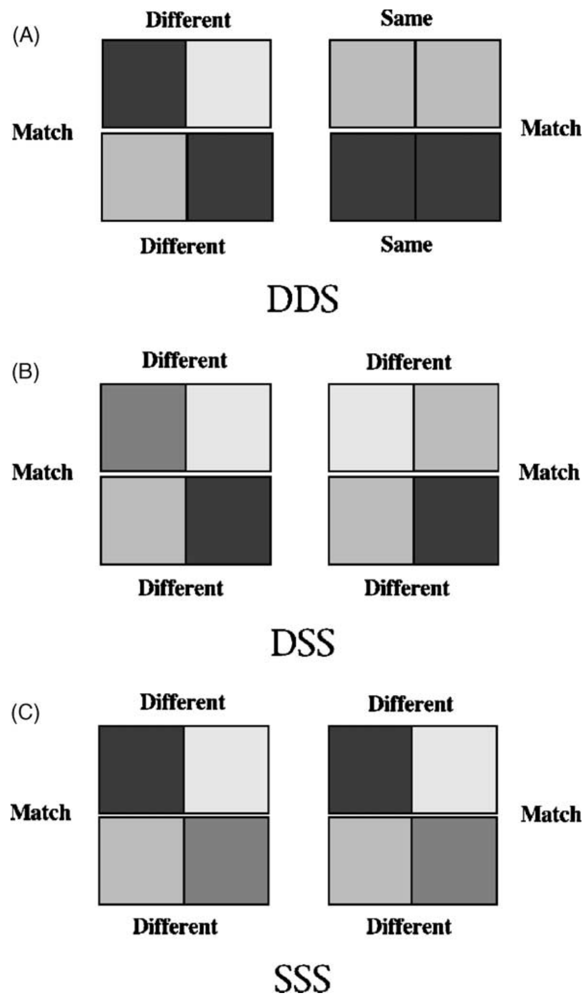

Fig. 4. Analyzing cooperative processing. Relation-level and system-level no-change trials for which the answer at the attribute and relation levels of complexity would be (A) change, change (DDS), (B) no change, change (DSS), and (C) no change, no change (SSS).

For each of the two higher complexity levels, we examined RTs for types of trials that varied in the degree of support that was provided by lower levels for the correct decision. Fig. 5 presents mean correct RTs for the various item types.[1] The results provide clear support for inter-dependent processing. At the relational level, RT was lower for SSS than for DSS trials (1,088 ms vs. 1,726 ms), $F(1, 31) = 38.80$, MSE = 167,430, $p < .0001$. In SSS trials, but not DSS trials, a lower level attribute match between probe and sample supports the required relational match. Similarly, at the system level RT increased across SSS, DSS, and DDS trials (1,207, 1,961 and 2,371 ms, respectively), $F(2, 70) = 71.36$, MSE = 175,928, $p < .0001$. Lower level matches between probe and sample led to faster matches at the most abstract level; and when both lower levels of complexity supported a match at the system level, decisions were faster than when only the relational level provided support. It appears that multiple levels

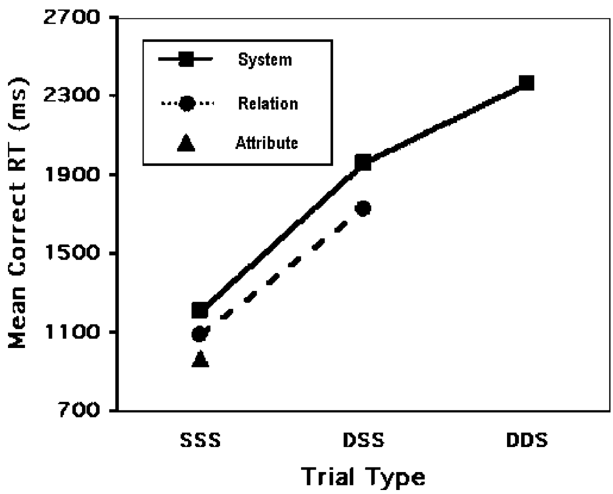

Fig. 5. Mean correct RTs for no-change trials as a function of criterial complexity level (attribute, relation, system) and trial type (SSS, DSS, DDS).

of representation operate cooperatively; decision efficiency at the system level increases with the number of lower levels that support the higher level decision.

A more serial processing model might be proposed as an alternative to cooperative processing. Perhaps participants became aware of the conditional implications relating matches at lower levels to matches at higher levels, and adopted a strategy of always serially checking for matches from the least to most complex level, responding at the first level that yielded a firm decision. This strategy could be applied for all three types of designated changes. For example, if a match is found at the attribute level (SSS trials), then a “match” response would be made immediately, regardless of the nominal complexity level. This strategy, rather than cooperative processing at multiple levels, could explain the reduction in RTs observed when a higher level match was supported by lower levels.

