

Running Head: EXPERTS AND NOVICES

PROBLEM PROCESSING BY EXPERTS AND NOVICES:
THE CASE OF ARITHMETIC

Vladimir M. Sloutsky

and

Aaron S. Yarlas

The Ohio State University

PAPER IN PROGRESS

Address correspondence to

Vladimir M. Sloutsky
21 Page Hall
1810 College Avenue
The Ohio State University
Columbus, OH 43210
Phone: (614) 688-5855
Fax: (614) 292-0321
Email: Sloutsky.1@osu.edu

Abstract

It has been well established that experts and novices focus on different aspects of problems, with novices focusing more on surface features rather than on deep relational features. What is less clear are the mechanisms that underlie these differences in the construction of problem representation. The article describes three studies that use an 'old/new' recognition procedure to examine expert and novice representation of arithmetic equations for which the deep relational properties (i.e., principles of commutativity and associativity) were familiar to both groups. Results from these studies indicate that both novices and experts encode and access both surface and principled features, and that they encode and access these features in a serial manner, with surface features preceding principled features. At the same time, only for novices but not for experts, surface features compete with deep principled features, thus requiring additional resources to inhibit this response competition.

PROBLEM PROCESSING BY EXPERTS AND NOVICES:
THE CASE OF ARITHMETIC

It has been well established that, when presented with a problem, experts and novices focus on different aspects of the problem (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Larkin, 1983; Simon & Simon, 1978; Reed, Ackinclose, & Voss, 1990). While experts are more likely to focus on theoretically-important structural properties of a problem, such as equations that can be used to solve the problem, novices are more likely to focus on less important surface features of a problem, such as the problem's storyline. By deep relational, or structural, properties we are referring to those aspects of the problem that (1) are not directly perceptible, (2) are relational, and (3) afford formulation of principles in a knowledge domain (e.g., Boyle's law as opposed to merely accidental properties of gases).

These types of novice-expert differences, with experts focusing on deep principles and novices focusing on surface features, have been demonstrated in a variety of knowledge domains, including chess (Chase & Simon, 1973), mathematics (Blessing & Ross, 1996; Hinsley, Hayes, & Simon, 1977; Bassok, 1996, 1997; Novick, 1988; Reed, et al, 1990; Ross & Kilbane, 1997; Schoenfeld & Herman, 1982; Silver, 1981), physics (Chi, et al 1981; Simon & Simon, 1978; Larkin, 1983; Larkin, McDermot, Simon, & Simon, 1980), and computer programming (Adelson, 1984). Similar effects have been observed in a variety of knowledge-lean domains, such as deductive and inductive inference. When presented with deduction problems, untrained reasoners often ignore the argument's logical form (i.e., its deep structure) while relying on the argument's surface features, such as content and believability (Cheng & Holyoak, 1985; Evans, Newstead, & Byrne, 1993; Johnson-Laird & Byrne, 1991; Byrne, 1989). When presented with induction and analogy problems, novices and young children also often

ignored deep relational structure while relying on the surface features (Gentner, 1989; Gentner & Toupin, 1986; Kotovsky & Gentner, 1996; Holyoak & Koh, 1987).

However, while these expert-novice differences have been well established, it remains unclear what accounts for these differences. Do differences occur because deep relational properties are familiar to experts but not to novices? Do they occur because novices are less intelligent than experts are, and they cannot grasp deep relational properties? Do experts and novices differ in processes underlying the construction of a problem representation? Or do differences stem from a combination of these factors?

In recent research focused on expert-novice differences, it was demonstrated that when tasks were sufficiently simple and deep relational properties were familiar even to novices, neither differences in knowledge nor in intelligence could account for the observed differences between novices and experts (Yarlas & Sloutsky, 2000). In a series of experiments designed to distinguish among these possibilities, Yarlas and Sloutsky (2000) constructed tasks that included principles of arithmetic familiar to novices and surface features that were completely superfluous with respect to deep relational features. In particular, they asked participants varying in age, general intelligence, and degree of mathematical expertise to sort arithmetic equations that could have common surface elements (e.g., commonality of numbers or the same number of constituent addends in the equation) or common deep mathematical principles (e.g., commutativity or associativity). Results indicated that only mathematics experts consistently focused on principles, whereas novices, regardless of age and general ability, focused mostly on surface features. However, holding shared surface features constant across problems, such that some problems shared only principles and others shared neither principles nor surface features, led to a substantial increase in novices' focusing on these principled properties. Interestingly,

however, the reintroduction of varying surface features reduced novice participants' focus on principles to their original low levels. These and other manipulations support these researchers' position that differences between novices and experts stem from differences in problem representation rather than familiarity with principles, age, or intelligence. In short, even when principles are familiar, novices fail to focus on these principles if the principles compete with salient surface features, whereas experts focus almost exclusively on the principles regardless of the competing surface features. Therefore, several questions arise about processes underlying problem representations in novices and experts. Do novices initially encode both deep and surface features, but later discard the deep relational properties, or do they simply fail to encode the deep relational properties? And what are the processing mechanisms underlying problem representations in experts: do experts encode and discard surface features, or do they ignore these features from the very beginning?

The current studies attempt to determine whether novices and experts represent only one type of feature, or both surface and structural features of problems within a simple task. As discussed above, there is much evidence that novices fail to represent structural aspects of problems. However, if novices have knowledge of and are familiar with the structural relations in question, under simple task conditions, it is possible that they will represent relations in addition to surface features. There is also evidence suggesting that experts may attend to both surface and structural features. In particular proficient problem solvers as defined by their high SAT scores (the criterion that is not necessarily indicative of mathematical expertise) do not ignore surface features of problems, but attend to them as well as to structural features (Novick, 1988). The current experiments will examine which features novices and experts represent for familiar

relations for a simple and unpressured task condition, as well as a task in which time pressure is enforced during encoding of problems.

Previous research has indicated that processing of surface and structural features follows a time course, with surface features processed prior to relations (Goldstone & Medin, 1994; Ratcliff & McKoon, 1989). For example, Goldstone & Medin (1994) found that similarity judgments for object arrangements were initially based upon matching surface features, but after a delay, these judgments were driven more by shared relational correspondences among objects. Findings such as these provide evidence that representations of surface features precede those of structural relations. However, given that these findings have been based upon knowledge-lean tasks (e.g., similarity of objects and object arrangements), it has not been established whether the same time course of processing would hold for novices and experts within a knowledge-rich domain. For example, it is viable to believe that experts might process structural features prior to surface features, that they may process the two types of features in parallel, or that they may, like novices, process surface features prior to structural features with less of a delay between the two. The current research attempts to examine these possibilities by closely examining processes of encoding and accessing of surface and structural features of knowledge-rich problems by experts and novices.

One way to examine encoding and access of features is to use recognition or recall memory tasks that have been used in the past to examine problem representations of both adults and young children (e.g., Siegler, 1978). Consistent recall or recognition of a component of a task is indicative of the fact that the component has been encoded. Hence, to address the issue of feature encoding we used an 'old/new' recognition paradigm. This paradigm affords the creation

of a set of foils, such that patterns of hits and false alarms indicate which aspects of problems have been encoded and committed to memory and which aspects have been left out.

The recognition paradigm used in the three experiments described here consisted of study and recognition phases. In the study phase, participants were presented with a set of arithmetic equations. These equations all utilized a principled property of mathematics, either associativity or commutativity. The former states that for addition, subtraction, and multiplication, constituent parts can be decomposed and recombined in different ways (e.g., $a + b = [a - c + c] + b$). The latter states that the order of elements is irrelevant for addition and multiplication (e.g., $a + b + c = b + c + a$). In addition, these equations all used consistent levels of two surface elements: all equations used numbers ranging between 1 and 9, and all used either 5 or 6 numbers in the equation. In the recognition phase, in addition to 'old' items that had been presented in the study phase, four combinations of 'new' equations were presented as foils. Half of these foils, which we refer to as 'feature +' foils, maintained the same levels of surface features as used in the learning phase (i.e., numbers ranging between 1 and 9, and either 5 or 6 numbers in the equation), while the other half of the foils, which we refer to as 'feature -' foils, violated these categories (i.e., numbers greater than 9, and either 4 or 7 numbers in the equation). Also, half of the foils, which we refer to as 'principle +' foils, maintained the use of one of the two principled properties, while the other half, which we refer to as 'principle -' foils, did not use any principled properties in the equation. The two levels of the two kinds of properties (feature being either + or -, and principles being either + or -) were fully-crossed, thus creating four combinations of foils: feature + /principle + (F+/P+), feature + /principle - (F+/P-), feature -/principle + (F-/P+), and feature -/principle - (F-/P-).

An inherent assumption of this paradigm is that participants (both novices and experts) will not encode and store the specific items presented in the study phase, due to the fact that the items used in the study phase are not highly differentiated from each another. Rather, it is expected that participants will encode study items categorically based on their surface and/or principled features, such that they will reject recognition items that violate these categories and accept those that do not. The accuracy of this assumption would be evidenced by findings that participants do not differentiate between F+/P+ foils and Old targets, and thus would inaccurately judge F+/P+ foils to be “Old” at above-chance levels.

Recall that the paradigm is used to determine which aspects of arithmetic problems experts and novices encode and which aspects they leave out. At the same time, in addition to feature encoding, recognition also involves feature access and retrieval. Therefore, processing models considered below will address both the encoding and accessing of surface and deep principled features of arithmetic problems. As discussed above, across a variety of tasks and knowledge domains, novices tend to focus on surface features of the problem, while experts tend to focus on deep relational properties of the task. Therefore, while the range of possible processing models is rather large, we consider only those models that are compatible with prior findings. In particular, given the reported expert-novice differences, it seems implausible that novices encode only deep relational features, or that experts encode only surface features. Those processing models that are compatible with prior findings are presented graphically in Figures 1 and 2 for novices and experts, respectively.

