If dancers ate their shoes: Inductive reasoning with factual and counterfactual premises

ROBERT J. STERNBERG and JOYCE GASTEL
Yale University, New Haven, Connecticut

This experiment addressed the effect of precue information, which may be either familiar or novel, and either relevant or irrelevant, on the solution of inductive reasoning problems. Sixty undergraduate students each completed 216 verbal inductive reasoning problems and five psychometric ability tests. The reasoning problems were equally divided among analogies, classifications, and series completions, with half of each kind of item presented in a standard, uncued format, and half presented with a precue. With respect to internal validation, it was found that for analogies and classifications, subjects take longer to process irrelevant than relevant information if the precue is familiar, but they take longer to process relevant than irrelevant information if the precue is novel. For series completions, this relation does not hold; rather, both novelty and irrelevance add time to the processing of information, with the time for irrelevance greater than that for novelty. The utility of precues for different tasks was explored, and it was found that familiar relevant precues facilitated solution of the more difficult kinds of items (classifications and series completions), but hampered solution of the easier, more automatically solved items (analogies). With respect to external validation, it was found that the nonentrenched induction tasks overlapped with psychometric tests in terms of abilities measured, that the abilities measured were fluid rather than crystallized, and that the precued (more nonentrenched) items were better measures of fluid abilities than were the uncued items.

In our everyday lives, we must frequently make inductive judgments. Most of these judgments are routine, requiring standard reasoning from familiar premises. Occasionally, however, we must assume that something is true about which we may have little or no information, or which may even be counterfactual. Suppose, for example, we view Nicaragua as having a socialist government interested only in self-determination. Does the standard U.S. government analogy between Nicaragua and Cuba still hold? Or suppose we believe that a leading clinical psychologist has three maladjusted children. Can this psychologist's clinical advice still be trusted?

An example of the need for such potentially counterfactual inductive reasoning occasionally arises in detective work, as shown in Isaac Asimov's (1979) short story "The Singing Bell." The question faced by the world-famous professor of extraterrestrial phenomena, Dr. Urth, is whether Louis Peyton was on the moon recently, where he allegedly murdered Albert Cornwell. Solving the problem requires Dr. Urth to reason about the effects of unfamiliar gravitational levels.

The motive for the murder would have been to obtain a large quantity of a rare and valuable life form called the Singing Bell. Peyton denies both the murder and having been on the moon. Dr. Urth asks the accused, Pey-

ton, whether he respects Singing Bells. Peyton replies that he does—too much to break one. At this point, Peyton gently strokes the Singing Bell he is holding. Dr. Urth suddenly commands Peyton to toss the Singing Bell to him. Peyton automatically tosses the Bell. It travels a short arc one-third the distance to Urth, curves downward, and shatters on the floor. Dr. Urth then concludes,

> Surely the matter is now obvious. The fact that Mr. Peyton could so egregiously misjudge the toss of an object he obviously valued so highly could mean only that his muscles are not yet readjusted to the pull of Earthly gravity. It is my professional opinion ... that the prisoner has, in the last few days, been away from Earth. He has either been in space or on some planetary object considerably smaller in size than the Earth—as, for example, the Moon. (Asimov, 1979, p. 237)

In the above story, Dr. Urth thinks in terms of what would follow if an unfamiliar situation were true. Such reasoning with unfamiliar information can be readily incorporated into inductive reasoning problems. For example, assume that dinosaurs are kinds of fruit juices. Given this premise, what word would come next in the following series: *substance, liquid, drink, fruit juice,* (1) *brontosaurus,* (2) *coffee,* (3) *vegetable,* (4) *glass?*

Our attempt to explore reasoning with different kinds of information is part of an ongoing program of research in which we explore the relations between task novelty and information processing in tasks requiring human intelligence. This program of research is motivated by the notion that the ability to deal with relative task novelty is a particularly crucial aspect of human intelligence (Sternberg, 1981, 1985; see also Raaheim, 1974). In other

words, tasks that present subjects with intermediate amounts of novelty may be good measures of intelligence.

This assertion is part of the triarchic theory of intelligence (Sternberg, 1985). According to this theory, intelligence comprises components of information processing that people employ in deciding how to adapt to, shape, or select their environment. Particularly relevant to the present experiment is the experiential subtheory, which asserts that intelligence is measured to the extent that a particular task is relatively novel (as in the present experiment) or requires automatization of information processing. In this experiment, we sought to test the novelty aspect of this subtheory through the use of counterfactual novelty in inductive reasoning problems.

In this experiment, we also explored the ability to sift relevant from irrelevant information. This ability, which we call selective encoding, is one of three abilities (the others are selective combination and selective comparison) that we have proposed as basic to insight (Sternberg & Davidson, 1982, 1983). This theory of insight is also part of the experiential subtheory of the triarchic theory of intelligence.