It is important to distinguish this specific self-terminating strategy from other strategies involving serial processing of properties and relations. There are logical reasons to believe that some form of (at least partially) serial strategy plays a crucial role in this task. Presumably observers process colors prior to processing higher order relations. Thus, at least a minimal amount of serial, “bottom-to-top” processing seems to be intrinsic to the task. The self-terminating account is more specific in that, in addition to serial “bottom-to-top” processing, it also holds that subjects always stop processing at the lowest level at which they find a match. The serial self-terminating strategy therefore predicts that RTs will depend solely on the lowest complexity level that generates a decision. If participants checked for a change first at the attribute level, then the relational level, and then the system level of complexity, they would never need to process any problem at a higher level than necessary to determine an answer, no matter in which complexity condition trials occurred. In particular, if they answered trials in all conditions based on the least complex representation that would provide an answer, RTs for “no-change” responses on SSS trials should be equal across complexity conditions.

We therefore compared means for SSS trials performed in the attribute, relation, and system conditions. This comparison involves the three trial types that appear on the left in Fig. 5. As the pattern depicted in Fig. 5 suggests, RTs for SSS trials increased monotonically with complexity level of the criterion (986, 1,074 and 1,194 ms, respectively), $F(2, 72) = 11.83$, MSE = 33,841, $p < .0001$. Thus, "no-change" decisions for perceptually identical stimuli varied with complexity level, even though it would have been possible in principle to always generate a correct response at the attribute level. The combination of variations in RTs for different trials types within complexity levels and variations in RTs for the same trial type across complexity levels jointly support a model based on cooperative processing across complexity levels.

3. Experiment 2

It might be argued that the monotonic increase in RTs across complexity levels observed in Experiment 1 could be partly attributable to some sort of automatic perceptual "pop-out" that facilitated detection of changes at the attribute level. The brief ISI (150 ms) might have allowed iconic memory for the sample pattern to be retained long enough to provide a basis for automatic change detection when the probe stimulus appeared. As our focus is on the influence of complexity level on deliberate, attentional processing, automatic change detection must be ruled out as an explanation for the faster response times obtained in Experiment 1 for the less complex trials.

3.1. Method

Experiment 2 replicated Experiment 1 exactly except for a single modification: the ISI was increased to 5 s, ensuring that attribute-level decisions would have to rely on representations being maintained in working memory, rather than direct visual traces. Thus, comparisons could only be accomplished by deliberately comparing the probe stimulus to a maintained representation of the sample. A 5-s delay is comparable to that typically used in match-to-sample studies that have implicated prefrontal working memory (e.g., Fuster, 1997).

Twenty-five UCLA undergraduates participated in the experiment as part of the requirements for an introductory psychology course. Instruction and administration of trial blocks occurred just as it did in Experiment 1.

3.2. Results and discussion

The pattern of results observed in Experiment 1 was fully replicated in Experiment 2. Fig. 6 (comparable to Fig. 3 for Experiment 1) depicts the monotonic increase in RTs that was observed as complexity level increased (1,577, 1,830, and 2,054 ms for the attribute, relational, and system conditions, respectively, collapsed across change and no-change trials), $F(2, 48) = 18.40$, MSE = 154,798, $p < .0001$. Mean accuracy in the three conditions was 88, 83, and 83%, respectively. As in Experiment 1, accuracy decreased significantly across complexity levels, $F(2, 48) = 9.06$, MSE = 47.794, $p < .0005$. No-change trials were significantly faster than change trials, $F(1, 24) = 27.69$, MSE = 127,183, $p < .0001$; this

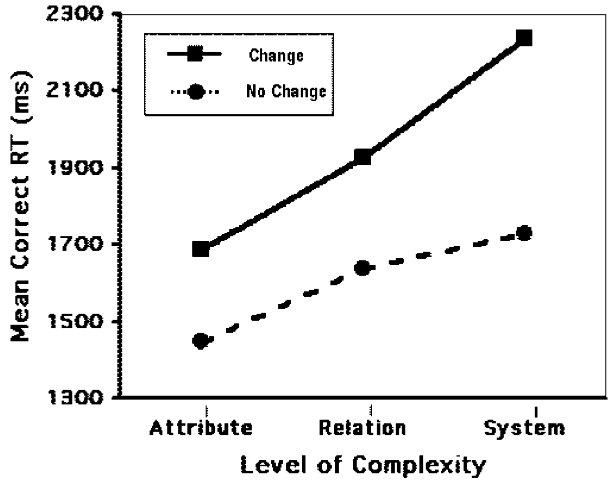

Fig. 6. Mean correct RTs for the three levels of complexity in Experiment 2. Means for no-change trials are shown separately from means for change trials.

factor interacted with complexity, $F(2, 48) = 3.55$, MSE = 44,903, $p < .05$, reflecting a smaller no-change advantage at the attribute level. The close replication of Experiment 1 supports the robustness of our findings. The use of a 5-s delay in Experiment 2 indicates that the relatively fast decisions in the attribute conditions are not attributable to automatic visual processes, but rather to deliberative processes applied to working-memory representations of low complexity.