If novices encode only surface features (Figure 1, Model 1), they should respond "Old" when surface features are present and they should respond "New" when surface features are absent, regardless of the presence or absence of principles. On the other hand, if novices encode both

surface and deep features (Figure 1, Model 2), they should respond “Old” when both surface and principled features are present (i.e., F+/P+ foils and Old targets) and respond “New” when surface and/or principled features are absent (i.e., F+/P-, F-/P+, and F-/P- foils).

Similar reasoning could be applied to experts. If experts encode only deep principles (Figure 2, Model 1), they should respond "Old" when principles are present and they should respond "New" when principles are absent. On the other hand, if experts encode both surfaces and principled features (Figure 2, Models 2 and 3), they should respond “Old” when both surface and principled features are present (i.e., F+/P+ foils and Old targets) and respond “New” when surface and/or principled features are absent (i.e., F+/P-, F-/P+, and F-/P- foils). The critical difference between Models 2 and 3 is as follows. According to Model 2, experts access deep principled properties first, and, therefore, their responses to P- foils should be faster than their responses to P+ foils. On the other hand, according to Model 3, experts access surface features first, and, therefore, their responses to F- foils should be faster than their responses to F+ foils. Of course, it is also possible that experts access both types of features in a parallel manner, in which case their response latencies should exhibit small or no differences across the foils.

Note that processing models presented in Figures 1 and 2 depict possibilities of feature encoding and the order of feature accessing in the course of recognition. The analysis of feature accessing is important because it affords differentiation between Models 2 and 3 in experts. In particular, if participants access features in a serial manner then each additional step in processing leads to increase in latencies. Therefore, we can derive testable predictions from each of the models presented in Figures 1 and 2. These predictions are presented in Table 1. The predictions are qualitative, in that they do not specify accuracy or latency across the conditions,

but rather point to (a) patterns of recognition responses and (b) directions of differences in latencies.

Several foils are critical for testing these predictions. Responses to F+/P- foils afford either corroboration or elimination of Model 1 for novices, whereas responses to F-/P+ foils afford corroboration or elimination of Model 1 for experts. In addition, the patterns of differences in latencies afford the selection of the more plausible model as well as the description of specific processing components. Specifically, if participants access principled features first, while encoding both surface and principled features (Model 2 for experts), P- foils should generate faster responses than P+ foils. This is because P- foils can be rejected in a single step, whereas P+ foils would require additional steps. Therefore, the following pattern of latencies should be observed: Old targets = F+/P+ = F-/P+ > F+/P- = F-/P-. On the other hand, if participants access surface features first, while encoding both surface and principled features (Model 2 for novices and Model 3 for experts), F- foils should generate faster responses than F+ foils. Therefore, the following pattern of latencies should be observed: Old targets = F+/P+ = F+/P- > F-/P+ = F-/P-. Finally, if participants access both surface and principled features in a parallel manner, latencies should not differ across the foils types, thus leading to the following pattern: Old targets = F+/P+ = F+/P- = F-/P+ = F-/P-. The goal of Experiment 1 was to test processing predictions presented in Table 1 in order to determine which of the proposed processing models for novices and experts were most accurate in predicting their encoding and access of surface and principled features.

EXPERIMENT 1

Method

Participants

Two samples, representing novices and experts, were used in this study. The novice group included 23 undergraduates in an introductory psychology course at a large Midwestern university who participated for partial course credit. This sample had an average age of 19.2 years ($SD = 0.9$ years), with 12 women and 11 men. The expert group included 12 graduate students in the Mathematics Department at the same university who participated for a payment of twenty dollars. This sample had an average age of 27.6 years ($SD = 5.8$ years), with 3 women and 9 men.

Materials and Procedure

The materials and procedures used in this study were identical for participants in both the novice and expert samples. All participants were run individually with stimuli presented by a personal computer using SuperLab software (Cedrus Corporation, 1999).

The experiment consisted of three phases: the study phase, the distraction phase, and the recognition phase. In the study phase, participants were presented with 30 arithmetic equations, which they had been instructed to memorize. All 30 equations incorporated the mathematical operation of addition, used numbers ranging from 1 to 9, contained either 5 or 6 numbers, and used either the associative or commutative principle (half for each). Each equation was centered and presented in black type on a white background for ten seconds, with a two-second interval between each equation, during which only the white background was seen. The order of presented equations was randomized. The list of all 30 equations presented in the study phase is given in the Appendix.

A distraction phase followed the study phase for the purpose of clearing participants' short-term memory. For the distractor task, participants were presented with 90 letters, for which they had been instructed to indicate whether the letter was a vowel (by pressing the "Z" key on the keyboard) or a consonant (by pressing the "M" key). Each letter was centered and presented in black type on a white background. This phase took approximately three minutes.

Following the distractor phase was the recognition phase. Participants were instructed that they would be presented with a number of arithmetic equations, some of which had been presented to them earlier and some which had not been presented earlier. They were further instructed to indicate whether each equation had been presented earlier or not, by pressing the "Z" key on the keyboard if the item was presented earlier, or the "M" key if it had not been presented earlier.

There were a total of 60 equations presented in the recognition phase. Each equation was centered and presented in black type on a white background. The order of equations presented in this phase was randomized. These equations fell into five categories, with 12 exemplars for each category. The first category contained "Old targets", which consisted of equations that had been randomly selected from those that had been presented earlier in the learning phase. The remaining four categories were foils, in that they contained new problems that had not been presented in the study phase. The first type of foil consisted of "feature + /principle +" (F+/P+) equations that used similar surface feature as the original equations (i.e., numbers between 1 and 9, and either 5 or 6 numbers), and used either the commutativity or associativity principle (e.g., $7 + 1 + 4 = 1 + 7 + 4$ and $9 + 7 = 5 + 4 + 7$). The second type of foil consisted of "feature + /principle -" (F+/P-) equations that used similar surface features as the original equations but did not use either the commutativity or associativity principle (e.g. $3 + 1 + 9 = 2 + 5 + 6$ and $4 + 9 =$

3 + 8 + 2). The third type of foil consisted of "feature - /principle +" (F-/P+) equations that used surface features different from those used in the original equations (i.e., numbers greater than 9, and either 4 or 7 numbers), and used either the commutativity or associativity principle (e.g., $7 + 12 = 12 + 7$ and $5 + 1 + 15 = 3 + 2 + 1 + 15$). The fourth type of foil consisted of "feature - /principle -" (F-/P-) equations that used surface features different from those used in the original equations and did not use either the commutativity or associativity principle (e.g., $8 + 3 + 12 = 5 + 1 + 4 + 13$ and $6 + 11 = 2 + 15$). The list of all 60 equations presented in the recognition phase, categorized by foil type, is given in the Appendix.

Results and Discussion

In this section, we will first discuss the accuracy of recognition and latencies of responses for novices, and then for experts. For each group, we will first examine overall accuracy of response to the foils (i.e., correct acceptance of Old targets and correct rejection of all foils). We will then compare participants' "Old" responses and latencies across the foil types. Note that for all foils except F+/R+, we compared latencies for correct responses only. Because false alarms were the dominant response for both experts and novices for the F+/R+ foil (both $\chi^2(1) > 36$, both $ps < .001$), latencies for incorrect responses for this foil were used in the analyses.

Novices exhibited high overall accuracy for most of the foils, exhibiting significantly greater than chance acceptance of Old targets and rejection of F-/P+, F-/P-, and F+/P- foils (all $t_s(22) > 2.7$, $ps < .02$). They gave mostly false alarms, however, on F+/P+ foils, as indicated by less than chance level accuracy, $t(22) = -2.6$, $p < .02$. This latter finding supports the assumption that participants relied upon categorical memory, rather than memory for specific items, for recognition: F+/P+ foils were categorically indistinguishable from Old targets, since both surface features and principled features present in Old targets were also present in F+/P+ foils. These

results indicate that these participants took the task seriously and were providing rather systematic responses.

Percentages of "Old" responses for novices are presented in the left-hand side panel of Figure 3. A one-way repeated measures ANOVA points to significant differences among foils for novices ($F(4, 88) = 53.9$, $MSE = 5.4$, $p < .001$). Paired-samples post-hoc t-tests with Bonferroni adjustments indicated that foils where both surface and principled features were present (i.e., Old targets and F+/P+) generated a greater number of "Old" responses than foils where either of the features was absent (i.e., F+/P-, F-/P+, and F-/P-), all $t_s(22) > 3$, all $p_s < .05$ for differences. More specifically, Old targets ($M = 0.84$, $SD = 0.15$) and F+/P+ foils ($M = 0.65$, $SD = 0.26$) generated "Old" responses significantly above chance, one-sample $t_s(22) > 2.6$, $p_s < .02$. At the same time, F+/P- ($M = 0.30$, $SD = 0.35$), F-/P- ($M = 0.07$, $SD = 0.20$), F-/P+ ($M = 0.03$, $SD = 0.16$) foils generated "Old" responses significantly below chance, one-sample $t_s(22) < -2.7$, $p_s < .02$. In short, Old targets and F+/P+ foils generated mostly "Old" responses, whereas all the other foils generated mostly "New" responses. This pattern of responses is consistent with predictions of Model 2 for novices (see Table 1), indicating that novices encoded both surface and principled features of arithmetic problems.