The scheme of the program of research is shown in Table 1. Our research on induction has proceeded through four overlapping phases. In the first phase, we and others attempted task decomposition to understand the information processing components underlying performance on familiar, IQ-test-like inductive reasoning items (Sternberg, 1977; Sternberg & Gardner, 1983; see also Mulholland, Pellegrino, & Glaser, 1980). Performance on tasks such as analogies, classifications, and series completions was decomposed into its elementary components of processing, enabling us to ascertain the latencies of such components, and the strategies into which they were combined. In a second phase, we investigated tasks that were, in some sense, at the opposite extreme: unfamiliar and un-IQ-test-like. In one kind of problem, the conceptual projection problem, subjects were required to predict the future state of an object, given incomplete information about its present state. Objects in these experiments were other-worldly. For example, the people of the planet Kyron could either be born young and die young, be born old and die old, be born young and die old, or be born old and die young (Sternberg, 1982; see also Tetewsky & Sternberg, 1986). In a second kind of problem, the insight problem, subjects had to solve what appeared to be

standard mathematics problems, but what were in fact insight problems that could not be solved by routine mathematical formulae (Davidson & Sternberg, 1984; Sternberg & Davidson, 1982). In a third phase, we investigated problems that were familiar but un-IQ-test-like: inductive predictions into the future (when will a bottle of milk spoil?) and postdictions into the past (when did the bottle of milk spoil?) (Kalmar & Sternberg, 1985). The present experiment represents work in a fourth phase of this research.

In the fourth phase, we have been investigating problems that are IQ-test-like, but unfamiliar. Such problems have the theoretical advantage that they may measure the ability to cope with relative novelty, which seems quite important to intelligence (Sternberg, 1985), and may also be practical for use both in laboratory experiments and on tests of human intelligence, broadly defined. These problems involve a mix of items, some requiring reasoning based on facts and others requiring reasoning based on counterfactual (novel) premises. In the first experiment in this series, Marr and Sternberg (1986) found that both gifted and nongifted students in grades 6, 7, and 8 gave significantly more attention to novel information than to familiar information in the test problems, which were analogies preceded by a cue that could be either novel (e.g., sparrows play hopscotch) or familiar (e.g., pistols are weapons), and either relevant or irrelevant to analogy solution. Gifted students, however, gave significantly less attention to irrelevant novel information than did nongifted students, but did not differ from the nongifted students in their attention to relevant novel information.

Marr and Sternberg's (1986) experiment addressed differences in outcomes of information processing of factual and counterfactual analogies for gifted and nongifted children, but did not examine the question of exactly what information processing is used, and it did not investigate information processing in other types of problems. Moreover, the work was limited to grade-school children. In the present experiment, we sought to confront head-on the question of how adults process information in solving factual and counterfactual analogies, classifications, and series problems. These three kinds of test items were not chosen arbitrarily: They are the three inductive item types most often used to measure general intelligence, and, particularly, so-called fluid intellectual abilities (see Cattell, 1971; Horn & Cattell, 1966). By investigating three related but distinct item types, we could investigate whether the results generalize or whether they are particular to each item type.

Table 1
Taxonomy of Induction Items in Four Phases of Induction Research

Tasks	Similarity to IQ-Test Items	
	IQ-Test-Like	Un-IQ-Test-Like
Familiar	Analogies Classifications Series Completions	Everyday Inductions
Unfamiliar	Novel Analogies Novel Classifications Novel Series Completions	Conceptual Projections Insight Problems

METHOD

Subjects

The subjects were 60 Yale undergraduates (17 males and 43 females) in an introductory psychology course. The subjects participated toward fulfillment of a course requirement.

Materials

Materials were of two basic types, experimental tasks and psychometric tests.

Experimental tasks. Three different experimental tasks were used: analogies, classifications, and series completions. In each task, half of the items were uncued and half were precued. The subjects saw each item in either precued or uncued form, but not both, with half the subjects seeing a given item in precued form and half in uncued form. Each task item was associated with a precue of one of four precue types—familiar relevant (FR), familiar irrelevant (FI), novel relevant (NR), or novel irrelevant (NI)—which appeared only in the precued condition. Items were not the same across precue conditions. Examples of each of these types of items for each task are shown in Table 2.

Each subject received a total of 216 inductive reasoning items. Those items that appeared in uncued form were simply preceded by a blank field. The subjects had as long as they wanted to look at the blank field, and then pressed a button that caused the induction item to appear. Those items that appeared in precued form were preceded by a premise that could be either familiar or novel, and either relevant or irrelevant. The subjects were first presented with the precue, and were given as long as they wished to read it. They then pressed a button, which resulted in the disappearance of the precue and the immediate appearance of the induction item.

Items were equally divided among verbal analogies, classifications, and series completions; equally divided again between uncued and precued; and equally divided again (for the precued items) among FR, FI, NR, and NI precues. An item was classified as either familiar, if its precue was factual and well-known, or novel, if its precue was counterfactual. The item was also classified as either relevant, if its precue gave information helpful to item solution, or irrelevant, if its precue was unhelpful. Irrelevant precues, although unhelpful, did contain information associatively related to the item stem, so that the precue could not be immediately recognized as irrelevant.

Psychometric tests. Five psychometric tests were administered to each subject: the Verbal Reasoning subtest of the Differential Aptitudes Test (Form T; Bennett, Seashore, & Wesman, 1973), the Cattell Culture Fair Test of g (Scale 3; Cattell & Cattell, 1963), the insight problems used by Sternberg and Davidson (1982), the Crossing Out A's subtest of the French Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963), and the Vocabulary subtest (V-3) of the French Kit.