It is worth noting that RTs across the conditions were approximately one-half second slower than those in Experiment 1 across all conditions, and error rates were slightly higher in the delay condition. The slower RTs in the delay experiment were most pronounced for more complex conditions. It appears participants were able to retain the most complex representations across the delay period but that maintaining the sample representation while forming and comparing the probe representation was more demanding after a 5-s delay than a 150-ms delay. This extra difficulty is possibly due to greater need for rehearsal after a longer delay.

Experiment 2 provided another opportunity to test for cooperative processing at multiple levels of complexity. As in Experiment 1, we compared system-level trials in which the probe and sample were the same at all levels of complexity (SSS), were the same at the highest two levels but different at the attribute level (DSS), or were the same only when considered at the system-level of complexity (DDS). We also compared relation-level no-change trials in which there was a change at the attribute level (DSS) with those in which there was not (SSS). Mean RTs in the relation-level condition for SSS and DSS trials were 1,631 and 2,165, respectively; system-level means for SSS, DSS, and DDS trials were 1,810, 1,929, and 2,455, respectively. RTs decreased as a function of the number of lower levels at which the answer was in agreement with the nominally correct answer for both relation-level trials, $F(1, 22) = 13.82$, MSE = 237,368, $p < .0015$, and system-level trials, $F(2, 28) = 13.21$, MSE = 133,722, $p < .0001$. As in Experiment 1, cooperative processing at multiple levels of complexity is supported by these results.

To ensure that the decrease in response time in trials where the answer at lower levels was in agreement with the nominal answer was not solely due to use of a "shortcut" strategy, in which participants simply responded in accord with the least complex level that provided an answer, trials in which the probe and sample did not differ at any level of relational complexity (SSS trials) in attribute, relation, and system conditions were compared, as in Experiment 1. Correct SSS mean RTs for the three conditions were 1,421, 1,597, and 1,709, respectively. As in Experiment 1, there was a monotonic increase in response time for SSS trials in increasingly complex conditions, $F(2, 48) = 4.739$, MSE $= 110{,}669$, $p = .01$. The complexity of the condition in which SSS trials were performed thus affected RTs, making it unlikely that in relation-level and system-level conditions SSS trials were performed faster than DSS or DDS trials simply because participants always responded according to an analysis at the attribute level of complexity.

4. Experiment 3

In Experiments 1 and 2, subjects compared a figure to the one preceding in order to determine a response, then stored appropriate information about that figure for comparison to the following figure. Thus, each figure served as the probe for one trial, to be compared to the already seen sample, and then as the sample for the subsequent trial. The effort of maintaining relationally complex representations over a 5-s delay led to an increase in RTs in Experiment 2 relative to those in Experiment 1, presumably because the longer delay increased the need for rehearsal to avoid decay of the representation. It may be assumed, then, that maintaining the representation of a sample placed demands on working memory and that these might conflict with the demands of encoding the probe and comparing it to the sample when the two were separated by only a 150-ms interval, since subjects might still be consolidating their representation of the sample when presented with the probe. This conflicting demand on processing capacity would impact response times most in the complex conditions, and thus serves as an alternate explanation of the pattern of response times observed in Experiments 1 and 2.

To control for this possibility, it would be useful to temporally separate the comparison of probe and sample from the encoding of the sample, but without adding a delay as in Experiment 2. In this way we might observe the effect of complexity on the activity of comparing a representation consolidated in working memory to a new stimulus without confounding influences from encoding the previous stimulus. Experiment 3 was designed to accomplish this separation of the encoding of the initial pattern from its comparison with the subsequent pattern.

4.1. Method

The design of Experiment 3 closely followed that of Experiment 1, but each successive pattern served only as a sample or as a probe. The ISI was 150 ms, as in Experiment 1. The first pattern presented was a sample. Participants were instructed to determine the information about the sample appropriate to the current complexity condition and then press a key when they were ready for the probe pattern. When the key was pressed, the sample was replaced by a blank screen, and after the 150 ms ISI the probe appeared. The participant was then to press the "n" or "c" key to indicate whether the probe had changed or had not changed relative to the sample. *RT* was measured from the time the probe appeared until the participant pressed a key indicating their response. Following a response, a blank screen again appeared for the 150 ms inter-trial interval, and the next sample was presented.

Thirty-five UCLA undergraduates participated in the experiment as part of the requirements for an introductory psychology course. Aside from modifications just mentioned, instructions and administration of trial blocks were the same as in Experiment 1. Participants performed 30 trials in each of the nine blocks for a total of 270 trials across the three conditions.