The right-hand side panel of Figure 3 presents novices' latencies to correct responses across F-/P- ($M = 1125$ ms, $SD = 549$ ms), F-/P+ ($M = 1201$ ms, $SD = 894$ ms), F+/P- ($M = 3653$ ms, $SD = 3372$ ms) foils, and Old targets ($M = 2375$ ms, $SD = 1955$ ms), and incorrect responses for F+/P+ foils ($M = 2283$ ms, $SD = 907$ ms). These measures were also subjected to a one-way repeated measures ANOVA. This analysis indicates significant differences among the foils, $F(4, 72) = 44.65$, $p < .001$. Paired-samples post-hoc t-tests with Bonferroni adjustments yielded the following order among latencies: F+/P- > Old targets = F+/P+ > F-/P+ = F-/P-, all $p_s < .05$

for differences. This pattern of latencies (F+ foils being slower than F- foils) further support predictions of Model 2 for novices (see Table 1), indicating that novices access surface and principled features in a serial manner, with surface features being accessed first. However, one aspect of this pattern deserves special consideration. Note that F+/P- elicited the slowest responses. These increased latencies may be indicative of a feature-principle competition that takes place in the course of accessing principled features.

In general, these data allow us to rule out Model 1 for novices presented in Figure 1 -- novices did not base their responses solely on the presence or absence of surface features. When surface features were absent (F-/P- and F-/P+ foils), novices produced fast and accurate "New" responses; however, when surface features were present, novices did not always produce "Old" answers. Rather, novices' responses were mediated by the presence or absence of principled features. In particular, when both surface and principled features were present (Old targets and F+/P+ foils) novices generally responded "Old". These responses were slower than those for F-/P- and F-/P+ foils. Finally, when surface features were present but principles were absent (F+/P- foils), novices in general accurately rejected these foils, but responses for these correct rejections were significantly slower than responses to Old targets. These findings indicate that, when surface features are present, there might be a feature-principle response competition. In this case, the delay indicates that participants have to inhibit the salient surface feature to correctly reject the attractive foil that has the same surface feature Old Targets, but does not have the principled feature.

Similar to novices, experts exhibited high overall accuracy for most of the foils, with greater than chance acceptance of Old targets and rejection of F-/P+, F-/P-, and F+/P- foils, all $t_s(11) > 26.5$, $p_s < .001$. They too gave mostly false alarms on F+/P+ foils, as indicated by less than

chance level accuracy, $t(11) = -12.9$, $p < .001$, indicating that like novices, experts also relied on categorical memory of the study items.

Percentages of "Old" responses for experts are presented in the left-hand side panel of Figure 4. A one-way repeated measures ANOVA points to significant differences among foils for experts, $F(4, 44) = 768.4$, $MSE = .34$, $p < .001$. As for novices, paired-samples post-hoc t-tests with Bonferroni adjustments indicated that foils where both surface and principled features were present (i.e., Old targets and F+/P+) generated a greater number of "Old" responses from experts than foils where either of the features was absent (i.e., F+/P-, F-/P+, and F-/P-), all $t_s(11) > 23$, all $p_s < .001$ for differences. More specifically, Old targets ($M = 0.9$, $SD = 0.05$) and F+/P+ ($M = 0.85$, $SD = 0.09$) foils generated "Old" responses significantly above chance, one-sample $t_s(11) > 12.9$, $p_s < .001$. At the same time, F+/P- ($M = 0.04$, $SD = 0.06$), F-/P- ($M = 0.00$, $SD = 0.00$), and F-/P+ ($M = 0.02$, $SD = 0.04$) foils generated "Old" responses significantly below chance, one-sample $t_s(11) < -28.2$, $p_s < .001$. In short, Old targets and F+/P+ foils generated mostly "Old" responses, whereas all the other foils generated mostly "New" responses. This pattern of responses is consistent with predictions of both Models 2 and 3 for experts (see Table 1), indicating that experts encoded both surface and principled features of arithmetic problems.

The right-hand side panel of Figure 4 presents experts' latencies to correct responses across F-/P- ($M = 1380$ ms, $SD = 586$ ms), F-/P+ ($M = 1318$ ms, $SD = 489$ ms), F+/P- ($M = 2736$ ms, $SD = 1008$ ms), and Old targets ($M = 2677$ ms, $SD = 1489$ ms), and incorrect responses for F+/P+ foils ($M = 2726$ ms, $SD = 1307$ ms). These measures were also subjected to a one-way repeated measures ANOVA. The analysis indicates significant differences among the foils, $F(4, 44) = 18.7$, $p < .001$. Paired-samples post-hoc t-tests with Bonferroni adjustments yielded the following order among latencies: Old targets = F+/P+ = F+/P- > F-/P+ = F-/P-, all $p_s < .05$ for

differences. This pattern of latencies (F+ foils being slower than F- foils) supports predictions of Model 3 over Model 2 for experts (see Table 1), indicating that, like novices, experts also access surface and principled features in a serial manner, with surface features accessed first.

The analysis of patterns of "Old" responses and response latencies allows us to eliminate Model 1 of expert responses presented in Figure 2. According to this model, experts should have responded "New" when principles were absent, and responded "Old" when principles were present. However, the F-/P+ foils almost invariably generated "New" responses, thus eliminating Model 1. Similarly, the analysis of latencies affords the elimination of Model 2. Recall that according to this model, experts should have more rapidly answered "New" when the principle was absent than when the feature was absent. However, the observed findings are consistent with Model 3 and not with Model 2, given that F-/P+ foils were rejected faster than F+/P- foils. Therefore, results of Experiment 1 support Model 2 for novices (Figure 1) and Model 3 for experts (Figure 2).

Experiment 1 establishes that novices and experts encode both surface and principled features of the problems. The results also indicate that in the course of recognition, surface and principled features are accessed in a serial manner. Finally, novices, but not experts, were found to exhibit a feature-principle response competition.

However, Experiment 1 leaves an important question unanswered. While this experiment elucidates the order of feature access, it does not reveal the order of feature encoding. The order of access might be indicative of the order of encoding, but it provides only indirect evidence. Therefore, although results of Experiment 1 suggest that both experts and novices encode surface features first, we deemed it necessary to generate more direct evidence. One way of generating such evidence is to drastically reduce the exposure time in the study phase. If under these

conditions participants will encode only one kind of feature, this would represent strong evidence that encoding is a serial process and that this feature is encoded first. For example, if novices encode surface features prior to principled features, then a shorter encoding time should result in performance depicted in Model 1 for novices (see Figure 1), such that all F+ foils should be accepted, whereas all F- foils should be rejected. To provide a direct examination of the issue of encoding, we conducted Experiment 2.

EXPERIMENT 2

The current experiment has the following goals. First, we deemed it necessary to determine whether participants encode features in a parallel or serial manner, and if the latter, to determine the order in which features are encoded. Another goal was to further examine the feature-principle response competition in novices that was found in Experiment 1.

Experiment 2 used the same recognition procedure as Experiment 1, with the exception that the length of presentation of stimuli in the study phase was reduced from 10 seconds to 1 second. If both types of features are encoded in parallel, then the patterns of accuracy and latency among all foils should be identical to those in Experiment 1. If features are encoded in a serial manner, however, these patterns could be different. Specifically, if participants encode surface features first, then (when encoding time is limited) they should respond "New" to the foils in which surface features are absent, while responding "Old" when surface features are present, regardless of the presence and absence of principles. Therefore, the proportion of "Old" responses for the foil in which surface features are present but the principles are absent (F+/P-) should increase comparatively to the level of "Old" responses observed in Experiment 1. If participants encode principled properties first, however, then the proportion of "Old" responses for the foil in which

surface features are absent but a principle is present (F-/P+) should increase comparatively to the proportions observed in Experiment 1.

The 1-second presentation time was established in a pilot experiment whose goal was to ascertain that participants could encode all elements of the equation in this limited presentation time, which is a necessary (though of course not sufficient) condition to encoding the principled features. For this pilot study, 12 undergraduates were presented in the study phase with the same 30 study equations as in Experiment 1, with each equation presented on screen for 1 second. In the recognition phase (following the distraction phase), participants were presented with 56 equations. Twelve of these equations consisted of Old Targets, while another 12 were F+/P+ foils. The remaining 32 equations were all F-/P- foils. Of these 32 foils, half violated the surface feature of number of elements (i.e., used 4 or 7 numbers in the equation), while the other half used the same number of elements as in the study equations. Additionally, all of these foils violated the surface feature of number (i.e., used a number greater than 9), with 8 of these foils violating this feature in only the first term of the equation, 8 violating this feature in only the term directly preceding the equals sign, 8 violating this feature in only the term directly following the equals sign, and 8 of these foils violating this feature in only the last term of the equation. If in the limited encoding time participants were unable to encode the entire equation, then the degree to which these different E-/R- foils are judged as “New” should differ. For example, if participants are only able to encode the first term, then they should be more likely to judge as “New” foils in which surface features are violated in the first term of the equation than for foils in which the surface features are violated later in the equation. The results of this pilot experiment indicated that there was in fact no significant differences among these E-/R- foils (all $t_s < 2.6$, all Bonferroni adjusted $p_s > .7$), and that the percentage of “New” judgments for all E-

/R- foils were significantly above chance (the means ranged from 81.3% to 95.8%, all $t_s > 4.1$, all Bonferroni adjusted $p_s < .02$). This pilot thus establishes that a 1-second presentation time was sufficient for participants to process all elementary terms of the study phase equations.

Method

Participants

Two samples, again representing novices and experts, were used in this study. The novice group included 25 undergraduates in an introductory psychology course at a large Midwestern university who participated for partial course credit. This sample had an average age of 20.4 years ($SD = 2.1$ years), with 13 women and 12 men. The expert group included 11 graduate students in the Mathematics Department at the same university who participated for a payment of twenty dollars. This sample had an average age of 27.3 years ($SD = 3.1$ years), with 2 women and 9 men.

Materials and Procedure

The materials and procedures for this experiment were identical to those of Experiment 1, with one critical exception. In the current experiment, each of the 30 equations in the study phase was presented on the screen for 1 second, as opposed to 10 seconds as in the previous experiment.