Design

Dependent variables were response times and error rates for the individual items. Independent variables, all within-subjects, were test type (analogies, classifications, series completions), cuing (uncued, precued), and, for precued items, precue type (familiarity: familiar, novel; and relevance: relevant, irrelevant). Each item had one cue associated with it, which was either FR, FI, NR, or NI.

Procedure

The subjects first filled out an informed consent form. Instructions were then administered orally. The subjects received the experimental tasks, followed by the psychometric tests. The three experimental tasks (analogies, classifications, and series completions) were administered via an Apple IIe microcomputer, in counterbalanced order across subjects, and psychometric tests followed in a fixed order in a later session. The subjects were informed that on precued items, the precue should be accepted as true and should be applied, if appropriate, to the item that followed. The subjects were told to respond as quickly and as accurately as possible. Items within a task were blocked by precuing versus no precuing, and

Table 2
Sample Stimulus Task Items

Analogies

Familiar Relevant
Chalk is used for writing.
INK : PAPER :: CHALK : (a) WORD, *(b) BLACKBOARD, (c) ERASER, (d) CLASSROOM

Familiar Irrelevant
Beds have sheets.
BUILDING : DOME :: BED : *(a) CANOPY, (b) PILLOW, (c) HOTEL, (d) ROOM

Novel Relevant
The hand is the organ of hearing.
EYE : BLINDING :: HAND : (a) SMOOTH, (b) TOUCHING, *(c) DEAFENING, (d) WORN
(Uncued condition: DEAFENING is replaced by NUMBING)

Novel Irrelevant
Sparrows play hopscotch.
TROUT : SCALY :: SPARROW : (a) GRACEFUL, *(b) FEATHERY, (c) SMALL, (d) ENERGETIC

Classifications

Familiar Relevant
Many fish are edible.
TUNA, SARDINE, SOLE, FLOUNDER, (a) BAIT, (b) TADPOLE, (c) SEAFOOD, *(d) SALMON

Familiar Irrelevant
Soup is mainly water.
BOUILLON, EGG DROP, BROTH, MINESTRONE, (a) SALAD, *(b) CHOWDER, (c) LIQUID, (d) CRACKERS

Novel Relevant
Flowers grow underground.
TULIP, DAFFODIL, ROSE, DAISY, (a) DIRT, (b) PETAL, *(c) ONION, (d) FLOWERPOT
(Uncued condition: ONION is replaced by BUTTERCUP)

Novel Irrelevant
Water boils at room temperature.
FOG, STEAM, VAPOR, CLOUD, (a) PUDDLE, (b) ICE, *(c) MIST, (d) RAIN

Series Completions

Familiar Relevant
Some people are heavier than others.
SKINNY : SLIM : AVERAGE : PLUMP : (a) HUNGRY, (b) THIN, *(c) FAT, (d) ATHLETIC

Familiar Irrelevant
Feet are bigger than hands.
THUMB : INDEX FINGER : MIDDLE FINGER : RING FINGER : (a) WRIST, (b) TOE, *(c) PINKY, (d) HAND

Novel Relevant
Furniture is eaten at the end of a meal.
APPETIZER : SOUP : SALAD : MAIN COURSE : *(a) TABLE, (b) BREAD, (c) MENU, (d) ENTREE
(Uncued conditon: TABLE is replaced by DESSERT)

Novel Irrelevant
People always sleep standing up.
AWAKE : FATIGUED : DROWSY : ASLEEP : (a) ALERT, (b) TIRED, (c) CONFUSED, *(d) UNCONSCIOUS

Note—The correct answer is preceded by an asterisk in the table. Subjects were instructed to answer items as though the premise were true, assuming nothing else out of the ordinary other than the information in the premise.

each task consisted of four blocks of trials, with precued and un-cued blocks alternating. Within the precued condition, the four kinds of cues were mixed so that a subject could not know on a given trial what kind of cue to expect. The first two blocks of each task were each preceded by four practice items. On uncued items, the subject initially saw a blank screen; on precued items, the subject initially saw the precue. When the subject pressed RETURN, the precue (if any) disappeared, and the test item appeared. By press-ing one of four response buttons, the subject terminated the trial. The computer recorded response times and responses, and then presented the next item.

MODELS OF INFORMATION PROCESSING

How do subjects solve the kinds of factual and coun-terfactual induction items presented in this experiment? Such a question can be addressed by considering alterna-tive models of information processing. Our goal in this article is not to provide a relatively complete model of information processing for beginning-to-end solution of analogies, classifications, and series completions: such a task has been undertaken in an earlier article (Sternberg & Gardner, 1983; see also Sternberg, 1985). We have not included the necessary task manipulations to test the Sternberg-Gardner model on the present data. It is possi-ble, of course, that the task manipulations we included in the present study changed information processing so as to vitiate the applicability of our earlier model, although we have no reason to believe that this was the case. Our goal in this paper is to discuss those aspects of informa-tion processing that are distinctive to the solution of in-duction items with familiar or novel precues that may be either relevant or irrelevant to problem solution.

Three basic kinds of models could account for infor-mation processing during solution of the induction problems: the null model, additive models, and interac-tive models. All of the proposed models are serial. We recognize, of course, that actual processing may not be strictly serial, and hence our models may oversimplify the complex information processing subjects actually did; however, we believe that the models provide at least good approximations to possible subject strategies. We also as-sume in all but the null model that processing of novel information takes longer than processing of familiar in-formation, in that novel information involves temporary overriding of a subject's previous knowledge. As we will show, this assumption is amply supported by the data in Table 4. We consider each of these models in turn.