4.2. Results and discussion

The results replicated the pattern observed in Experiments 1 and 2. Mean correct RTs for attribute, relation, and system levels in no-change trials were 1,003, 1,376, and 1,686 ms, respectively; and in change trials 1,113, 1,467, and 1,760 ms, respectively. As depicted in Fig. 7, when participants were comparing a representation in working memory to a new pattern there was a monotonic increase in RTs as complexity level increased, $F(2, 68) = 81.434$, MSE = 95,364, $p < .0001$. Mean accuracy in the three conditions was 98, 95, and 94%, respectively. No-change trials were faster than change trials, $F(1, 24) = 6.759$, MSE = 65,817, $p < .05$, and there was no interaction with complexity level, $F(2, 68) = 0.21$, MSE = 26,517, $p = .81$. Thus, when consolidation of the representation of the sample was removed as a potential contributor to RT, the monotonic increase in RT observed in Experiments 1 and 2 as relational complexity increased was again replicated.

To further investigate the contribution of cooperative processing, a comparison was made in the relation-level condition between no-change trials for which the attribute-level response would also be no change, and those in which a change occurred at the attribute level. Mean correct RTs were 1,226 and 1,807 ms, respectively, $F(1, 33) = 94.7$, MSE = 60,545, $p < .0001$.

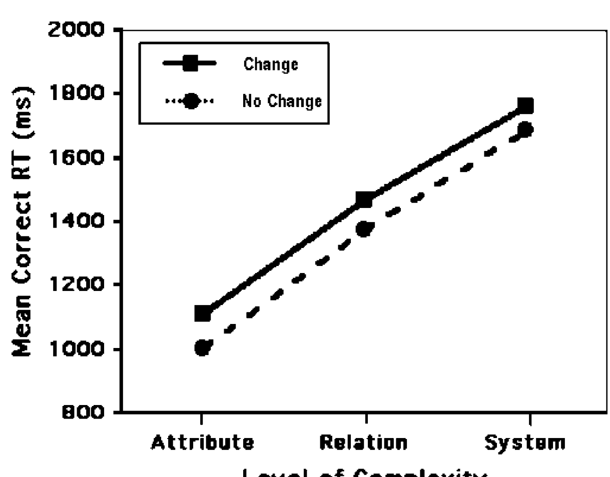

Fig. 7. Mean correct RTs for the three levels of complexity in Experiment 3. Means for no-change trials are shown separately from means for change trials.

A comparison was also made between no-change system-level trials in which no change occurred at the attribute and relation levels of complexity (SSS), those in which a change occurred at the attribute level but not the relation level (DSS), and those in which a change occurred at both the attribute and relation levels of complexity (DDS). Mean RTs for this comparison were 1,415, 1,943, and 2,118, respectively, $F(2, 66) = 39.93$, MSE = 114,125, $p < .0001$. Thus, as in Experiments 1 and 2, participants were faster when the answer considered at lower levels of complexity was in agreement with the nominal no-change answer for both relation-level judgments, and system-level judgments. These results further support the presence of cooperative processing in trials performed at the relation and system levels of complexity.

As in the previous two experiments, we compared mean RTs for SSS trials across the three complexity conditions to discover whether savings for SSS and DSS trials were due to a shortcut or least-complexity strategy rather than cooperative processing at multiple levels of relational complexity. Means were 1,002, 1,226, and 1,414 ms, respectively, revealing a monotonic increase in RT for SSS trials as the complexity of the condition increased, $F(2, 66) = 45.5$, MSE = 31,895, $p < .0001$. Thus, the processing participants engaged in to solve higher complexity trials was not the same as the processing they applied to solution of SSS trials in the attribute condition.

5. Experiment 4

Because Experiment 3 used a short (150 ms) ISI, we cannot rule out the possibility of some contribution of iconic memory to the pattern of results. Accordingly, in Experiment 4 we replicated the design of Experiment 3 but used a 5-s ISI, as in Experiment 2. To further assure that iconic memory for the initial pattern was eliminated, a colored pattern mask was displayed for the entire 5-s ISI. This mask would be expected to obliterate any iconic storage.