Results and Discussion

As for our discussion of results for Experiment 1, we will first discuss the accuracy of recognition and latencies of responses for novices, and then for experts. We will then compare participants' "Old" responses and latencies across the foil types. For E- foils and Old targets, we compare latencies for correct responses only. Because false alarms were the dominant response for novices for both F+/P- and F+/P+ foils (both $\chi^2(1) > 6.2$, $p_s < .02$), latencies for incorrect

responses for these foils were used in the analyses for novices. Also, because false alarms were the dominant response for experts for F+/P+ foils ($\chi^2(1) = 120.0, p < .001$), latencies for incorrect responses for this foil were used in the analyses for experts.

A one-way repeated measures ANOVA points to significant differences in novices' recognition accuracy of different foils, $F(4, 96) = 56.1, MSE = .05, p < .001$. Paired-samples post-hoc comparisons using Bonferroni adjustments indicate the following direction of accuracy among the foils: F-/P- = F-/P+ = Old targets > F+/P- = F+/P+ foils, all $ps < .002$ for differences. Thus, in contrast to findings from Experiment 1, where accuracy for F+/P- foils was equivalent to that for Old targets and significantly greater than for F+/P+ foils, in the current study accuracy for F+/P- foils dropped below that of Old targets and was equivalent to that for F+/P+ foils. This points to the fact that with the decreased encoding time, novices were less likely to encode principled properties and to use these properties in their recognition judgments.

The percentages of novices' "Old" responses across foil types are presented in the left-hand side panel of Figure 5. These results indicate that the proportion of "Old" responses for F-/P- ($M = 0.15, SD = 0.27$) and F-/P+ ($M = 0.06, SD = 0.11$) foils were significantly below chance (both $t_s(24) < -6.6, ps < .001$), that the proportion of "Old" responses for F+/P+ foils ($M = 0.80, SD = 0.17$) and Old targets ($M = 0.84, SD = 0.14$) were significantly above chance (both $t_s(24) > 8.8, ps < .001$), and that the proportion of "Old" responses for F+/P- foils ($M = 0.57, SD = 0.34$) was not significantly different from chance, $t(24) < 1$. These findings point to a number of important phenomena. First, novices were highly accurate at judging foils as "New" when surface features were absent (F-/P+ and F-/P- foils), indicating that even with a one-second exposure time, they were able to encode surface features. Second, when surface features were present but principles were absent (F+/P- foils), the proportion of incorrect "Old" responses increased compared to

Experiment 1 (30% in Experiment 1 vs. 57% in Experiment 2, $t(46) = 3.0$, $p < .01$). While in Experiment 1 participants were likely to correctly reject F+/P- foil as "New," in Experiment 2 they were equally likely to accept this foil as "Old." In short, novices tended to reject F- foils and to accept F+ foils, thus exhibiting processing presented in Model 1 for novices (see Figure 1). Taken together, these results indicate that with a one-second presentation of stimuli in the study phase, novices were able to encode the surface features, but were less likely to encode the principled properties. Therefore, it seems likely that (a) novices encode features in a serial manner and (b) surface features are encoded first.

The right-hand side of Figure 5 presents novices' latencies to correct responses across F-/P+ ($M = 1086$ ms, $SD = 462$ ms), F-/P- ($M = 1118$ ms, $SD = 507$ ms), and Old targets ($M = 1477$ ms, $SD = 437$ ms), and incorrect responses for F+/P- ($M = 1643$ ms, $SD = 935$ ms) and F+/P+ foils ($M = 1635$ ms, $SD = 566$ ms). A one-way repeated measures ANOVA indicates significant differences among the foils, $F(4, 84) = 14.0$, $p < .001$. Paired-samples post-hoc t-tests with Bonferroni adjustments indicated that latencies for F+ foils and Old Targets were significantly greater than for F- foils, all $t_s > 3.4$, $p_s < .05$, but there was no differences among latencies for F+ foils and Old Targets (all $t_s < 2.4$, $p_s > .25$). A planned comparison between latencies for correct responses to F+/P- foils and Old targets was conducted to determine whether response-competition persisted for these foils when novices did attend to the absence of principles. This test yielded that when novices correctly answered "Old" to F+/P- foils, their responses were significantly slower than for correct answers for Old targets, $t(22) = 3.74$, $p < .01$. This finding provides additional support for the feature-principle response competition. As found in Experiment 1, it appears that correct rejections of F+/P- foil required participants to inhibit a salient surface feature and an attractive "Old" response to the foil.

While novices' patterns of responses for Experiment 2 differed markedly from those in Experiment 1, for experts the patterns were similar across the two experiments. A one-way repeated measures ANOVA points to significant differences in accuracy among foils, $F(4, 40) = 12.2$, $MSE = .05$, $p < .001$. Paired-samples post-hoc comparisons using Bonferroni adjustments indicate the same patterns of accuracy among the foils as in Experiment 1: $F-/P- = F-/P+ = \text{Old targets} = F+/P- > F+/P+$ foils, all $ps < .05$ for differences.

The percentages of experts' "Old" responses across foil types are presented in the left-hand side panel of Figure 6. These results indicate that the proportion of "Old" responses for $F-/P-$ ($M = 0.12$, $SD = 0.22$), $F-/P+$ ($M = 0.08$, $SD = 0.12$), and $F+/P-$ foils ($M = 0.23$, $SD = 0.31$) were significantly below chance (all $ts(10) < -2.8$, all $ps < .05$), and that the proportion of "Old" responses for Old targets ($M = 0.80$, $SD = 0.21$) and $F+/P+$ foils ($M = 0.68$, $SD = 0.24$) were significantly above chance (both $ts(10) > 2.5$, $ps < .05$). These results indicate that, unlike novices, experts were able to encode both surface and principled features of problems even when the exposure time was reduced to one second. These findings by themselves, however, do not provide a conclusive answer as to whether experts encode the two types of features in parallel, or if they encode them in a serial manner, with both features being encoded very quickly. This problem should be addressed in future experiments that use even shorter encoding times.

The right-hand side of Figure 6 presents experts' latencies to correct responses across $F-/P+$ ($M = 1386$ ms, $SD = 483$ ms), $F-/P-$ ($M = 1497$ ms, $SD = 703$ ms), $F+/P-$ ($M = 2850$ ms, $SD = 1113$ ms), and Old targets ($M = 2445$ ms, $SD = 1433$ ms), and incorrect responses for $F+/P+$ foils ($M = 2138$ ms, $SD = 983$ ms). A one-way repeated measures ANOVA indicates significant differences among the foils, $F(4, 40) = 6.2$, $p < .001$. Paired-samples post-hoc t-tests with Bonferroni adjustments yielded that the patterns of differences for latencies among foils in the

current experiment are qualitatively identical to those in Experiment 1. Again, in contrast to novices, F+/P- latencies for correct responses for experts were not significantly different from those for Old targets, $t(10) < 1$. Additionally, the latencies of F-/P- and F-/P+ foils were again significantly lower than those for F+/P- and F+/P+ foils, all $t_s > 3.9$, all $p_s < .05$. The results of the current experiment, then, corroborate those in Experiment 1 in supporting Model 3 (Figure 2) as the most accurate model of expert processing of features.

The results of Experiment 2 indicate that for novices, feature encoding is a serial process, with surface features being encoded prior to deep relational properties. The results also corroborate findings of Experiment 1, indicating that accurate rejection of F+/P- items takes novices significantly more time to answer than any other item. Therefore, those novices who encoded both surface features and principled properties during the one-second presentation time exhibited the same pattern of competition between the "Old" and "New" responses as participants in Experiment 1. In both experiments, to produce the correct "New" response to F+/P- foils, it appears that novices needed to suppress the erroneous "Old" response. This competition between the two responses ("Old" vs. "New") points to a difficulty for participants to suppress the salient surface feature and reject the item.

Unlike for novices, the performance of experts in Experiment 2 does not lend clear support for the manner in which the two types of features are encoded. The fact that accuracy was high for F-/P-, F-/P+, and F+/P- foils indicate that even with a greatly reduced exposure time for problems, experts were capable of encoding both surface and principled features. The fact that experts access features in a serial manner, as indicated in Experiment 1, represents suggestive evidence that they may encode features in the same serial manner. However, results of the present experiment do not rule out the possibility of parallel feature encoding in experts.

Therefore, further study is needed to more conclusively determine the manner in which experts encode these features. Regardless, the expert data from Experiment 2 clearly corroborate the findings of Experiment 1 in supporting Model 3 as the most predictive of expert processing of problem features.

Having established the order of property encoding, we deemed it important to replicate these findings using a variation of the recognition procedure. This new variation, which is incorporated in Experiment 3, involves limiting the amount of stimuli presentation time during the recognition phase. When recognition items are presented, a participant must (a) encode the recognition item, and (b) compare this item to an accessed representation of properties of study items. Thus, by reducing the amount of time during recognition, we can better understand both encoding and accessing of elementary and relational properties.

EXPERIMENT 3

Experiment 3 used the same recognition procedure as Experiments 1 and 2, with the exception that the length of presentation of stimuli in the recognition phase was reduced from an unlimited amount of time (i.e., until the participant indicated their choice of the item as “Old” or “New”) to 1 second.

Method

Participants

Only a novice sample was used in this study. This novice group included 25 undergraduates in an introductory psychology course at a large Midwestern university who participated for partial course credit. This sample had an average age of 19.3 years ($SD = 0.8$ years), with 16 women and 9 men.

Materials and Procedure

The materials and procedures for this experiment were identical to those of Experiment 1, with one critical exception. In the current experiment, each of the 60 equations in the recognition phase was presented on the screen for 1 second, rather than until the participant made their response. Prior to the recognition phase, participants were instructed that each item would be presented on screen for a brief time, and after the item disappeared, they should indicate whether the item had appeared or had not appeared earlier in the experiment by pressing a corresponding key on the keyboard. Each item in the recognition phase was presented on screen for 1 second, and then replaced on the screen by the following instruction: “Press Z if the item was presented before, or press M if the item was not presented before.”