Null Model
Model 0. Model 0 would be the most plausible model of information processing if the precuing manipulation failed. According to this model, the precuing manipula-tion has no effect unless the precue is NR, because only in this condition is the correct answer changed by the in-formation in the precue. Hence, it is possible that the results would show a constant effect of the precue unless

there is actually some need of it to compute a new an-swer. This model predicts that NR items will be harder than all the others, and that the others—FR, FI, and NI—will be equal in difficulty:

$$FR = FI = NI < NR. \qquad (0)$$

Additive Models
Model 1: Relevant < Irrelevant. In Model 1, there are separate time charges for two elements of informa-tion processing: dealing with novelty and dealing with ir-relevant information. The motivating notion is that dur-ing item solution, subjects are delayed by having to take account of novel information, and also by having to deal with irrelevant information, which is recognized as irrele-vant only after a series of successive failed self-terminating relevance tests. Relevance is determined by checking whether the conceptual relation in the precue matches that in the item of the problem. Irrelevance takes longer than relevance because subjects discontinue the relevance tests as soon as they see that the precue information is rele-vant to solving the given problem. Furthermore, to the extent that relevant precues act as hints, they may actu-ally facilitate problem solution.

Model 1 can be subdivided into two submodels, 1a and 1b. In 1a, it is presumed that the time charge for novelty is greater than that for irrelevance. In 1b, it is presumed that the time charge for irrelevance is greater than that for novelty. Thus, the response-time and error-rate predic-tions of Models 1a and 1b for the four precue conditions—FR, FI, NR, and NI—are

$$FR < FI < NR < NI \qquad (1a)$$

and

$$FR < NR < FI < NI. \qquad (1b)$$

Model 2: Irrelevant < Relevant. Model 2 is like Model 1, except that there is an incremental time charge for *relevance* rather than for irrelevance. According to this model, novel information takes longer to process than familiar information, as in Model 1. But relevant infor-mation is presumed to take longer to process than irrele-vant information, because only relevant information needs to be incorporated into item solution. If a subject deter-mines that information is irrelevant, that information can be disregarded. If the information is relevant, however, then it is integrated into the knowledge representation used to solve the item, and this integration plus the subsequent use of the extra information adds a time charge to item processing, and also increases the probability of an er-ror, due to the addition of an extra step of information processing.

Model 2, like Model 1, can be subdivided into two sub-models. In Model 2a, it is presumed that the time charge for novelty is greater than that for relevance, whereas in Model 2b, it is presumed that the time charge for

relevance is greater than that for novelty. Thus, the response-time and error-rate predictions are

$$FI < FR < NI < NR \qquad (2a)$$

and

$$FI < NI < FR < NR. \qquad (2b)$$

Interactive Models

The interactive models differ from the additive models in their assumption that whether relevant or irrelevant information takes longer to process depends upon whether that information is familiar or novel.

Model 3: Relevant < Irrelevant for Familiar Only. In Model 3, as in Models 1 and 2, there is an incremental time charge for novelty. However, it is assumed that for familiar items, successive self-terminating searches result in irrelevant information being slower to incorporate than relevant information (as in Model 1). Moreover, because FR information does not change the keyed response on familiar items, such information can immediately be discarded without further processing. On NR items, however, relevant information does change the keyed answer, and hence must be incorporated into solution processing, resulting in an added time charge for relevant information in items with novel precues. Thus,

$$FR < FI < NI < NR. \qquad (3)$$

Model 4: Relevant < Irrelevant for Novel Only. In Model 4, as in the preceding models, there is an incremental time charge for novelty. However, it is assumed that on items with familiar precues, incorporating precue information into item solution causes these items to take longer to solve than items with familiar irrelevant precues (as in Model 2). On items with novel precues, it is assumed that irrelevant information takes longer to process, because it is more difficult to recognize the precue information as irrelevant. In other words, whereas the relevance of familiar information can be readily assessed, the relevance of information is hard to assess when the information is novel in the first place; hence, there is a delay for irrelevant information when processed in the context of a novel precue. Thus,

$$FI < FR < NR < NI. \qquad (4)$$

RESULTS

Reliability of Data

A fundamental, preliminary issue that needs to be addressed before subsequent data analysis is whether the data for the experimental task are reliable. The task is a rather unusual one, and it is possible that, as a result, the data will lack internal consistency, either with respect to items (each item measuring a different ability) or with respect to subjects (each subject employing a different strategy). Item coefficient-alpha reliabilities are equivalent to reliabilities for all possible split halves of items, whereas subject coefficient-alpha reliabilities are equivalent to reli-

abilities for all possible split halves of subjects. These internal-consistency reliabilities are shown in Table 3. All of the reliabilities are in the high .80s and low .90s, which are highly satisfactory for data-analytic purposes.

Correlations Between Response Times and Error Rates

A second fundamental and preliminary issue concerns correlations between response times and error rates. Across subjects, this correlation was .62 ($p < .001$), indicating that there was no speed–accuracy tradeoff, but rather that subjects who were faster were also more accurate. Across items, the correlation was .21 ($p < .001$), indicating that items that took longer to solve were also more susceptible to error. Thus, we need concern ourselves no further either with speed–accuracy tradeoff or with the possibility that errors tended to be "quick" ones due to subjects giving up without fully attempting item solution.