A second objective of Experiment 4 was to separately evaluate the time to encode stimuli at the different levels of relational complexity, and the time to perform comparisons for different complexity levels. In Experiments 1–2, stimuli served first as probe to be compared to the previous stimulus, then as the sample to be compared to the subsequent stimulus; thus subjects were presented with a series of stimuli to which binary “sameness” judgments at varying levels of relational complexity were made. Implicit in this design was the assumption that a representation at the appropriate level of relational complexity was calculated for each stimulus once, and that the reaction time on each trial resulted from this processing together with the time to compare the representation to that of the previous stimulus. However, it could be argued that the time to process a stimulus for purposes of comparing it to the previous stimulus, and processing the stimulus for purposes of storing a representation to compare to the subsequent stimulus, might have separately contributed to the reaction times observed. It is therefore unclear whether the RT patterns observed resulted from differences in encoding times or comparison times. In addition, interactions between encoding and comparison times might affect the relative RTs observed (e.g., longer comparison times may reduce the time necessary for subsequent encoding of a stimulus for storage). For these reasons, Experiment 4 attempted to measure the time to encode patterns at each level of complexity, and separate this time from a measure of the time to compare the two patterns. This was done by recording two reaction times on each trial: the time to encode the first pattern and prepare for the delay (the sample RT), and the time to make the required comparison after the second pattern was presented (the probe RT). If the increase in comparison times across complexity levels observed in the previous experiments were entirely attributable to increases in encoding time across the levels, we would expect to observe the same pattern of increase in the sample RT as in the probe RT.

5.1. Method

The method of Experiment 4 was identical to that of Experiment 3, except that (1) the ISI was increased to 5 s; (2) a colored pattern mask was presented throughout the ISI; and (3) the sample RT, the time to encode the first pattern and prepare for the delay, was measured in addition to the probe RT (comparison time after the second pattern was presented). The mask used on a trial was randomly selected from a set of six different pattern masks, each composed of randomly-sized and randomly-oriented overlapping ovals. The ovals were in any of eight different shades, which did not overlap with the colors used in the stimulus patterns.

Twelve undergraduate and graduate students at UCLA and Princeton University served as volunteer subjects.

5.2. Results and discussion

5.2.1. Sample RT

Fig. 8A depicts the pattern of sample *RT*s across the three levels of complexity. Sample RT was measured from the onset of the first pattern until the subject pressed a button to terminate its presentation, which was then replaced by the mask, followed after 5 s by presentation of the second pattern. Mean sample RTs for attribute, relation, and system levels were 2,278, 1,655, and 1,690 ms, respectively, $F(2, 22) = 7.199$, MSE = 203,913, $p < .005$.

This pattern is extremely different from that observed for comparison times in the previous experiments (or the present Experiment 4; see discussion of probe RTs below). Rather than showing a monotonic increase with complexity level, sample RTs were much higher for the attribute level than for the relational or system level, which did not differ reliably from each other. Although not anticipated, this pattern is quite sensible given the nature of the task. Sample RT includes both time to encode the first pattern and time to prepare for the 5-s ISI, during which a colored mask would be present. As we pointed out earlier, the number of separate pieces of information that must be maintained over the delay interval is greatest for the attribute level (four colors), followed by the relation level (two first-order same/different decisions) and the system level (one second-order same/different decision). Given that the limit on the capacity of visual working memory appears to be four to five elements (Bundesen, 1998; Luck & Vogel, 1997; Sperling, 1960), only the attribute level strained the capacity of working memory. In addition, given that the mask consisted of colored patterns, the attribute level was likely most vulnerable to interference from the mask (as it required that four colors be remembered over the ISI).

5.2.2. Probe RT

The results for mean correct probe RT—the time to make a match/mismatch decision after the second pattern was presented—replicated the pattern of comparison times observed in

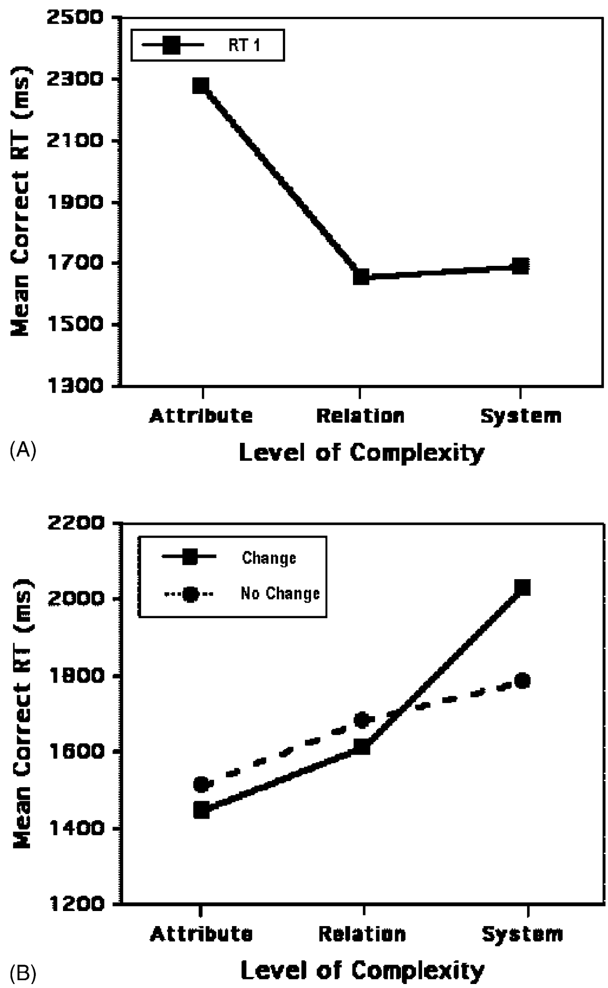

Fig. 8. (A) Mean sample RTs for the three levels of complexity in Experiment 4. (B) Mean correct probe RTs. Means for no-change trials are shown separately from means for change trials.