Results and Discussion

In this section, we will first discuss the accuracy of participants’ responses across foils. We will then compare participants’ “Old” responses and latencies across the foil types. For E- foils and Old targets, we compare latencies for correct responses only. Because false alarms were the dominant response for F+/P+ foils ($\chi^2(1) = 43.6, p < .001$), latencies for incorrect responses for these foils were used in the analyses. Also, because both “Old” and “New” responses were equally likely for F+/P- foils ($\chi^2(1) = .34, p > .5$), latencies for all responses for these foils were used in the analyses.

A one-way repeated measures ANOVA indicates significant differences in accuracy among foils, $F(4, 96) = 71.2, \text{MSE} = .03, p < .001$. Paired-samples post-hoc comparisons using Bonferroni adjustments indicate the following direction of accuracy among the foils: F-/P- = F-/P+ > Old targets > F+/P- = F+/P+ foils, all $ps < .005$ for differences. Thus, in contrast to findings from Experiment 1, where accuracy for F+/P- foils was equivalent to that for Old

targets, in the current study (as in Experiment 2) accuracy for F+/P- foils dropped below that of Old targets. This points to the fact that with the decreased presentation time for items in the recognition phase, participants were less likely to access principled properties when performing recognition judgments.

The percentages of participants' "Old" responses across foil types are presented in the left-hand side panel of Figure 7. As in Experiment 2, these results indicate that the proportion of "Old" responses for F-/P- ($\underline{M} = 0.06$, $\underline{SD} = 0.09$) and F-/P+ ($\underline{M} = 0.04$, $\underline{SD} = 0.06$) foils were significantly below chance (both $t_s(24) < -24.0$, $p_s < .001$), that the proportion of "Old" responses for F+/P+ foils ($\underline{M} = 0.69$, $\underline{SD} = 0.22$) and Old targets ($\underline{M} = 0.75$, $\underline{SD} = 0.14$) were significantly above chance (both $t_s(24) > 4.4$, $p_s < .001$), and that the proportion of "Old" responses for F+/P- foils ($\underline{M} = 0.52$, $\underline{SD} = 0.25$) was not significantly different from chance, $t(24) < 1$. These findings point to a number of important phenomena. First, novices were highly accurate at judging foils as "New" when surface features were absent (F-/P+ and F-/P- foils), indicating that even with a one-second exposure time in the recognition phase, they were able to access surface features. Second, when surface features were present but principles were absent (F+/P- foils), the proportion of incorrect "Old" responses increased in comparison to Experiment 1 (30% in Experiment 1 vs. 52% in Experiment 3, $t(46) = 2.43$, $p < .02$). Whereas in Experiment 1 novices were likely to correctly reject the F+/P- foil as "New," in Experiment 3 (as in Experiment 2) they were equally likely to accept this foil as "Old." These findings indicate that when stimuli exposure in recognition phase was reduced to one second, novices tended to reject F- foils and to accept F+ foils, thus exhibiting processing presented in Model 1 for novices (see Figure 1). In short, these results indicate that with a one-second presentation of stimuli in the recognition phase, novices were able to retrieve the surface features, but were less likely to

retrieve the principled properties. Therefore, as suggested in Experiment 1, feature access for novices is similar to feature encoding, such that novices access features in a serial manner and surface features are accessed first.

The right-hand side of Figure 7 presents participants' latencies to correct responses across F-/P- ($M = 566$ ms, $SD = 291$ ms), F-/P+ ($M = 595$ ms, $SD = 324$ ms), and Old targets ($M = 817$ ms, $SD = 368$ ms), incorrect responses for F+/P+ foils ($M = 911$ ms, $SD = 432$ ms), and all responses for F+/P- foils ($M = 1002$ ms, $SD = 560$ ms). A one-way repeated measures ANOVA indicates significant differences among the foils, $F(4, 96) = 18.3$, $p < .001$. The patterns of differences among foils in the current experiment are qualitatively identical to those in Experiments 2. Paired-samples post-hoc t-tests with Bonferroni adjustments indicated that latencies for F+ foils and Old Targets were significantly greater than for F- foils, all $t_s > 4.2$, $p_s < .001$, but there was no differences among latencies for F+ foils and Old Targets (all $t_s < 2.5$, $p_s > .1$). A planned comparison between latencies for correct responses to F+/P- foils and Old targets was again conducted to determine whether response-competition persisted for these foils when novices did attend to the absence of principles. As in Experiment 2, this test yielded that when novices correctly answered "Old" to F+/P- foils, their responses were significantly slower than for correct answers for Old targets, $t(24) = 3.3$, $p < .01$. This finding provides further evidence for the surface feature-principal response competition. Similarly to Experiments 1 and 2, it seems that correct rejections of F+/P- foil required participants to inhibit a salient surface feature and an attractive "Old" response to the foil.

The results of Experiment 3 corroborated findings of Experiments 1 and 2, indicating that accurate rejection of F+/P- items takes significantly more time to answer than any other item. Therefore, those novices who encoded and accessed both surface features and principled

properties exhibited the same pattern of competition between responses as participants in the two previous experiments. In all experiments, to produce the correct "New" response to F+/P- foils, it appears that novices needed to suppress the erroneous "Old" response. This competition between the two responses ("Old" vs. "New") points to a difficulty for participants to suppress the salient surface feature and reject the item.

General Discussion

Results of the three experiments are as follows. Experiment 1 establishes that both novices and experts encode both surface and structural features of arithmetic problems, and that for both experts and novices, surface and structural features are accessed in a serial manner with surface features accessed first. In addition, novices, but not experts, were found to exhibit a feature-principle response competition. Experiment 2 indicates that novices encode surface features and principles in a serial manner, with surface features being encoded first. Experiment 2 also provided further evidence for the feature-principle response competition for novices that was found in Experiment 1. Experiment 3 provided additional evidence that features are encoded and accessed in a serial manner, with surface features being encoded and accessed first. Further, Experiment 3 once again found the feature-principle response competition for novices.

These findings point to important processing similarities in experts and novices. First, experts and novices encode both surface and principled features, exhibiting serial access to these features in the course of recognition. In addition, both experts and novices access surface features first.

However, there are important differences in processing between novices and experts. One difference is that the shorter encoding time used in Experiment 2 affected encoding for novices but not for experts. Novices, but not experts, were less likely to encode principles with this

shorter encoding time, as evidenced by the fact that accuracy for F+/P- foils fell to a chance level, while for experts, accuracy for this foil remained above chance. It thus appears that experts encode principles much faster than do novices. Another important difference is that while novices experience competition between salient surface features and less salient deep principles, experts do not exhibit such competition. For the majority of novices, well known deep principles of commutativity and associativity end up winning the competition; however, the competition takes time and effort. At the same time, experts represent both deep and surface features of the problem and do not experience such attentional competition. Therefore, processing in experts is compatible with Model 3 for experts (Figure 2), whereas processing in novices conform to a revised version of Model 2 for novices, as presented in Figure 8. The critical component of this revised model is the presence of response competition between salient surface features and less salient principled features. In the course of response competition, participants suppress an attractive "Old" response to F+/P- foil and reject this foil. However, it remains unclear whether or not they can suppress salient features under more demanding task conditions. In particular, it is possible that in more resource demanding tasks, such as categorization, reasoning, or problem solving, deep relational features may lose attentional competition to salient surface features for novices. This loss would manifest itself in novices' tendency to focus on surface features, while ignoring deep relational features (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Gentner & Toupin, 1986; Kotovsky & Gentner, 1996; Larkin, 1983; Simon & Simon, 1978; Yarlas & Sloutsky, 2000).

The findings presented in this paper have several potential implications. First, they lead to a better understanding of expertise, indicating that expert-novice differences persist even with simple tasks (although it is reasonable to expect that more complex tasks would result in more

dramatic expert-novice differences). Second, the results have important instructional implications, suggesting that salient surface features may prevent learners from focussing on less salient deep relational features, thus deterring the acquisition of these deep relational features in the course of learning. Finally, it seems that the task used in the reported experiments is both simple and informative as to processing differences, and therefore could be used for neuroscientific analyses of problem representations in experts and novices.

There are several issues that need to be addressed in future research. First, it is important to establish whether or not feature encoding in experts is a serial process. While feature access in experts suggests that the process is indeed serial, the reduction of exposure to one second did not result in a decline in accuracy of experts' responses to F+/P- or F-/P+ foils. Therefore, additional experiments with even shorter exposure times are needed to address this issue. Second, it seems important to establish whether or not the reported response competition is a property of this specific task, of only arithmetic tasks, or whether it is a property of any task within any domain of knowledge that includes salient surface features and less salient relational properties. It has been established that even in knowledge-lean tasks, relational features (e.g., monotonic increase) typically elicit more difficulties in both children and adults than non-relational object features (e.g., shapes or colors of constituent objects) (Gentner & Medina, 1998; Kotovsky & Gentner, 1996; Gentner & Toupin, 1986). Therefore, it seems plausible that the observed response competition might be a general property of problems that have both relational and object properties.

The reported findings indicate that even when a task is very simple, experts and novices construct problem representations differently. While both experts and novices encode deep as well as surface features of the problem, only for novices and not for experts, surface features

compete with deep features, thus requiring additional resources to inhibit this response competition. These findings may or may not hold for less familiar deep principles or more complicated tasks. However, these results allow us to conclude that even when a task is very simple and deep principles are familiar, experts and novices differ in processes underlying the construction of problem representations.