Basic Statistics

Basic statistics for the various experimental conditions are shown in Table 4. Means are shown for both correct response times and error rates as a function of item type (analogies, classifications, series), cuing condition (cued or uncued), and type of precue information (FR, FI, NR, NI). (Patterns of results are essentially identical if all response times, including those for errors, are analyzed.)

Table 5 shows an analysis of variance (ANOVA) upon these means. The results are clear-cut: All of the main effects and interactions are statistically significant for the response times, and all of the main effects and all but one of the interactions are statistically significant for error rates. These results indicate that the experimental manipulations affected performance, but, more importantly, that these effects were interactive, as would be predicted from either Model 3 or Model 4. For the main effects, series completions take longer to solve than classifications, which in turn take longer to solve than analogies. Items with novel precues take longer to solve than items with familiar precues, and irrelevant precues take longer than relevant ones. ANOVAs were also calculated for each task, and they were essentially the same as the overall ANOVAs.

Table 3
Reliabilities of Response Times

	Overall	Uncued	Cued
Analogies			
Item	.89	.89	.89
Subject	.89	.89	.89
Classifications			
Item	.87	.87	.88
Subject	.91	.91	.91
Series			
Item	.88	.85	.90
Subject	.92	.92	.92

Note—Only correct response times were used.

Table 4
Basic Statistics: Raw Scores

	Analogies		Classifications		Series		Overall	
	RT	ER	RT	ER	RT	ER	RT	ER
				Familiar Relevant				
Cued	3.91	2.4	3.81	2.6	5.15	5.4	4.29	3.5
Uncued	3.61	0.7	4.48	6.1	5.55	11.5	4.55	6.1
				Familiar Irrelevant				
Cued	4.18	3.5	4.26	4.1	6.22	5.7	3.89	4.4
Uncued	3.52	2.2	3.87	2.2	5.17	4.3	4.19	2.9
				Novel Relevant				
Cued	4.97	9.1	6.64	23.5	5.22	5.6	5.61	12.7
Uncued	3.82	3.7	4.26	6.9	5.01	4.8	4.36	5.1
				Novel Irrelevant				
Cued	4.32	6.9	4.42	8.1	6.65	13.0	5.13	9.3
Uncued	3.23	2.4	4.09	3.0	5.02	8.1	4.11	4.5

Note—Response times (RTs) are expressed in seconds, error rates (ERs) in percentages. Response times are for correct responses only.

The various main effects and interactions can be better understood in the context of Table 4, which includes mean response times and error rates on the various item types for each of the four conditions of precuing. Reaction times and error rates show identical patterns: Overall means, as well as the means for the analogy and classifications tasks, clearly support Model 3, the interactive model in which irrelevance adds to response times and error rates for items with familiar precues, but in which relevance adds to response times and error rates for items with novel precues. In the series completions, the response times and error rates for novel relevant items are relatively low; these means support Model 1b rather than Model 3. Thus, the individual task results and the overall means for response times and error rates support Model 3, except for the series completions task, the results of which support Model 1b.

Effect of Precues

So far in our models we have considered only simple scores. It is also informative, however, to look at the pattern of differences between cued and uncued scores to examine the effect of cues on problem solution (Table 6). The expected, obvious patterns occur clearly: the difference scores for items with irrelevant precues and/or novel precues are all positive; that is, both novel and irrelevant precued items have a higher response time and error rate than the same items uncued. It is most interesting, however, to look at the scores for items having familiar relevant precues, because solution of these items is sometimes hindered and sometimes facilitated by the presence of precues.

The familiar relevant precues make analogies more difficult, while facilitating solution of classifications and series completions. There are two possible reasons for this effect. The first reason relates to the difficulty of the type of item. As can readily be seen from Table 4, analogies are clearly the easiest uncued items. This is proba-

bly because solving analogies is a well-practiced skill in the population sampled (Yale undergraduates). The analogies in this experiment, which employ only common words and concepts, can be solved almost automatically by these subjects, whereas the classifications and series completions are less familiar and hence more difficult. The effect of a precue—even a familiar relevant one—on analogy solving is to cause an interruption of processes that would otherwise proceed automatically. Even if the precue gives helpful information, incorporating it expends more time than it saves. On the other hand, the precues for classifications and series completions are potentially more helpful because the subjects need more help on the more difficult items.

The second possible reason for the differential helpfulness of familiar relevant cues is that the more flexible the structure of an item, the more difficult it is to determine the relevance of a precue. Analogies are highly flexible in that there can be a wide variety of relations between terms. Consider, for example, the relation between bagel and doughnut, between hero and admiration, and between eye and blinding (all from test items). In classifications, the relations between terms are more constrained: all terms in an item are the same part of speech, and they are all related by group membership. In series problems, the relations between terms are even further restricted; in addition to satisfying the constraints mentioned above for classifications, the terms must also be arranged in a progression. Thus, in analogies it should be fairly difficult to determine whether or not a precue contains relevant information. In other words, nearly any precue that mentions any aspect of any of the analogy terms could conceivably be relevant to problem solution.