the previous experiments. Mean correct probe RTs for attribute, relation, and system levels in no-change trials were 1,515, 1,683, and 1,788 ms, respectively; and in change trials 1,446, 1,614, and 2,032 ms, respectively. As depicted in Fig. 8B, when participants were comparing a representation in working memory to a new pattern there was a significant main effect of comparison level on reaction time, $F(2, 22) = 12.274$, MSE $= 37{,}612$, $p < .001$. Mean accuracy in the three conditions was 97, 96, and 97%, respectively.

The results of Experiment 3, together with those of Experiment 4, make it clear that the increase in RT that accompanied added relational complexity in Experiments 1 and 2 was not solely due to the use of each stimulus as first a probe and then as a sample. In Experiment 3, the probe stimuli never served as samples, and in Experiment 4, the time to encode each stimulus when acting as sample was measured separately from the time to compare each stimulus acting as probe to the previous stimulus. The patterns of probe RTs in Experiments 3 and 4 were consistent with the pattern of RTs observed in Experiments 1 and 2.

It is possible that overall, subjects needed to compare more of the four stimulus squares to make responses in the more relationally complex trials. For example, in some attribute-level trials for which the answer was "change," subjects could answer after considering a single square, if that square differed between sample and probe. Thus, subjects could potentially respond and terminate a trial after considering on average fewer squares for the relation-level trials than for the system-level trials, and for the attribute-level trials than for the relation-level trials. While we cannot fully discount any influence of such processing, three observations suggest this description does not portray the way subjects typically completed trials. First, by this logic, the slowest trials would be no-change trials, since for these trials subjects would always need to perform an exhaustive comparison of all squares, regardless of relational complexity level. This was clearly not the case, as no-change trials were always faster than change trials. Second, for no-change trials, RT increased with relational complexity. There is no reason to think this increase resulted from a self-terminating serial processing of the four squares, as the only difference between no-change trials at the three relational complexity levels was the difference in relational complexity.

To further investigate the contribution of cooperative processing, a comparison was made in the relation-level condition between no-change trials for which the attribute-level response would also be no change, and those in which a change occurred at the attribute level. Mean correct RTs were 1,671 and 1,655 ms, respectively, $F < 1$. The lack of a difference between these two conditions (which differed in the three previous experiments) may have resulted because presentation of the visual mask made retention and use of attribute-level information more difficult, so that the contribution of attribute-level information to higher complexity judgments was attenuated in Experiment 4. A comparison was also made between no-change system-level trials in which no change occurred at the attribute and relation levels of complexity (SSS), those in which a change occurred at the attribute level but not the relation level (DSS), and those in which a change occurred at both the attribute and relation levels of complexity (DDS). Mean RTs for this comparison were 1,778, 1,940, and 2,337, respectively, $F(2, 22) = 9.66$, MSE = 103,020, $p < .001$. As in the previous experiments, participants were faster at making system-level judgments when the answer considered at lower levels of complexity was in agreement with the nominal no-change answer.

As in the previous experiments, we also compared mean RTs for SSS trials across the three complexity conditions to discover whether savings for SSS and DSS trials were due to a shortcut or least-complexity strategy rather than cooperative processing at multiple levels of relational complexity. Means were 1,521, 1,671, and 1,777 ms, respectively. A significant main effect of comparison level was observed, $F(2, 22) = 6.809$, MSE = 29,217, $p < .005$. As in the previous experiments, the processing in which participants engaged to solve higher complexity trials was not the same as the processing they applied to solution of SSS trials in the attribute condition.

In summary, the pattern of results for probe RTs in Experiment 4 essentially replicated the pattern of comparison times observed in the previous experiments, and was very different from the pattern of sample RTs. Although we assume that extra encoding time is required to abstract first-order relations and higher order relations from the visual pattern of four squares, sample

RT was clearly influenced by additional processes. As we noted above, sample RT presumably included not only the time to form a basic encoding of the first pattern, but also the time to prepare for the delay and pattern mask. It appears that the attribute level is most vulnerable to interference over the delay, and hence requires the most preparation time. Further studies will be required to fully disentangle the various contributions to processing time in this type of comparison task; but the present results demonstrate that the pattern of comparison times (probe RT) is not simply a reflection of the times required for those processes that determine sample RT.