References

- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. Journal of Experimental Psychology: Learning, Memory, & Cognition, 10, 483-495.
- Bassok, M. (1996). Using content to interpret structure: Effects on analogical transfer. Current Directions in Psychological Science, 5, 54-58.
- Bassok, M. (1997). Two types of reliance on correlations between content and structure in reasoning about word problems. In L. D. English (Ed.), Mathematical reasoning: Analogies, metaphors, and images, (pp. 221-246). Mahwah: NJ: Lawrence Erlbaum Associates.
- Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. Journal of Experimental Psychology: Learning, Memory, & Cognition, 22, 792-810.
- Byrne, R. M. (1989). Suppressing valid inferences with conditionals. Cognition, 31(1), 61-83.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.
- Chi, M. T. H., Feltovich, P. G., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.
- Evans, J. St. B. T., Newstead, S. E.; Byrne, R. M. J. (1993). Human reasoning: The psychology of deduction. Hove, England: Lawrence Erlbaum Associates.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning, (pp. 199-241). New York: Cambridge University Press.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. Cognition, 65, 263-297.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. Cognitive Science, 10, 277-300.

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

Hinsley, D. A., Hayes, J. R. & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. In P. A. Carpenter & M. A. Just (Eds.), Cognitive processes in comprehension (pp. 89-106). Hillsdale, NJ: Erlbaum.

Holyoak, K. J.; Koh, K. (1987). Surface and structural similarity in analogical transfer. Memory & Cognition, 15, 332-340.

Johnson-Laird, P., & Byrne, R. (1991). Deduction. Hove, UK: Lawrence Erlbaum.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. Child Development, 67, 2797-2822.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. Cognitive Science, 4, 317-345.

Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & A. Stevens (Eds.), Mental models, (pp. 75-98). Hillsdale, NJ: Erlbaum.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. Journal of Experimental Psychology: Learning, Memory, & Cognition, 14, 510-520.

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

Reed, S. K., Ackinclose, C. C., & Voss, A. A. (1990). Selecting analogous problems: Similarity versus inclusiveness. Memory & Cognition, 18, 83-98.

Reeves, L., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. Psychological Bulletin, 115, 381-400.

Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. Journal of Experimental Psychology: Learning, Memory, & Cognition, 23, 427-440.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, & Cognition, 8, 484-494.

Siegler, R. S. (1978). The origins of scientific thinking. In R. S. Siegler, (Ed), Children's thinking: What develops? (pp. 109-149). Hillsdale, NJ: Lawrence Erlbaum Associates.

Silver, E. A. (1981). Recall of mathematical problem information: Solving related problems. Journal of Research in Mathematics Education, 24, 117-135.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler, (Ed), Children's thinking: What develops? (pp. 325-348). Hillsdale, NJ: Lawrence Erlbaum Associates.

Yarlas, A. S., & Sloutsky, V. M. (2000). Problem representation in experts and novices: Part 1. Differences in the content of representation. In L. R. Gleitman & A. K. Joshi (Eds.), Proceedings of the 22nd Annual Conference of the Cognitive Science Society (pp. 1006-1011). Mahwah, NJ: Erlbaum.

Author Note

This research was supported in part by a grant from James S. McDonnell Foundation. We would like to thank Brad Morris and Aaron Rader for their helpful comments on earlier versions of this manuscript. Correspondence concerning the manuscript should be addressed to Vladimir M. Sloutsky (sloutsky.1@osu.edu) or Aaron S. Yarlas (yarlas.1@osu.edu), 21 Page Hall, 1810 College Avenue, The Ohio State University, Columbus, OH 43210.

Table 1

Patterns of responses and latencies predicted by alternative models for novices and experts

| Foil Types and Patterns of Responses | | | | | |
|--|-------------|-------------------|-------|-------|-------|
| Components of Processing | Old targets | F+/P+ | F+/P- | F-/P+ | F-/P- |
| Encoding | | Response Patterns | | | |
| Features only (Novices, Model 1) | Old | Old | Old | New | New |
| Principles only (Experts, Model 1) | Old | Old | New | Old | New |
| Features and Principles (Novices, Model 2; Experts Models 2 and 3) | Old | Old | New | New | New |
| Access | | Latencies | | | |
| Features first (Novices, Model 2; Experts, Model 3) | Slow | Slow | Slow | Fast | Fast |
| Principles first (Experts Model 2) | Slow | Slow | Fast | Slow | Fast |
| Principles and Features in parallel | Fast | Fast | Fast | Fast | Fast |

Figure Captions

Figure 1. Two proposed possibilities of processing mechanisms underlying novices' recognition of problems when surface features and principled features are present or absent.

Figure 2. Three proposed possibilities of processing mechanisms underlying experts' recognition of problems when surface features and principled features are present or absent.

Figure 3. Proportion of novices' "Old" responses and response times (in milliseconds) across foil types in the recognition phase. Error bars represent Standard Errors of the Mean. F+/P- = Feature +/Principle -, F+/P+ = Feature +/Principle +, F-/P- = Feature -/Principle -, F-/P+ = Feature -/Principle +, and Old = Old targets, Experiment 1.

Figure 4. Proportion of experts' "Old" responses and response times (in milliseconds) across foil types in the recognition phase. Error bars represent Standard Errors of the Mean. F+/P- = Feature +/Principle -, F+/P+ = Feature +/Principle +, F-/P- = Feature -/Principle -, F-/P+ = Feature -/Principle +, and Old = Old targets, Experiment 1.

Figure 5. Proportion of novices' "Old" responses and response times (in milliseconds) across foil types in the recognition phase. Error bars represent Standard Errors of the Mean. F+/P- = Feature +/Principle -, F+/P+ = Feature +/Principle +, F-/P- = Feature -/Principle -, F-/P+ = Feature -/Principle +, and Old = Old targets, Experiment 2.

Figure 6. Proportion of experts' "Old" responses and response times (in milliseconds) across foil types in the recognition phase. Error bars represent Standard Errors of the Mean. F+/P- = Feature +/Principle -, F+/P+ = Feature +/Principle +, F-/P- = Feature -/Principle -, F-/P+ = Feature -/Principle +, and Old = Old targets, Experiment 2.

Figure 7. Proportion of novices' "Old" responses and response times (in milliseconds) across foil types in the recognition phase. Error bars represent Standard Errors of the Mean. F+/P- =

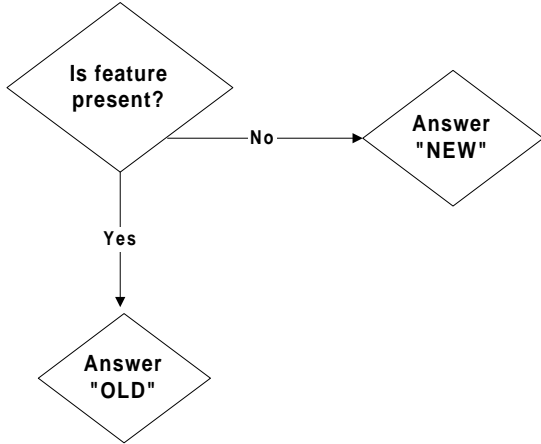
Feature +/Principle -, F+/P+ = Feature +/Principle +, F-/P- = Feature -/Principle -, F-/P+ =

Feature - /Principle +, and Old = Old targets, Experiment 3.

Figure 8. Revised model of problem processing for novices.

Figure 1

Novices: Model 1



Novices: Model 2

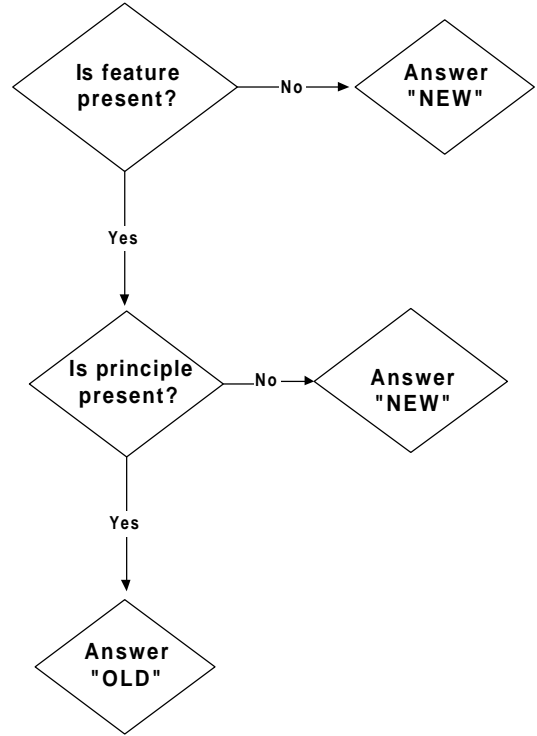
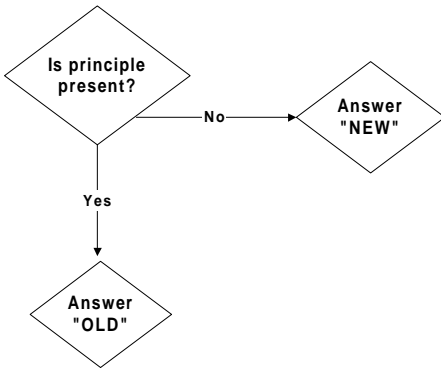
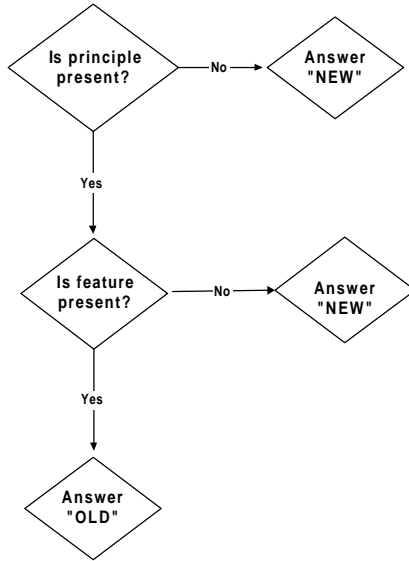