In classifications, because the terms must be related by group membership, it should be fairly easy to spot a relevant precue, and a relevant precue should be fairly helpful, in that it gives away the rule regarding to what group the terms belong. The precue in series problems is even

Table 5
Analysis of Variance

Factors	df	MS	F
Reaction Time (Correct Only)			
Test Type	2	302.1	106.42†
Test Type × Subject	118	2.8	
Cue	1	163.8	54.97†
Cue × Subject	59	3.0	
Novelty	1	36.7	53.84†
Novelty × Subject	59	0.7	
Relevance	1	5.0	10.88*
Relevance × Subject	59	0.5	
Test × Cue	2	1.2	1.15
Test × Cue × Subject	118	1.1	
Test × Novelty	2	18.5	28.29†
Test × Novelty × Subject	118	0.7	
Test × Relevance	2	39.7	51.10†
Test × Relevance × Subject	118	0.8	
Cue × Novelty	1	78.7	153.06†
Cue × Novelty × Subject	59	0.5	
Cue × Relevance	1	12.4	14.71†
Cue × Relevance × Subject	59	0.8	
Novelty × Relevance	1	18.3	25.50†
Novelty × Relevance × Subject	59	0.7	
Test × Novelty × Relevance	2	16.3	24.91†
Test × Novelty × Relevance × Subject	118	0.7	
Cue × Novelty × Relevance	1	28.9	42.44†
Cue × Novelty × Relevance × Subject	59	0.7	
Test × Cue × Relevance × Novelty	2	20.8	23.95†
Test × Cue × Relevance × Novelty × Subject	118	0.9	
Error Rate			
Test Type	2	.1760	23.75†
Test Type × Subject	118	.0074	
Cue	1	.2871	28.08†
Cue × Subject	59	.0102	
Novelty	1	.4899	89.59†
Novelty × Subject	59	.0055	
Relevance	1	.0879	16.69†
Relevance × Subject	59	.0053	
Test × Cue	2	.1416	10.40†
Test × Cue × Subject	118	.0068	
Test × Novelty	2	.0907	19.95†
Test × Novelty × Subject	118	.0045	
Test × Relevance	2	.1385	18.09†
Test × Relevance × Subject	118	.0076	
Cue × Novelty	1	.4109	44.01†
Cue × Novelty × Subject	59	.0093	
Cue × Relevance	1	.0046	.88
Cue × Relevance × Subject	59	.0052	
Novelty × Relevance	1	.0068	1.15
Novelty × Relevance × Subject	59	.0062	
Test × Cue × Novelty	2	.0586	8.62†
Test × Cue × Novelty × Subject	118	.0068	
Test × Cue × Relevance	2	.0635	6.70†
Test × Cue × Relevance × Subject	118	.0095	
Test × Novelty × Relevance	2	.2327	29.78†
Test × Novelty × Relevance × Subject	118	.0078	
Cue × Novelty × Relevance	1	.1121	17.29†
Cue × Novelty × Relevance × Subject	59	.0064	

*$p < .01$.   †$p < .001$.

easier to recognize as relevant and is more helpful, in that if it is relevant, it really "gives away" the answer, and if not, it is readily perceived as unhelpful, and can be discarded quickly.

According to the above rationale, the following would be expected: For analogies, both relevant and irrelevant cues are unhelpful and time-consuming, in that they mainly distract, whereas for classifications and series completions, relevant precues are helpful, and it is less difficult to recognize a precue as relevant. These expectations do describe the pattern of results obtained for items with familiar precues. As shown by both response times and

error rates, solution of analogies is impeded by precues, whereas solution of classifications and series completions is facilitated by precues.

Similar patterns hold for the novel precues for analogies and series completions; however, the results for novel classifications are anomalous. The reason seems to relate to the actual novel relevant cues used in the classifications. The most obvious type of novel relevant precue that could have been constructed for classification problems turned out to be trivially easy, such as,

> Celery is a fruit.
> apple, pear, peach, grapes,
> 1. orchard
> 2. produce
> 3. celery
> 4. jam

To avoid such items, which merely entail substitution rather than reasoning, we constructed the items as follows:

> Fruits have long green stalks.
> apple, pear, peach, grapes,
> 1. orchard
> 2. produce
> 3. celery
> 4. jam

Thus, in the novel precues for classifications (unlike those for analogies and series completions), the change induced by the precue is a change in the properties of the words in the stem, not in the word that completes the item. The effect of this extra inductive step was to make the novel relevant precues more difficult to reason with for classifications than for either analogies or series completions. Consequently, the novel relevant classifications have comparatively high reaction times and error rates.

In summary, the effect of precues seems to be as follows: Precues hinder item solution if the precue is novel and/or irrelevant, if the item is easy and automatized in the uncued state, if the structure of the item is flexible enough that it is difficult to determine whether or not the precue is relevant, or if the precue is nonobvious by being several inductive steps away from the problem stem. Conversely, a precue can facilitate item solution if the precue is familiar and relevant, if the item is fairly difficult to start with, if the item is of a rigidly structured type

where the precue can be readily recognized as helpful, and if the precue is inductively close to the item stem.

Correlations of Experimental Task with Psychometric Test Scores

Table 7 shows correlations between experimental task and psychometric test scores. Correlations are shown both for the individual tests, and for approximation factor scores based on a varimax-rotated principal-axis factor analysis of the data. The factor analysis yielded two interpretable factors with eigenvalues greater than 1: a reasoning factor (DAT Verbal Reasoning, Cattell Culture-Fair Test of g, Insight Test) and a verbal/perceptual factor (Crossing Out A's, Vocabulary).