6. General discussion

The four experiments reported here establish a new method for investigating how people process relations at different levels of complexity. Decisions about whether or not two figures matched took increasingly more time as the "match" was defined at progressively higher levels of complexity. At the same time, matches at lower levels contributed to decisions at higher levels, providing evidence for an inter-dependent, cooperative mode of processing. The overall pattern of relational processing was virtually identical regardless of whether the ISI was brief (150 ms in Experiments 1 and 3) or longer (5 s in Experiments 2 and 4). Adding a colored mask during the ISI (Experiment 4) did not alter the basic pattern. Responses tended to be slower when the interval was longer, and accuracy decreased slightly across conditions, consistent with previous findings concerning the effect of delay in match-to-sample studies (e.g., Bodner, Kroger, & Fuster, 1996).

Multi-level processing of the sort observed in the present study is consistent with computational models of analogical mapping (e.g., Falkenhainer et al., 1989; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997), as well as with interactive models of similarity judgments (Goldstone, 1994; Goldstone & Medin, 1994a). Simulations established that the LISA model (Hummel & Holyoak, 1997, 2003) of analogical mapping can be extended to account for the pattern of RTs to make sameness judgments at multiple levels of complexity (see supplemental information at http://cogsci.psy.utexas.edu/supplements/). Our results are broadly consistent with other evidence from tasks based on speeded and unspeeded similarity judgments, which also have found evidence that people abstract and use higher order relations derivable from perceptual displays (Goldstone & Medin, 1994b; Goldstone, Medin, & Gentner, 1991; Markman & Gentner, 1993).

The pattern of comparison times in our experiments demonstrated that lower level matches contributed to decisions at higher levels of abstraction. For example, matches at the relational level were made more rapidly for patterns that also matched at the attribute level (i.e., RTs were faster for SSS than for DSS trials). The design of the present experiments did not allow any clear test of whether inter-dependent processing includes influences in the opposite direction, as would be revealed if "mismatch" decisions at lower levels were slower for patterns that matched at higher levels than for those that did not (e.g., slower RT to detect an attribute level mismatch for DSS than for DDS trials).

Such "abstract-to-concrete" influences would most readily be detected in "mismatch" trials, as in the above example; however, the manner in which "mismatch" trials were constructed in the present set of experiments precluded making such comparisons. Love, Rouder, and Wisniewski (1999) reported a set of experiments that appear to show that shared perceptual relations slow "mismatch" decisions at the attribute level. For example, their subjects were slower to report a mismatch when a column composed of three circles was compared to a column composed of three triangles (i.e., when the relation "column of same forms" was shared between the patterns) than when the second stimulus consisted of a column composed of a circle above two triangles (i.e., "column of different forms"). Love et al. interpreted their findings as evidence of "global-to-local" influences on comparisons similar to those posited by models of analogy. However, their findings are open to alternative interpretations based on low-level perceptual processing. In particular, a column of identical elements may be grouped by relatively early perceptual mechanisms that group texture elements into perceptual units (e.g., Julesz, 1975, 1981), or on the basis of the Gestalt principles of similarity and proximity. By contrast, a column of three different elements would not be grouped in this manner. Thus, a column of three circles would be perceived as a column, as would a column of three triangles (rendering them perceptually similar), whereas two triangles and a circle would not be perceived as a column (rendering it dissimilar to either column). Further research will be required to determine whether and when truly abstract perceptual relations of the sort used in the present study influence processing at lower levels.

Future research may endeavor to link abstract same-different judgments to the neural substrate of relational reasoning. Recent neuroimaging studies (Christoff et al., 2001; Kroger et al., 2002; Prabhakaran et al., 1997) have revealed activity in dorsolateral prefrontal cortex accompanying mediation of relationally complex information. Deficits in relational reasoning by frontal patients (Waltz et al., 1999) have also been found. Fuster (1997) has argued that the human prefrontal cortex subserves abstract linkage of information across time and space. Future research may determine whether the frontal cortex specifically plays an important role in processing the concept of sameness at various levels of abstraction.

Note

1. In all experiments, analyses of RTs for selected item types excluded data from any subject who did not solve any trials correctly for some item type. For this reason the degrees of freedom and mean correct RTs reported in the text vary slightly across different analyses of overlapping sets of item types. These minor variations do not alter any ordinal comparisons.

Acknowledgments

Preparation of the paper was supported by NSF Grant SBR-9729023. The paper is based in part on a Ph.D. dissertation completed at UCLA by the first author under the direction of the second author. Preliminary analyses of Experiments 1 and 2 were presented at the 1997 conference of the Cognitive Science Society (Kroger & Holyoak, 1997), and a preliminary version of the LISA simulations was presented at the 1998 conference of the Cognitive Neuroscience

Society. We thank Abi Gopal for helping with data analyses and Bob Morrison for assistance with Experiment 4.

References

Bodner, M., Kroger, J. K., & Fuster, J. M. (1996). Auditory memory cells in dorsolateral prefrontal cortex. *Neuroreport*, *71*, 905–1908.