Figure 2

Experts: Model 1



Experts: Model 2



Experts: Model 3

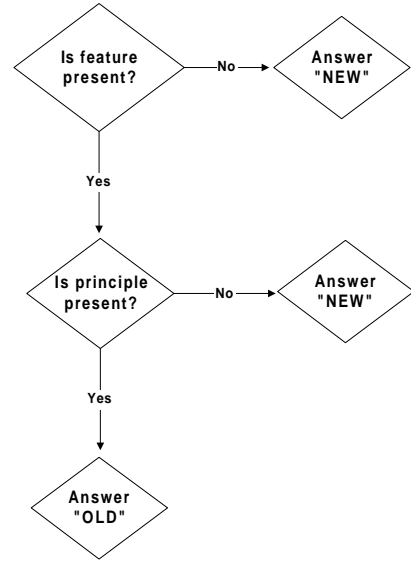


Figure 3

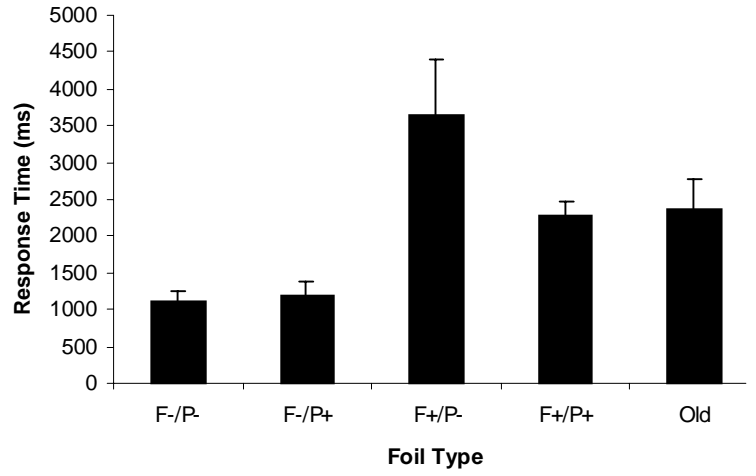
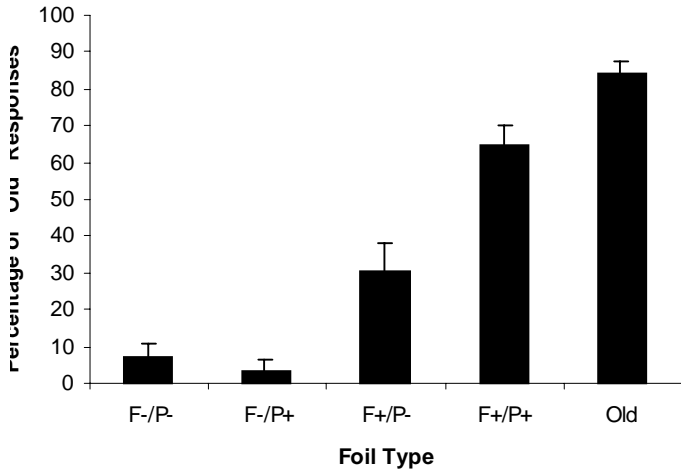


Figure 4

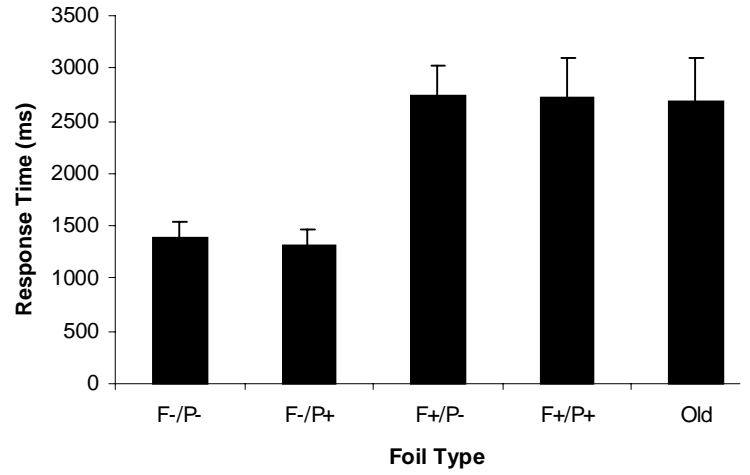
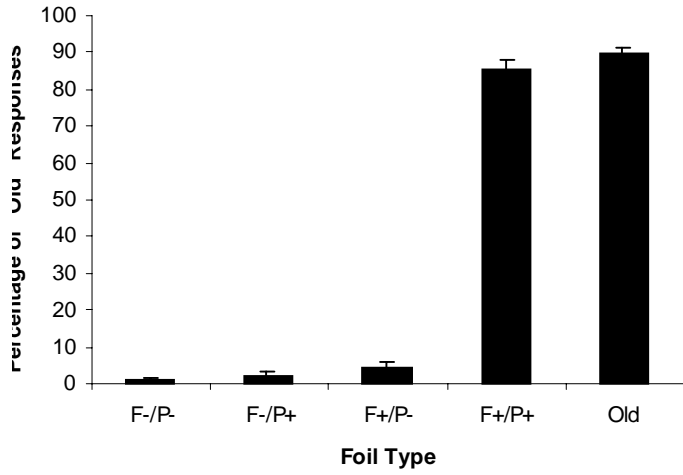


Figure 5

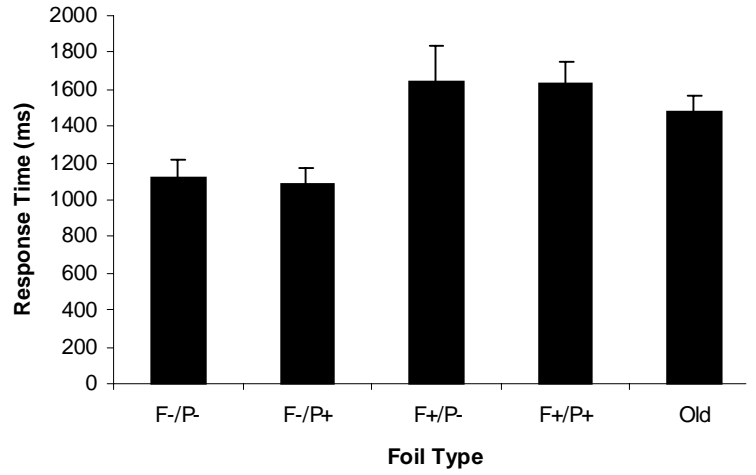
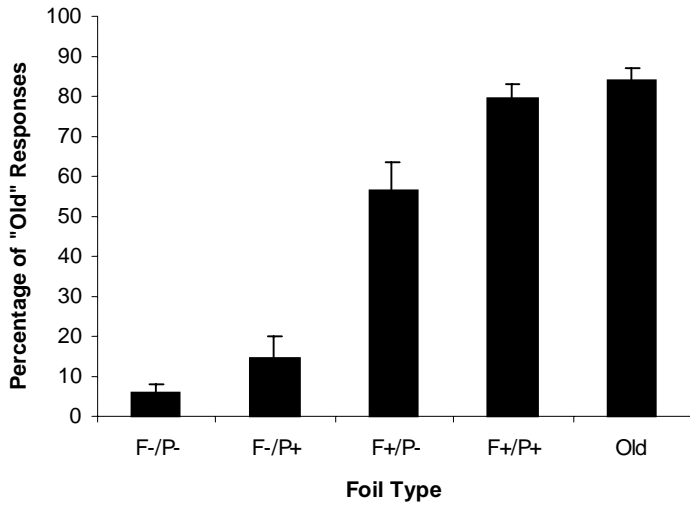


Figure 6

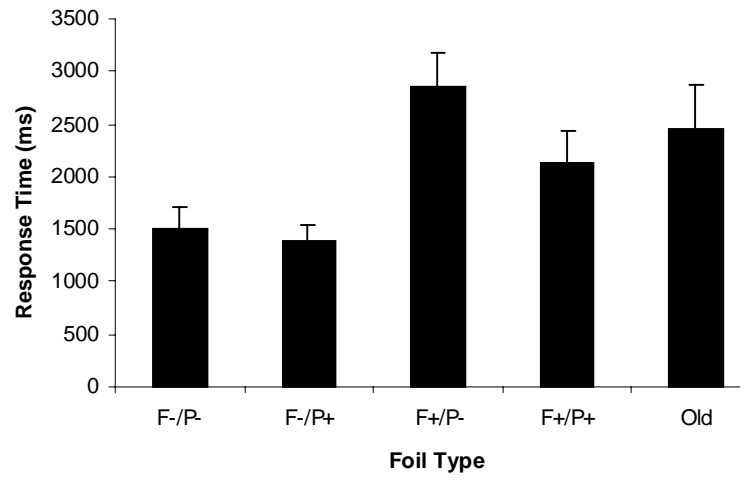
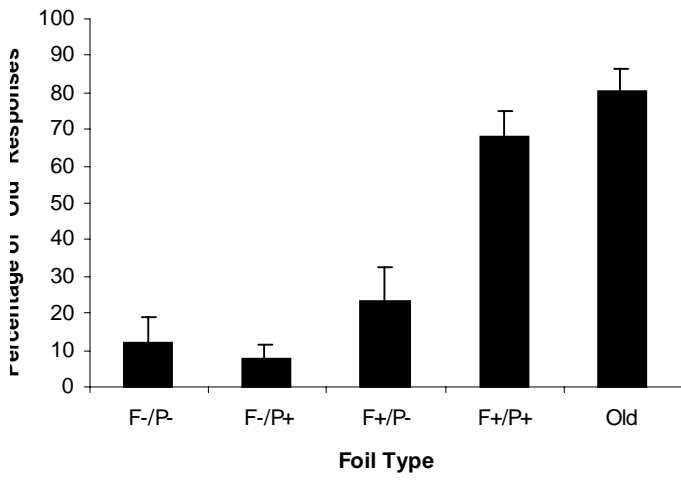