These correlations address three basic questions. The first question concerns convergent validation, whereas the second and third questions concern discriminant validation. First, are scores on the experimental tasks related to scores on the psychometric tests? One would expect most of the correlations to be statistically significant, but moderate, as the abilities tapped by the experimental tasks should be related but nonidentical to those tapped by the psychometric tests (see Hunt, Frost, & Lunneborg, 1973, for similar logic). Second, do the correlations with the reasoning tests (and factor) differ from those with the verbal/perceptual factor? In particular, the reasoning tests were chosen to measure fluid abilities similar to those tapped by the experimental tasks, whereas the verbal and perceptual tasks were chosen to measure crystallized abilities different from those tapped by the experimental tasks (see Hunt, Lunneborg, & Lewis, 1975, for an information processing analysis of verbal or crystallized abilities). Third, are the correlations for the uncued and cued conditions with the psychometric tests different? The triarchic theory predicts that the more nonentrenched (precued) items should be the better measures of intelligence, and the psychometric tests of fluid ability were chosen for being among the best measures of fluid reasoning ability. Hence, the correlations for the precued items might be expected to be higher than those for the uncued items.

The answers to these questions are fairly straightforward. First, most of the correlations in the table are statistically significant, and generally low to moderate in magnitude. Second, correlations of the experimental tasks with the reasoning tasks are clearly higher than those with the verbal/perceptual tasks. For the factor scores, for example, all of the correlations of reaction times and of error rates with the reasoning factor (ranging from −.31 to −.46) are statistically significant, whereas none of the correlations with the verbal/perceptual factor (ranging from −.05 to −.17) are significant. Thus, the experimental tasks do appear to tap abilities related to those tapped by the psychometric tests. Third, six of six comparisons between correlations of uncued versus precued response times and error rates with the psychometric reasoning tests come out higher for the precued than for the uncued items. In sum, the experimental tasks showed the predicted pat-

Table 6
Difference Scores (Precued − Uncued) for Reaction Time of Correct Responses RT (in sec); and for Error Rate ER (in percent)

	Analogies		Classifications		Series	
	RT	ER	RT	ER	RT	ER
Familiar Relevant	.30	1.7	−.67	−3.5	−.40	−6.1
Familiar Irrelevant	.66	1.3	.39	1.9	1.05	1.4
Novel Relevant	1.15	5.4	2.38	16.6	.21	0.8
Novel Irrelevant	1.09	4.5	.33	5.1	1.63	4.9

Table 7
Correlations of Response Times (RTs) and Error Rates (ERs) With Ability Tests

	Overall		Uncued		Cued	
	RT	ER	RT	ER	RT	ER
DAT Verbal Reasoning	−.21	−.58‡	−.19	−.45‡	−.23	−.57‡
Cattell Abstract Reasoning	−.38†	−.06	−.32*	−.02	−.41†	−.08
Insight Problems	−.24	−.41‡	−.20	−.36†	−.26*	−.39†
Crossing Out A's	−.09	−.06	−.16	−.10	−.04	.04
Extended Vocabulary (Level 3)	−.07	−.29*	−.10	−.26*	−.04	−.26*
Reasoning Factor	−.36†	−.46‡	−.31*	−.36†	−.39†	−.45‡
Verbal/Perceptual Factor	−.10	−.14	−.17	−.10	−.05	−.15

Note—Only correct response times were included. $*p < .05.$ $†p < .01.$ $‡p < .001.$

terns of convergent-discriminant validity with respect to the psychometric tests.

DISCUSSION

This experiment investigated information processing during the solution of analogies, classifications, and series completions either taking the standard form or preceded by precuing information that could be novel or familiar, and relevant or irrelevant. The data analyses investigated both the internal and external validities of the tasks.

With respect to internal validation, it was found that for analogies and classifications, subjects take longer to process irrelevant than relevant information if the precue is familiar, but they take longer to process relevant than irrelevant information if the precue is novel. For series completions, however, both novelty and irrelevance add time to the processing of information, with the time for irrelevance greater than that for novelty. Figure 1 shows a general information processing model—formalized as a flow chart—that provides a plausible account of information processing in all three tasks. The tasks differ only in the proportion of time the various steps take, which is a factor of item type, cue difficulty, and so on. These variables affect what specific information processing model applies to a given task. For instance, as previously discussed, applying novel relevant precues is easy for series completions relative to the other two tasks, so the resulting model is additive rather than interactive.

In the general model shown in the flow chart (which applies to precued items only; for standard models, see Sternberg & Gardner, 1983), subjects first read the precue. If the precue is familiar, then the subjects access category information needed for problem solution. Subjects then request the item stem. They then do a series of relevance checks, self-terminating as soon as they find how the information is relevant to item solution. At this point, they incorporate the precue if it is relevant, solve the problem, and then respond (see Sternberg & Gardner, 1983, for how similar uncued problems are solved). If successive tests fail to show the relevance of the precue, subjects discontinue relevance checking, reject the precue as irrelevant, solve the problem, and respond.