Bundesen, C. (1998). Visual selective attention: Outlines of a choice model, a race model and a computational theory. *Visual Cognition*, *5*, 287–309.

Christoff, K., Prabhakaran, V., Dorfman, J., Kroger, J. K., Zhao, Z., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage*, *14*, 1136–1149.

D'Amato, M. R., Salmon, D. P., Loukas, E., & Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (*Cebus apella*) and pigeons (*Columba livia*). *Journal of the Experimental Analysis of Behavior*, *44*, 35–47.

D'Amato, M. R., Salmon, D. P., Loukas, E., & Tomie, A. (1986). Processing of identity and conditional relations in monkeys (*Cebus apella*) and pigeons (*Columba livia*). *Animal Learning and Behavior*, *14*, 365–373.

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

Fujita, K. (1982). An analysis of stimulus control in two-color matching-to-sample behaviors of Japanese monkeys *Macaca fuscata fuscata*. *Japanese Psychological Research*, *24*, 124–135.

Fuster, J. M. (1997). *The prefrontal cortex* (3rd ed.). New York: Raven Press.

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

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

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

Goldstone, R. L., & Medin, D. L. (1994a). Similarity, interactive activation, and mapping: An overview. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2. Analogical connections* (pp. 321–362). Norwood, NJ: Ablex.

Goldstone, R. L., & Medin, D. L. (1994b). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 29–50.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, *23*, 222–262.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.

Halford, G. S., & Wilson, W. H. (1980). A category theory approach to cognitive development. *Cognitive Psychology*, *12*, 356–411.

Hofstadter, D. R., & Mitchell, M. (1994). The Copycat project: A model of mental fluidity and analogy-making. In K. J. Holyoak & J. A. Barnden (Eds.), *Analogical connections. Advances in connectionist and neural computation theory* (Vol. 2, pp. 31–112). Norwood, NJ: Ablex.

Holyoak, K. J., & Hummel, J. E. (2000a). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 229–263). Mahwah, NJ: Erlbaum.

Holyoak, K. J., & Hummel, J. E. (2000b). The proper treatment of symbols in a connectionist architecture. In E. Deitrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 229–263). Mahwah, NJ: Erlbaum.

Holyoak, K.J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. Cognitive Science, 13, 295–355. Reprinted in T.A. Polk & C.M. Siefert (Eds.) (2002), *Cognitive modeling*. Cambridge, MA: MIT Press.

Holyoak, K.J., & Kroger, J.K. (1995). Forms of reasoning: Insight into prefrontal functions? In J. Grafman, K.J. Holyoak & F. Boller (Eds.), *Structure and functions of the human prefrontal cortex*. New York: New York Academy of Sciences.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge: MIT Press.
Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.
Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220–264.
Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American*, *232*, 34–43.
Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature*, *290*, 91–97.
Keane, M. T., & Brayshaw, M. (1988). The incremental analogical machine: A computational model of analogy. In D. Sleeman (Ed.), *European working session on learning* (pp. 53–62). London: Pitman.
Kroger, J. K., & Holyoak, J. J. (1997). What it means to be "the same": The impact of relational complexity on processing efficiency. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th conference of the cognitive science society* (pp. 406–411). Mahwah, NJ: Erlbaum.
Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of dorsolateral and anterior prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex*, *12*, 477–485.
Love, B. C., Rouder, J. N., & Wisniewski, E. J. (1999). A structural account of global and local processing. *Cognitive Psychology*, *38*, 291–316.
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.
Nissen, H. W., Blum, J. S., & Blum, R. A. (1948). Analysis of matching behavior in chimpanzee. *Journal of Comparative and Physiological Psychology*, *41*, 62–74.
Oden, D. L., Thompson, R. K., & Premack, D. (1988a). Spontaneous transfer of matching by infant chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 140–145.
Oden, D. L., Thompson, R. K., & Premack, D. (1988b). Can an ape reason analogically? Comprehension and production of analogical problems by Sarah, a chimpanzee (*Pan troglodytes*). In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 471–497). Cambridge, MA: MIT Press.
Piaget, J., Montangero, J., & Billeter, J. (1977). La formation des correlats. In J. Piaget (Ed.), *Recherches sur l'abstraction reflechissante I* (pp. 115–129). Paris: Presses Universitaires de France.
Prabhakaran, V., Smith, J. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology*, *33*, 43–63.
Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences*, *6*, 125–167.
Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, *27*, 125–140.
Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 146–178). New York: Cambridge University Press.
Sperling, G. (1960). The information available in brief presentations. *Psychological Monographs*, *74* (whole no. 498).
Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 31–43.
Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*, 119–125.
Wright, A. A., Shyan, M. R., & Jitsumori, M. (1990). Auditory same/different concept learning by monkeys. *Animal Learning and Behavior*, *18*, 287–294.