Figure 7

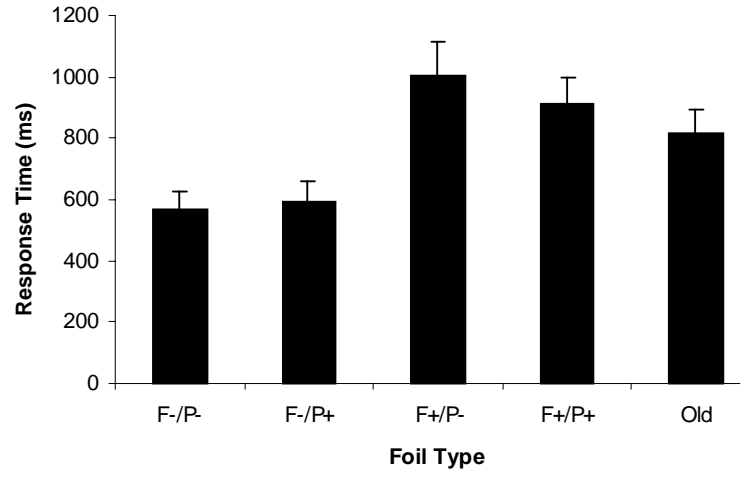
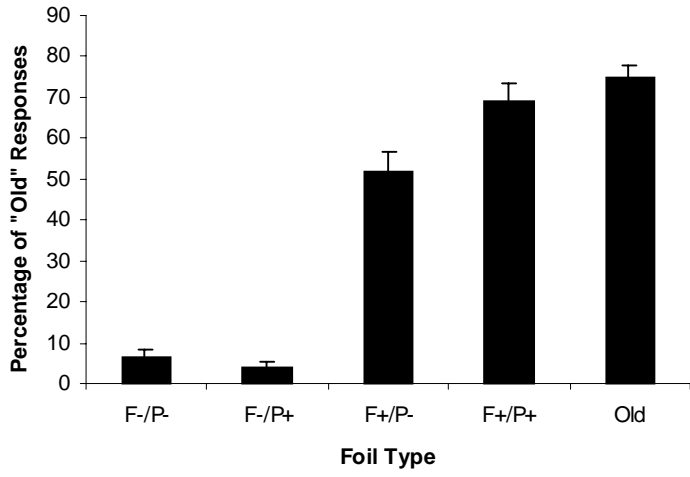
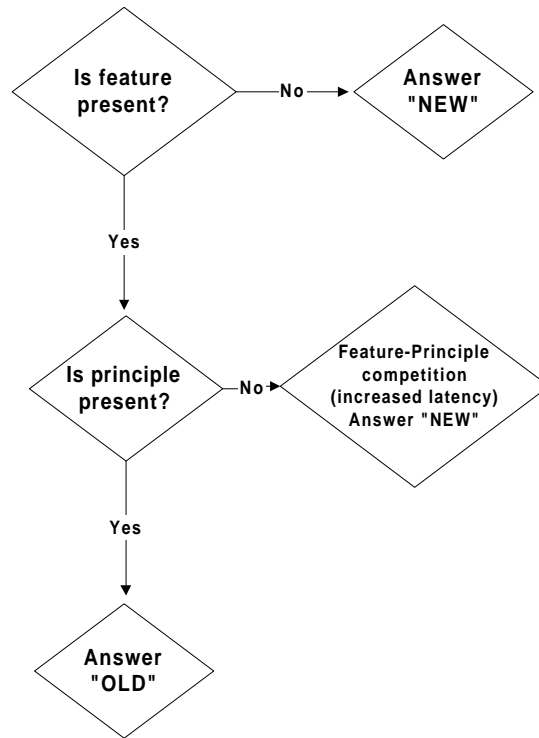


Figure 8



Appendix

Study Phase Problems (N = 30)

$$4 + 5 + 1 = 1 + 5 + 4$$

$$5 + 7 + 3 = 3 + 7 + 5$$

$$6 + 1 + 4 = 4 + 1 + 6$$

$$8 + 4 + 3 = 4 + 3 + 8$$

$$9 + 2 + 4 = 2 + 9 + 4$$

$$6 + 3 + 8 = 3 + 8 + 6$$

$$1 + 6 + 9 = 9 + 6 + 1$$

$$2 + 5 + 6 = 6 + 5 + 2$$

$$7 + 9 + 5 = 5 + 9 + 7$$

$$4 + 7 + 2 = 2 + 7 + 4$$

$$1 + 5 + 8 = 8 + 5 + 1$$

$$3 + 6 + 1 = 6 + 1 + 3$$

$$8 + 7 + 2 = 2 + 8 + 7$$

$$9 + 1 + 8 = 1 + 8 + 9$$

$$8 + 2 + 6 = 6 + 8 + 2$$

$$3 + 5 = 2 + 1 + 5$$

$$9 + 4 = 3 + 6 + 4$$

$$7 + 4 = 2 + 5 + 4$$

$$8 + 6 = 3 + 5 + 6$$

$$6 + 3 = 4 + 2 + 3$$

$$5 + 2 = 1 + 4 + 2$$

$$3 + 6 = 2 + 1 + 6$$

$$8 + 2 = 3 + 5 + 2$$

$$5 + 4 + 3 = 9 + 3$$

$$1 + 7 + 2 = 8 + 2$$

$$3 + 4 + 1 = 7 + 1$$

$$2 + 3 + 6 = 5 + 6$$

$$1 + 2 + 8 = 3 + 8$$

$$4 + 1 + 9 = 5 + 9$$

$$3 + 5 + 6 = 8 + 6$$

Old targets (N = 12)

$$4 + 5 + 1 = 1 + 5 + 4$$

$$8 + 4 + 3 = 4 + 3 + 8$$

$$1 + 6 + 9 = 9 + 6 + 1$$

$$4 + 7 + 2 = 2 + 7 + 4$$

$$8 + 7 + 2 = 2 + 8 + 7$$

$$9 + 1 + 8 = 1 + 8 + 9$$

$$3 + 5 = 2 + 1 + 5$$

$$8 + 6 = 3 + 5 + 6$$

$$3 + 6 = 2 + 1 + 6$$

$$5 + 4 + 3 = 9 + 3$$

$$2 + 3 + 6 = 5 + 6$$

$$3 + 5 + 6 = 8 + 6$$

F+/P+ Foils (N = 12)

$$6 + 8 + 3 = 3 + 8 + 6$$

$$5 + 1 + 2 = 2 + 1 + 5$$

$$7 + 1 + 4 = 1 + 7 + 4$$

$$9 + 3 + 8 = 3 + 8 + 9$$

$$2 + 5 + 1 = 1 + 5 + 2$$

$$7 + 8 + 5 = 5 + 8 + 7$$

$$9 + 7 = 5 + 4 + 7$$

$$7 + 1 = 3 + 4 + 1$$

$$8 + 2 = 3 + 5 + 2$$

$$2 + 7 + 4 = 9 + 4$$

$$3 + 1 + 2 = 4 + 2$$

$$1 + 5 + 9 = 6 + 9$$

F+/P- Foils (N = 12)

$$4 + 9 = 3 + 8 + 2$$

$$6 + 2 = 3 + 4 + 1$$

$$7 + 4 = 6 + 3 + 2$$

$$8 + 7 = 5 + 6 + 4$$

$$5 + 7 + 3 = 2 + 9 + 4$$

$$3 + 1 + 9 = 2 + 5 + 6$$

$$8 + 7 + 3 = 4 + 9 + 5$$

$$5 + 7 + 4 = 2 + 8 + 6$$

$$8 + 4 + 3 = 9 + 6$$

$$1 + 7 + 2 = 4 + 6$$

$$3 + 4 + 1 = 6 + 2$$

$$2 + 5 + 4 = 8 + 3$$

F-/P+ Foils (N = 12)

$$10 + 4 = 4 + 10$$

$$11 + 3 = 3 + 11$$

$$7 + 12 = 12 + 7$$

$$1 + 13 = 13 + 1$$

$$14 + 9 = 9 + 14$$

$$6 + 15 = 15 + 6$$

$$11 + 2 + 6 = 7 + 4 + 2 + 6$$

$$6 + 3 + 12 = 5 + 1 + 3 + 12$$

$$14 + 2 + 8 = 5 + 9 + 2 + 8$$

$$10 + 1 + 2 = 4 + 6 + 1 + 2$$

$$13 + 7 + 2 = 8 + 5 + 7 + 2$$

$$5 + 1 + 15 = 3 + 2 + 1 + 15$$

F-/P- Foils (N = 12)

$$5 + 10 = 11 + 4$$

$$6 + 11 = 2 + 15$$

$$12 + 7 = 13 + 6$$

$$13 + 4 = 5 + 12$$

$$14 + 1 = 10 + 5$$

$$2 + 15 = 3 + 14$$

$$11 + 5 + 9 = 7 + 4 + 6 + 8$$

$$8 + 3 + 12 = 5 + 1 + 4 + 13$$

$$6 + 14 + 3 = 4 + 7 + 10 + 2$$

$$8 + 2 + 10 = 2 + 9 + 4 + 5$$

$$13 + 6 + 4 = 2 + 3 + 11 + 7$$

$$5 + 7 + 15 = 8 + 2 + 4 + 13$$