If the precue is novel rather than familiar, then subjects have to create new category information for the counterfactual. After requesting the item stem, they then do a series of relevance checks for the novel precue. If, after a series of checks, they are unable to see any relevance for the precue, they reject the cue, solve the problem, and respond. If, however, the precue is relevant, then they combine the precue with the stem, a nontrivial process since the precue not only overrides their previous knowledge, but also changes the answer to the problem. After combining the cue with the stem, the subjects solve the problem and respond.

With respect to external validation, it was found that the nonentrenched induction tasks overlapped with psychometric tests in terms of abilities measured, that the abil-
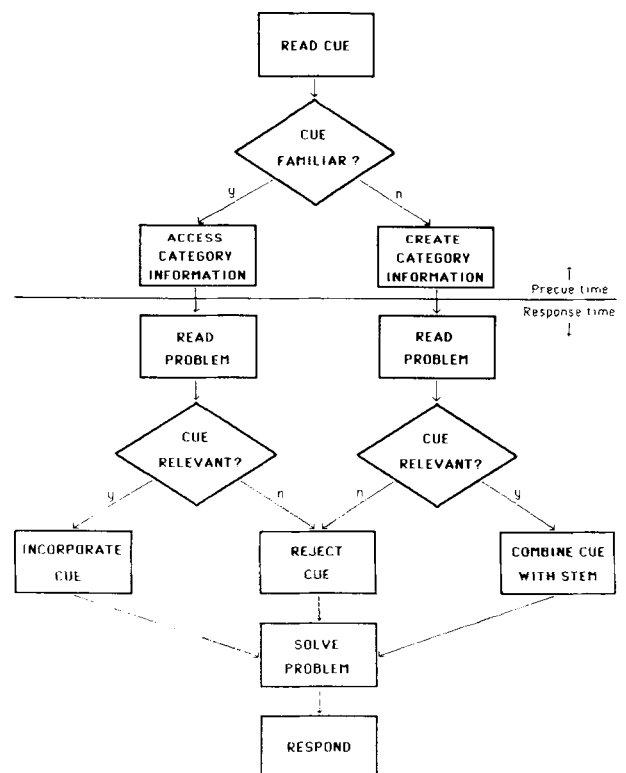
Figure 1. Model of information processing for precued induction problems.

ities measured were fluid rather than crystallized, and that the precued (more nonentrenched) items were better measures of fluid abilities than were the uncued ones. The nonentrenched induction tasks thus fulfill their original goal of being essentially IQ-test-like in their surface structure, in that they can be answered rather quickly with an unambiguous keyed answer. But they are unlike standard items on intelligence tests in being more nonentrenched—or unusual—in their information processing requirements.

The results of this experiment are consistent with the notions of Raaheim (1974), Snow (1980), and Sternberg (1982, 1985) that relatively nonentrenched tasks provide particularly apt measures of intelligence and, particularly, of fluid intelligence. The ability to cope with relative novelty is an important aspect of intelligence, and it can be measured efficiently and in a theoretically based way using the nonentrenched analogies, classifications, and series completions used in the present research.

REFERENCES

Asimov, I. (1979). The singing bell. In I. Asimov, M. H. Greenberg, & C. G. Waugh (Eds.), *The thirteen crimes of science fiction* (pp. 221-237). Garden City, NY: Doubleday.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1973). *Differential aptitude tests*. New York: Psychological Corporation.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston, MA: Houghton-Mifflin.

Cattell, R. B., & Cattell, A. K. (1963). *Test of g: Culture fair, Scale 3*. Champaign, IL: Institute for Personality and Ability Testing.

Davidson, J. E., & Sternberg, R. J. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly, 28*, 58-64.

French, J. W., Ekstrom, R. B, & Price, I. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized ability intelligences. *Journal of Educational Psychology, 57*, 253-270.

Hunt, E. B., Frost, N., & Lunneborg, C. (1973). Individual differ-

ences in cognition: A new approach to intelligence. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 87-122). New York: Academic Press.

Hunt, E. B., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology, 7*, 194-227.

Kalmar, D., & Sternberg, R. J. (1985, November). *When will the milk spoil? Everyday induction in daily life*. Paper presented at the 26th annual meeting of the Psychonomic Society, Boston, MA.

Marr, D. B., & Sternberg, R. J. (1986). Analogical reasoning with novel concepts: Differential attention of intellectually gifted and non-gifted children to relevant and irrelevant novel stimuli. *Cognitive Development, 1*, 53-72.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology, 12*, 252-284.

Raaheim, K. (1974). *Problem solving and intelligence*. Oslo, Norway: Universitetsforlaget.

Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Vol. 1. Cognitive process analyses of aptitude* (pp. 27-63). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1981). Intelligence and nonentrenchment. *Journal of Educational Psychology, 73*, 1-16.

Sternberg, R. J. (1982). Natural, unnatural, and supernatural concepts. *Cognitive Psychology, 14*, 451-458.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J., & Davidson, J. E. (1982, June). The mind of the puzzler. *Psychology Today, 16*, 37-44.

Sternberg, R. J., & Davidson, J. E. (1983). Insight in the gifted. *Educational Psychologist, 18*, 51-57.

Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112*, 80-116.

Tetewsky, S. J., & Sternberg, R. J. (1986). Conceptual and lexical determinants of nonentrenched thinking. *Journal of Memory & Language, 25*, 202-225.